

Genetic risk...

by Uopik Miskad

FILE	ISK_FACTORS_FOR_COLORECTAL_CANCER_IN_MULTIETHNIC_INDONESIANS.PDF (626.02K)	WORD COUNT	12388
TIME SUBMITTED	28-DEC-2020 10:59PM (UTC+0700)	CHARACTER COUNT	60112
SUBMISSION ID	1481702605		

4
Noname manuscript No.
(will be inserted by the editor)

1 Genetic Risk Factors for Colorectal Cancer in
2 Multiethnic Indonesians

3 Irawan Yusuf · Upik A. Miskad ·
4 Ronald E. Lusikooy · Arham Arsyad ·
5 Akram Irwan · George Mathew · Ivet
6 Suriapranata · Rinaldy Kusuma ·
7 Bens Pardamean · Muhamad Fitra
8 Kacamarga · Arif Budiarto · Tjeng
9 Wawan Cenggoro · Carissa I.
10 Pardamean · Christopher McMahan ·
11 Chase Joyner · James W. Baurley ✉

12 Received: date / Accepted: date

13 **Abstract Purpose:** Colorectal cancer is a common cancer in Indonesia, yet
14 it has been understudied. We conduct a genome-wide association study focused
15 on evaluation and discovery of colorectal cancer risk factors in Indonesians.

16 **Methods:** We administered detailed questionnaires and collecting blood sam-
17 ples from 162 colorectal cancer cases throughout Makassar, Indonesia. We also
18 established a control set of 193 healthy individuals frequency matched by age,
19 sex, and ethnicity. A genome-wide association analysis was performed on 84
20 cases and 89 controls passing quality control. We evaluated known colorectal
21 cancer genetic variants using logistic regression and established a genome-wide
22 polygenic risk model using a Bayesian variable selection technique.

23 **Results:** We replicate associations for rs9497673, rs6936461 and rs7758229
24 on chromosome 6; rs11255841 on chromosome 10; and rs4779584, rs11632715,
25 and rs73376930 on chromosome 15. Polygenic modeling identified 10 SNP as-
26 sociated with colorectal cancer risk.

27 **Conclusions:** This work helps characterize the relationship between variants

Irawan Yusuf · Upik A. Miskad · Ronald E. Lusikooy · Arham Arsyad · Akram Irwan
Faculty Medicine; Hasanuddin University, Indonesia

George Mathew · Ivet Suriapranata · Rinaldy Kusuma
Mochtar Riady Institute for Nanotechnology; Pelita Harapan University, Indonesia

Bens Pardamean · Muhamad Fitra Kacamarga · Arif Budiarto · Tjeng Wawan Cenggoro ·
Carissa I. Pardamean
Bioinformatics & Data Science Research Center; Bina Nusantara University, Indonesia

Christopher McMahan · Chase Joyner
School of Mathematical and Statistical Sciences; Clemson University, USA

James W. Baurley ✉
Bioinformatics and Data Science Research Center; Bina Nusantara University, Indonesia
E-mail: baurley@binus.edu; (62-21) 534-5830 ext. 1700

28 in the *SCL22A3*, *SCG5*, *GREM1*, and *STXBP5-AS1* genes and colorectal
29 cancer in a diverse Indonesian population. With further biobanking and inter-
30 national research collaborations, variants specific to colorectal cancer risk in
31 Indonesians will be identified.

32

33 **Keywords** colorectal cancer · genome-wide association study · Indonesia ·
34 polygenic risk score

35 1 Introduction

36 Colorectal cancer is one of the most common cancers in the world and a
37 leading cause of cancer-related deaths [1][2]. There is growing evidence that
38 colorectal cancer rates are changing in Asian countries, but the causes are still
39 under investigation [3][4]. Colorectal cancer is now one of the top three cancers
40 in many Asian countries [5]. Currently, Asia contributes to 48% of the total
41 number of new colorectal cancer cases in the world, of which the majority
42 are found in Eastern Asia [6]. Specifically in Indonesia, the age-standardized
43 incidence for males and females has been reported as 15.9 and 10.1 per 100,000
44 respectively [7].

45 The contribution of heritable factors towards colorectal cancer occurrence
46 is estimated to be between 12-35%. However, germline mutations that are
47 highly penetrant contribute less than 5% to colorectal cancer [8]. Nonetheless,
48 increasing evidence is finding that heritability plays a potential, crucial role in
49 colorectal cancer pathogenesis. Currently, mutations in 14 genes are suspected
50 to underlie different subtypes of colorectal cancer, including mutations in the
51 APC that increases predisposition to familial adenomatous polyposis (FAP)
52 and defects in mismatch repair genes associated with Lynch Syndrome [8]. Re-
53 cent genome-wide association studies have identified common genetic variants
54 linked to colorectal cancer predisposition, highlighting a greater association
55 between heritable risk and the disease. Thus far, over 40 genetic variants have
56 been identified, within several well-known biological pathways that have been
57 shown to be highly relevant to oncogenesis, including the TGF-beta/BMP
58 pathway and the mitogen-activated protein kinases (MAPK) pathway [8].

59 However, many of these colorectal cancer genetic associations were dis-
60 covered in European-ancestry populations but do not replicate well in other
61 ancestry groups, demonstrating the need for studies in diverse populations
62 worldwide [9]. The Asia Colorectal Cancer Consortium was initiated in 2009
63 among East Asian nations and has successfully identified novel relevant, ge-
64 netic regions [10, 11]. However, colorectal cancer cases from South East Asian
65 cohorts, have been under represented.

66 Given the changes in colorectal cancer rates in Asia and the differences in
67 risk factors present in ethnically diverse South East Asia, we present results
68 of the first genomic association study of colorectal cancer in Indonesia. We
69 present results from the initial phase of this study, focused on cases from
70 South Sulawesi, Indonesia.

71 2 Methodology

72 2.1 Study participants

73 162 colorectal cancer cases were recruited from 7 hospitals throughout Makas-
74 sar, Indonesia between 2014 and 2016. The hospitals were Wahidin Sudirohu-
75 sodo Hospital, Hasanuddin University Hospital, Ibnu Sina Hospital, Akademis
76 Hospital, 15 Restelina Hospital, Stella Maris Hospital, and Hikmah Hospital. 193
77 controls were frequency matched to cases on age category, sex, and ethnicity.
78 This research was approved by the Hasanuddin University Ethical Committee
79 (registration number: UH 15040389).

80 2.2 Data and DNA sample collection

81 Questionnaires and medical records were recorded into study data collection
82 forms and entered into a study database. The case forms contained 382 ques-
83 tions and the control forms contained 319 questions. The forms included in-
84 formation on demographics, cancer history in the family, smoking behavior,
85 alcohol use, and detailed dietary history. For colorectal cancer cases, the forms
86 collected information on cancer symptoms, staging (post operation), tumor,
87 location, histopathology, and type of surgery. The database was managed by
88 the Bioinformatics and Data Science Research Center (BDSRC) at Bina Nu-
89 santara University (Jakarta, Indonesia). A blood sample was collected from
90 the basilic/cephalic vein on all participants for genotyping. These blood sam-
91 ples were stored in Hasanuddin University Laboratory at minus 20 degrees
92 Celsius.

93 2.3 Genotyping and imputation

94 DNA was extracted from samples at Mochtar Riady Institute for Nanotech-
95 nology (MRIN) Laboratory (Tangerang, Indonesia). Genomic DNA was ex-
96 tracted from 200 μ L of whole blood sample using the QIAamp DNA Mini Kit
97 (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA
98 concentration was determined using NanoDrop ND-1000 spectrophotometer,
99 version 3.3 (Thermo Fisher Scientific, Wilmington, DE, USA) and adjusted
100 to a concentration of 20 ng/ μ L. The quality of DNA extracted was verified
101 by purity index of OD260/OD280 (1.8-2.0) and OD260/OD230 ($>$ 1.5). The
102 DNA was inspected through Gel Electrophoresis using 1% molecular biology
103 grade Agarose (Biorad, Hercules, CA, USA). Extracted DNA were sent to
104 RUCDR Infinite Biologics for genotyping (Piscataway, NJ, USA) under Mate-
105 rial Transfer Agreement (MTA) approved by the Indonesian Health Ministry
106 (registration number: LB.02.01/I/12749/2016).

107 DNA samples from study cases and controls were genome-wide genotyped
108 on the Smokescreen Genotyping Array [12]. Using 200 ng of genomic DNA, ar-
109 ray plates were prepared using the Axiom 2.0 Reagent Kits and then processed

110 on the GeneTitan MC instrument (Thermo Fisher Scientific, Wilmington, DE,
111 USA). Analysis of the raw data was performed using Affymetrix Power tools
112 (APT) v-1.16 according to the Affymetrix best practices workflow. 183 sam-
113 ples remained after completing these steps. Additional steps were performed
114 using SNPfilter to identify and select best performing probe sets and high
115 quality SNPs for downstream analysis. 524,765 SNPs remained after QC filter-
116 ing. Additional sample quality control included verifying concordance of study
117 replicates, checking for unintentional duplicates and unexpected relatives, and
118 verifying genetic versus reported gender. After filtering samples with miss-
119 ing covariates, 173 samples (84 cases and 89 controls) remained for statistical
120 analysis.

121 Genome-wide imputation was performed on the Michigan Imputation Server
122 v1.0.2 [13]. Briefly, quality controlled study genotypes were reported on the
123 forward strand and uploaded in vcf format. 1000 Genomes Phase 3 [14] was
124 selected as a reference panel, phasing was performed using Eagle v2.3 [15], and
125 allele frequencies were compared against the 1000 Genomes East Asian (EAS)
126 populations. The server automatically excludes variants with alleles other than
127 (A,C,T,G), variants with duplicate positions, indels, monomorphic sites, and
128 allele mismatches with the reference panel.

129 2.4 Statistical analysis

130 2.4.1 Ancestry analysis

131 Ancestry categories were estimated from 5,515 ancestry informative markers
132 contained on the Smokescreen Genotyping Array using fastStructure 1.0 [16].
133 Combining study and reference data from the 1000 Genomes Project Phase
134 3, we estimated the ancestry proportions of East Asian (EAS), South Asian
135 (SAS), European (EUR), and African (AFR).

136 2.4.2 Genome-wide association analysis

137 We filtered out variants with poor imputation quality (< 0.3) and rare variants
138 (minor allele $< 1\%$). We then performed a marginal analysis of the remain-
139 ing SNP genotype dosages fitting logistic regression models, with sex, age,
140 body mass index, smoking status and estimated ancestries proportions (i.e.,
141 SAS, EUR, AFR) as covariates. The threshold for statistical significance in the
142 discovery scan was set at the historical traditional genome-wide value of $5E-8$.

143 We queried the scan results for markers previously reported to be associ-
144 ated with colorectal cancer. These variants were identified through previous
145 genotyping in an independent sample of South Sulawesi colorectal cancer cases
146 (R. Kusuma, I. Suriapranata, personal communication) and a recent catalog
147 of colorectal cancer SNPs for a genome-wide association scan in Hispanics [17].
148 The source and annotation for these variants are provided in Supplementary

149 Table 3. Variants with evidence of replication (p -value < 0.05) were flagged
150 for further investigation.

151 We also developed a polygenic model considering the joint effect of mul-
152 tiple genetic variants on colorectal cancer. We selected the top 200 SNPs,
153 based on Bayes factors [18], as candidate predictors in this joint model. Bayes
154 factors were computed for the marginal versus the null models for each SNP
155 while controlling for gender, age, BMI, and smoking status. To jointly model
156 these variants, we use a Bayesian variable selection technique. In particular,
157 we fit a logistic regression model utilizing shrinkage priors for each of the
158 explanatory variables; i.e., the covariates listed above as well as the remain-
159 ing candidate SNPs. In this analysis, the generalized double Pareto shrinkage
160 prior of [19] was specified and the parameters of the joint model were esti-
161 mated via a maximum a posteriori (MAP) estimator [19] which was obtained
162 via an expectation-maximization (EM) algorithm [20]. The MAP estimator
163 under these specifications simultaneously completes parameter estimation and
164 variable selection by obtaining a sparse estimator [21]; i.e., some of the regres-
165 sion coefficients are estimated to be identically equal to zero thus removing
166 the effect of the corresponding explanatory variable. The EM algorithm was
167 developed following the techniques illustrated by [19,22] and the regulariza-
168 tion parameters were selected via the Bayesian information criterion [23]. All
169 statistical analysis was performed in R [24].

170 3 Results

171 3.1 Characteristics of study sample

172 The characteristics of the colorectal cancer cases and controls are summarized
173 in Table 1. The mean age of the colorectal cancer cases was 54 years. The
174 majority of cases were male (57%). Among ethnicities, most cases were self-
175 reported Bugis (44%) or Makassar ethnicity (27%). Controls appeared to be
176 adequately frequency matched to cases by age, sex, and ethnicity ($p > 0.05$).
177 Colorectal cancer cases had lower average body mass index (BMI) and were
178 more likely to be smokers than controls ($p < 0.01$). Estimated genetically, the
179 majority of both cases and controls were of East Asian ancestry. 87% of the
180 cases had late stage cancer (III or IV) which unfortunately is consistent with
181 recent reports in Indonesia [25]. As seen in other studies, the most common
182 colorectal cancer site was rectum (43%) [26,27].

183 3.2 Genome-wide association analysis

184 As expected given the sample size, no SNPs met the historical cutoff set for
185 genome-wide significance (Supplementary Figures 6 and 7). The summaries for
186 all variants with a marginal p -value $< 5E-5$ are included in the supplementary
187 materials (Table 4). These include two intergenic SNPs and two SNPs in the
188 *MRO* gene on chromosome 18.

189 Results for previously reported colorectal cancer SNPs are presented in
190 Figure 1 and Supplementary Table 3. There is evidence of replication for the
191 following genetic variants: rs9497673, rs6936461 and rs7758229 on chromosome
192 6; rs11255841 on chromosome 10; and rs4779584, rs11632715, and rs73376930
193 on chromosome 15. The regions are characterized in Figures 2, 3, 4, and 5.
194 The pattern of associations is rather diffuse in the *STXBP5-AS1* (*STXBP5*
195 Antisense RNA 1) and *SLC22A3* genes of chromosome 6, representing the
196 correlation among the variants in these regions (Figures 2 and 3). Similarly,
197 the association pattern tapers along chromosome 10. The strongest associa-
198 tion pattern can be found on chromosome 15. This region has a more defined
199 peak than the other regions with associations spanning two genes: *SCG5* (se-
200 cretogranin V) and *GREM1* (gremlin 1, DAN family BMP antagonist).

201 The polygenic analysis identified 10 SNPs which appear to have a relatively
202 strong association (i.e., large effect size) with the risk of developing colorectal
203 cancer. Five of these SNPs lie in intergenic regions; three lie in introns of
204 *ARHGEF3*, *PLCG2*, and *RGMB*; one is a deletion in *PIGN*; and one is an
205 insertion in *SHISA9*.

206 4 Discussion

207 This study represents the first genome-wide analysis of a South Sulawesi pop-
208 ulation in Indonesia. Strengths of the study include the building of a colorectal
209 cancer research program in Indonesia, the extensive questionnaire for assessing
210 non-genetic risk factors, and genome-wide genotyping across diverse ethnici-
211 ties. Limitations of the study include the sample size, which restricts the anal-
212 ysis to previously identified colorectal cancer markers and challenges shared by
213 case-control study designs. For instance, the controls may represent different
214 groups than cases. We attempted to account for this by frequency matching
215 on age, sex, and ethnicity. Additionally, the timing of assessments need to be
216 considered in interpreting the results. Given screening programs are still being
217 developed in Indonesia, the majority of the cases had late stage colorectal can-
218 cer, stage III and IV. When BMI was assessed in these patients they already
219 had significant weight lose, thus the direction of the effect is different than
220 what one might expect.

221 Several previously identified colorectal cancer associated SNPs replicated
222 in this population. And we can begin characterizing these regions by examining
223 neighboring variants.

224 rs7758229 within *SLC22A3* on chromosome 6 was originally identified and
225 subsequently replicated in large case-control study of a Japanese population
226 (OR of 1.3) [28]. Interestingly, in a subsequent study in a Chinese population,
227 this SNP was not associated with colorectal cancer (OR of 0.95) [29]. However,
228 in S. Sulawesi, we detect a statistically significant association with colorectal
229 cancer ($p=0.009$, OR of 2.2). Given these difference among East Asians, fur-
230 ther work to understand variation in *SLC22A3* and colorectal cancer is needed.
231 *SLC22A3* encodes for the protein OCT3, which is an organic cationic trans-

232 porter. While OCT3/SLC22A3 is well characterized within neurochemistry,
233 it has been found to play a role within oncology as well. The upregulation
234 of SLC22A3 in head and neck squamous cell carcinoma is associated with
235 improved prognosis while the downregulation of SLC22A3 leads to enhanced
236 metastasis and invasion of the tumor [30]. SLC22A3 has also been implicated
237 in the pathogenesis of prostate cancer and its expression is elevated in these
238 neoplastic tissues [31]. The level of OCT3/SLC22A3 expression has also been
239 linked to the level of patient responsiveness towards cancer treatments [32]; in
240 particular, platin-based cytotoxic cancer treatments in colorectal cancer [33]
241 patients, as well as head and neck squamous cell carcinoma patients [30].

242 Intergenic variant rs11255841 on chromosome 10 was identified in an col-
243 orectal cancer GWAS of European ancestry individuals [34] and has replicated
244 in a Japanese study and a large meta-analysis with nearly 37,000 cases [35,
245 36]. With the risk allele of T, this variant had an odds ratio of 2.2 in our study,
246 while previous reports had an odds ratio of 1.1-1.2.

247 The region on chromosome 15 nearby *SCG5* and *GREM1* have been flagged
248 in multiple GWAS, e.g., [37]. We replicated colorectal cancer associations for
249 rs4779588 (p=0.018), rs11632715 (p=0.004), and rs73376930 (p=0.010). Inter-
250 estingly, the smallest p-value in the region was rs10083612 within an intron of
251 *SCG5* (p=1.61e-5, see Figure 5). The role of *SCG5* in colorectal cancer has not
252 been well characterized, while much is known about its neighbor *GREM1*'s role
253 in colorectal cancer. *GREM1*, which is one of the antagonists of the bone mor-
254 phogenetic proteins (BMPs) found within the TGF-beta signaling pathway,
255 has been found to be important for the survival and proliferation of several
256 types of cancers [38]. In particular, modulated expression of *GREM1* is found
257 in cancer-associated stromal cells. *GREM1* is also found to be a proangio-
258 genic factor, suggesting a role in cancer development when it is upregulated
259 [39]. *SCG5* and *GREM1* genes have been found to be associated with poly-
260 posis syndromes that are associated with colorectal cancer [40]. A duplication
261 that spans the 3'end of *SCG5* and the immediate, adjacent upstream region
262 of *GREM1* is associated with hereditary mixed polyposis syndrome (HMPS)
263 as well as tumorigenesis in juvenile polyposis. This duplication results in a
264 40-kb extra segment that leads to the upregulation of *GREM1* expression.
265 The duplication is the basis for an autosomal dominant HMPS condition that
266 is prevalent among the Ashkenazi Jewish population and is a recommended
267 biomarker/genetic test to detect CRC in this population. Aberrant expres-
268 sion of *GREM1* has also been shown to underlie oncogenesis within the large
269 intestines and colon [41].

270 Two of the previously identified colorectal cancer markers replicate in this
271 study (rs6936461 and rs9497673; see Table 3). These SNPs are located in the
272 intronic regions of *STXBP5-AS1* on chromosome 6. Using bioinformatics tools,
273 it is predicted that changes from T to A in rs6936461 and A to G in rs9497673,
274 has the potential to alter the splicing of the gene [42]. *STXBP5-AS1* is a long
275 non-coding (lncRNA) gene. lncRNAs drive many important cancer phenotypes
276 through their interactions with other cellular macromolecules including DNA,
277 protein, microRNA and mRNA. The different expression of lncRNAs in col-

278 rectal cancer indicate that lncRNAs are involved in all stages of colorectal
279 cancer. In colorectal cancer pathogenesis, lncRNAs are implicated in a variety
280 of signaling pathways including the Wnt/-catenin signaling pathway, epider-
281 mal growth factor receptor (EGFR)/insulin-like growth factor type I receptor
282 (IGF-IR) signaling pathway, KRAS and phosphatidylinositol-3-kinase (PI3K)
283 pathways, transforming growth factor-beta (TGF-) signaling pathway, p53 sig-
284 naling pathway, and the epithelial-mesenchymal transition (EMT) pathway
285 [43]. While it is still unclear how *STXBP5-AS1* contributes to colon carcino-
286 genesis, in a study involving 1067 breast cancer samples, Guo et al. identified
287 *STXBP5-AS1* among lncRNA genes which play a role in predicting the prog-
288 nostic survival with good sensitivity and specificity. The lncRNAs may act
289 as competing endogenous RNAs (ceRNAs) and interfere in the binding of
290 miR-190b to certain targets such as ERG, STK38L, and FNDC3A and thus
291 contribute to breast cancer pathogenesis [44]. *STXBP5-AS1* may act similarly
292 in colorectal cancer; it may hinder the binding of microRNAs to their target
293 genes and subsequently modulate colorectal cancer tumorigenesis.

294 Interestingly, *STXBP5-AS1* was identified among genes that are methy-
295 lated in buccal samples in a genome-wide screen for cigarette smoke exposure,
296 indicating its possible role in smoking-related diseases [45]. Since there is a
297 significant difference in smoking status between cases and controls in our co-
298 hort, it is plausible that genetic variants associated with tobacco smoke are
299 also associated with the presence of colorectal cancer in our study population.

300 The polygenic model represents a strategy for jointly modeling SNP effects
301 in a GWAS and development of risk prediction models in a specific population.
302 These models can be used to estimate an individual's risk of colorectal cancer
303 based on easily obtainable genotypes. While most of the variants flagged in
304 the polygenic model are novel, the gene *ARHGEF3* has been implicated in
305 promoting nasopharyngeal carcinoma in Asians [46]. *RGMB* has been shown
306 to promote colorectal cancer growth [47]. Additional samples will enable us to
307 refine and validate a polygenic colorectal cancer risk model in Indonesians.

308 5 Conclusion

309 We demonstrate replication of several colorectal cancer genetic risk factors
310 in an Indonesian population. This study provides rationale for additional data
311 collection in this population to characterize these regions more precisely and
312 identify genetic risk factors unique to this diverse population.

313 **Acknowledgements** We would like to acknowledge Bina Nusantara and Hasanuddin Uni-
314 versity for funding this study, MRIN Laboratory for DNA Extraction, RUCDR Infinite
315 Biologics for DNA processing and genotyping, BioRealm for support of the Smokescreen
316 Genotyping Array, Research credits from Amazon Web Services (AWS) and generous contri-
317 butions from NVIDIA and the AI R&D Center at Bina Nusantara University for computing
318 and database support.

319 **6 Tables**

Table 1 Characteristics of South Sulawesi colorectal cancer cases and controls

		Cases	Controls	P
		N = 89	N = 84	
Age		53.8 (13.2)	50.5 (14.5)	0.12
Gender				>0.99
	Female	38 (42.7%)	36 (42.9%)	
	Male	51 (57.3%)	48 (57.1%)	
Ethnicity				0.68
	Bugis	39 (43.8%)	45 (53.6%)	
	Makassar	24 (27.0%)	23 (27.4%)	
	Mandar	2 (2.3%)	1 (1.2%)	
	Toraja	10 (11.2%)	8 (9.5%)	
	Non South Sulawesi	9 (10.1%)	4 (4.8%)	
	Non Sulawesi	5 (5.6%)	3 (3.6%)	
BMI		21.2 (3.1)	24.5 (3.6)	<0.01
Smoking Status				<0.01
	Smoker	39 (43.8%)	15 (17.9%)	
	Non smoker	50 (56.2%)	69 (82.1%)	
Ancestry (Estimated)				
	East Asian (EAS)	0.92	0.94	0.02
	South Asian (SAS)	0.07	0.05	0.15
	African (AFR)	<0.01	<0.01	0.02
	European (EUR)	0.01	0.01	0.36
Cancer Site				
	Right Colon	15 (16.9%)	-	
	Transversum	9 (10.1%)	-	
	Left Colon	1 (1.12%)	-	
	Sigmoid	26 (29.2%)	-	
	Rectum	38 (42.7%)	-	
Staging				
	I	3 (3.4%)	-	
	II	9 (10.1%)	-	
	III	62 (69.7%)	-	
	IV	15 (16.9%)	-	

Table 2 Polygenic risk model learned from colorectal cancer data. Presented results include the chromosome (Chr) and position of the significant genetic variants, the gene they lie on (Gene), reference allele (Ref), minor allele frequency (MaF), and estimated effect (Estimate).

Description	Chr	Position	Gene	Ref	MaF	Estimate
Intercept						0.90
Gender						0.00
Age						-3.75
BMI						0.00
Smoking						1.32
rs11919079	3	57086348	Intron:ARHGEF3	G	0.07	2.40
rs4888186	16	81947156	Intron:PLCG2	C	0.08	0.85
rs11016111	10	129963848	Intergenic	C	0.34	-1.32
rs77657157	5	98125016	Intron:RGMB	G	0.05	1.95
-	18	59822981	Deletion:PIGN	TC	0.19	-1.39
rs17066763	5	164113078	Intergenic	T	0.12	1.65
rs2446103	6	77328692	Intergenic	A	0.04	1.22
rs7219420	17	45800299	Intergenic	T	0.36	1.32
-	16	13018917	Insertion:SHISA9	C	0.11	1.67
rs78165118	3	12816282	Intergenic	A	0.03	2.13

320 **7 Figures**

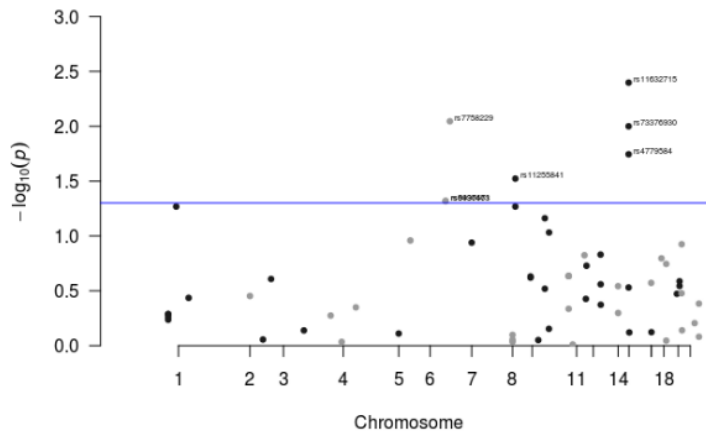


Fig. 1 Results for known colorectal cancer susceptibility SNPs. Variants with p-values < 0.05 were flagged for further investigation.

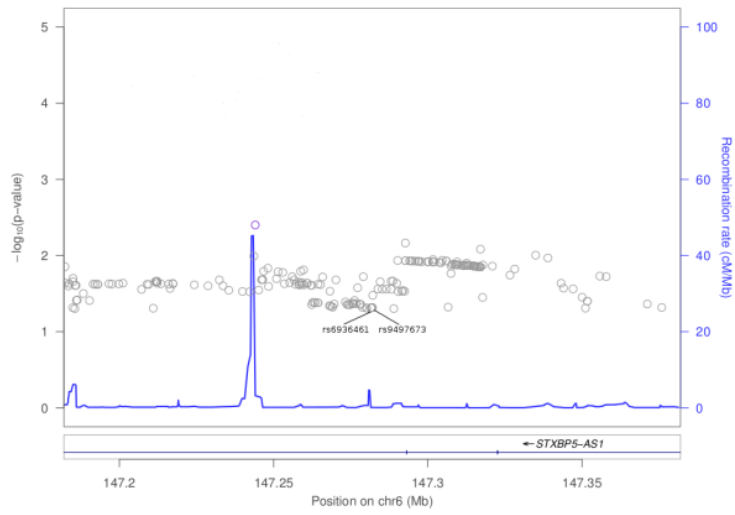


Fig. 2 Association plot for 100kb region flanking rs6936461 on chromosome 6

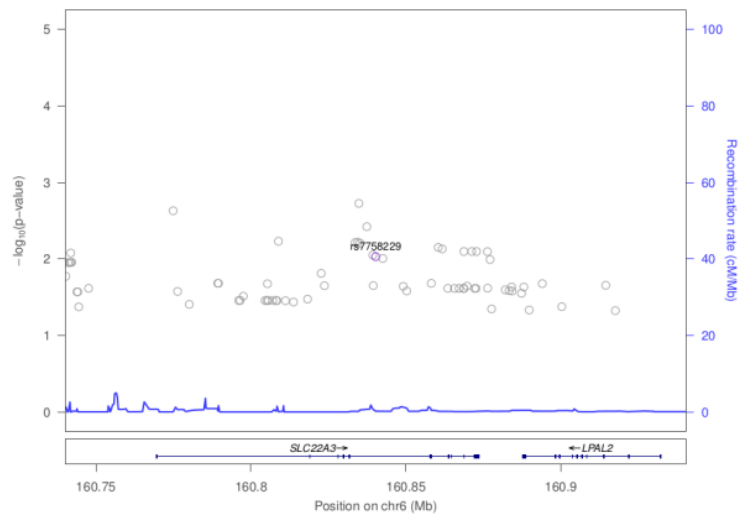


Fig. 3 Association plot for 100kb region flanking rs7758229 on chromosome 6

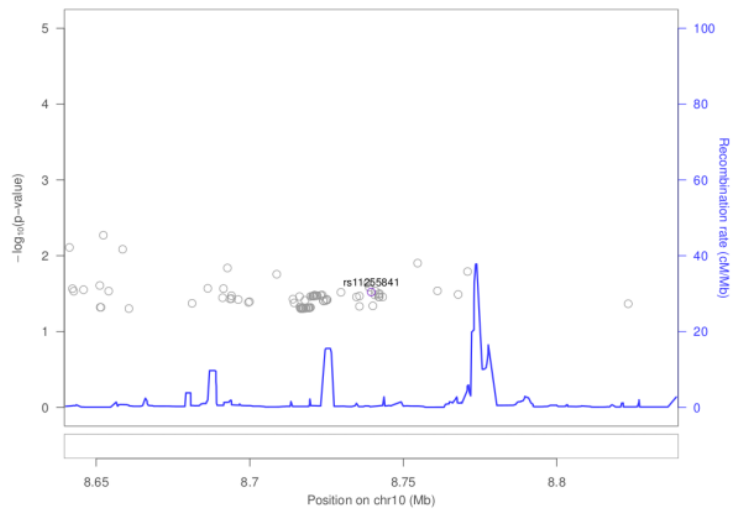


Fig. 4 Association plot for 100kb region flanking rs11255841 on chromosome 10

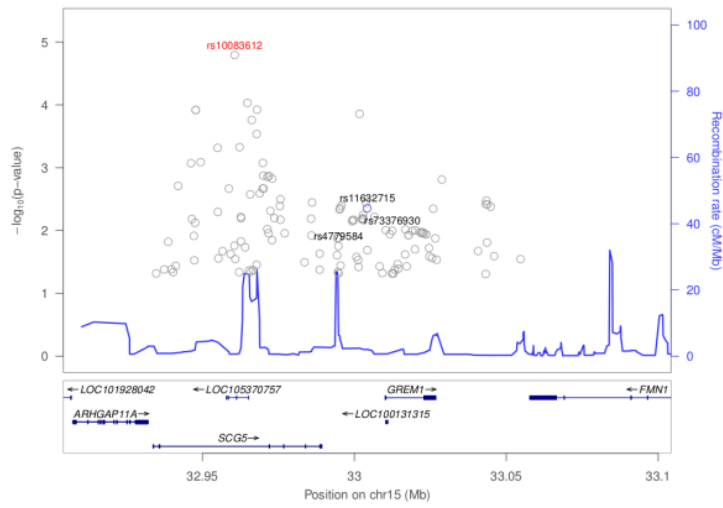


Fig. 5 Association plot for 100kb flanking rs11632715 on chromosome 15. The top associated SNP in the region was rs10083612.

321 **8 Supplementary materials**

Table 3 Results for previously identified colorectal cancer SNPs

Raid	Gene	Chr	Pos	Ref	Alt	Source	MaF	OR	SE	P
rs12124798	Intron:RP11-451O13.1	1	157906955	A	T	[48]	0.091	1.309	0.489	0.581
rs6084686	Intron:RP11-451O13.1	1	157909708	C	G	[48]	0.091	1.327	0.472	0.549
rs6701170	Intron:RP11-451O13.1	1	157913954	C	C	[48]	0.090	1.302	0.476	0.516
rs4971169	Intron:RP11-451O13.1	1	157914330	T	C	[48]	0.088	1.372	0.486	0.516
rs10911251	Intron:LAMC1	1	183081194	A	C	[49,50]	0.345	0.503	0.356	0.054
rs6087758	Intergenic	1	222164948	A	G	[51]	0.093	1.550	0.486	0.367
rs11903757	Intergenic	2	192587204	T	C	[49]	0.161	0.713	0.364	0.353
rs35360328	Intergenic	3	40924962	T	A	[52]	0.142	1.079	0.505	0.880
rs312481	Intron:LRIG1	3	66442435	C	G	[52]	0.158	0.607	0.432	0.247
rs10936599	Synonymous:MYNN	3	169492101	C	T	[51]	0.437	1.104	0.284	0.727
rs7356196	Intergenic	4	84173104	G	A	[48]	0.309	1.229	0.329	0.532
rs3987	Intron:AC108056.1	4	118759055	A	G	[53]	0.396	1.028	0.292	0.925
rs35509282	Intergenic	4	16333405	T	A	[17]	0.320	0.784	0.321	0.448
rs647161	Intron:CTC-203F4.1—CTC-349C3.1	5	134499092	C	A	[10]	0.419	0.923	0.284	0.776
rs1321311	Intron:PI16	6	36622900	C	A	[54]	0.332	0.621	0.298	0.110
rs9497673	Intron:STXBP5-AS1	6	147281518	G	A	[38]	0.289	1.955	0.340	0.048
rs7758229	Intron:SLC22A3	6	160840252	G	T	[55]	0.428	2.244	0.311	0.009
rs10499807	Intergenic	7	68459734	C	T	[48]	0.347	1.792	0.370	0.115
rs10505477	Intron:RP11-382A18.1	8	128407443	A	G	[56,57]	0.301	0.954	0.325	0.884
rs6983267	Intron:RP11-382A18.1	8	128413305	G	T	[56,58,59,60,55]	0.303	0.922	0.318	0.798
rs7014346	Intron:RP11-382A18.1	8	128424792	A	G	[61,51]	0.254	0.966	0.350	0.922
rs10795668	Intergenic	10	8701219	G	A	[59]	0.249	0.525	0.334	0.054
rs11255841	Intergenic	10	8739580	T	A	[50]	0.254	0.462	0.357	0.030
rs10763129	Intron:PCDH15	10	56468045	T	G	[48]	0.124	0.557	0.498	0.240
rs10825383	Intron:PCDH15	10	56473754	G	A	[48]	0.124	0.560	0.488	0.234
rs704017	Intron:RP11-202P11.1	10	80819132	A	G	[11]	0.272	0.958	0.311	0.890
rs1035209	Intergenic	10	101345366	C	T	[50]	0.159	2.105	0.410	0.069
rs11190164	Intergenic	10	101351704	A	G	[52]	0.251	1.460	0.367	0.303
rs12241008	Intron:VTHA	10	114280702	T	C	[62]	0.335	0.895	0.290	0.703
rs11196172	Intron:TCF7L2	10	114726843	G	A	[11]	0.481	1.674	0.307	0.093
rs174537	Intron:C1orf9—RP11-467L20.9	11	61552680	G	T	[11]	0.081	0.684	0.518	0.462
rs4246215	Utr5:FEN1	11	61564299	G	T	[11]	0.087	0.543	0.510	0.232
rs174550	Utr5:FADS1	11	61571478	T	C	[11]	0.087	0.543	0.510	0.232
rs1535	Intron:FADS2	11	61597972	A	G	[11]	0.087	0.543	0.510	0.232
rs3824999	Intron:POLD3	11	74345550	T	G	[54]	0.376	0.991	0.295	0.977
rs3802842	Intron:C1orf92—C1orf93	11	111171709	C	A	[61]	0.324	0.651	0.298	0.150
rs10774214	Intron:RP11-264F23.3—RP11-264F23.4	12	4368352	T	C	[10]	0.253	0.737	0.343	0.375
rs10849432	Intergenic	12	6385727	C	T	[11]	0.130	0.562	0.437	0.187
rs34245511	Intron:LIMA1	12	50573433	G	C	[50]	0.191	0.600	0.353	0.148
rs7136702	Intergenic	12	50880216	T	C	[51]	0.382	1.333	0.263	0.276
rs11169552	Intergenic	12	51155663	C	T	[51]	0.474	1.243	0.273	0.424
rs444235	Intergenic	14	54410919	T	C	[51,59]	0.497	1.352	0.283	0.287
rs1957636	Intergenic	14	54560018	T	C	[59]	0.350	1.267	0.354	0.504
rs16969681	Intergenic	15	32993111	C	T	[59]	0.445	1.356	0.291	0.295
rs4779584	Intergenic	15	32994756	T	C	[63,59]	0.130	0.358	0.433	0.018
rs11632715	Intergenic	15	33004247	G	A	[59]	0.241	4.728	0.546	0.004
rs73376930	Intron:GREM1	15	33012502	A	G	[50]	0.446	3.010	0.428	0.010
rs1851317	Intron:RP11-814P5.1	15	35077786	A	C	[48]	0.423	1.100	0.309	0.758
rs9929218	Intron:CDH1	16	68820946	G	A	[51]	0.410	0.729	0.285	0.268
rs12603526	Intron:NXN	17	800593	T	C	[11]	0.133	0.882	0.399	0.754
rs12458173	Intergenic	18	31430167	G	A	[48]	0.309	1.845	0.436	0.160
rs7229639	Intron:SMAD7	18	46450976	A	G	[11]	0.124	1.079	0.612	0.901
rs4939827	Intron:SMAD7	18	46453463	T	C	[64,61]	0.312	0.645	0.327	0.180
rs10411210	Intron:RHPN2	19	33532300	C	T	[51]	0.153	0.705	0.365	0.337
rs1800469	Intron:TMEM91	19	41860296	A	G	[11]	0.498	1.359	0.288	0.286
rs2241714	Nonsynonymous:B9D2	19	41869392	T	C	[11]	0.491	1.364	0.275	0.259
rs961253	Intergenic	20	6404281	C	A	[51,59]	0.179	0.693	0.379	0.333
rs4813802	Intergenic	20	6699595	T	G	[59,49]	0.239	0.594	0.334	0.119
rs2423279	Intergenic	20	7812350	T	C	[10]	0.416	1.106	0.287	0.725
rs6066825	Intron:PREX1	20	47340117	A	G	[52]	0.416	0.844	0.346	0.624
rs4925386	Intron:LAMA5	20	60921044	T	C	[51,49]	0.260	0.758	0.338	0.414
rs2427308	Intron:CABLES2	20	60969451	C	T	[50]	0.190	1.141	0.619	0.831

Chr: Chromosome
 Pos: Chromosome Position (build 37)
 Ref/Alt: Reference and alternate allele
 MaF: Minor allele frequency
 OR: Odds ratio
 SE: Standard error

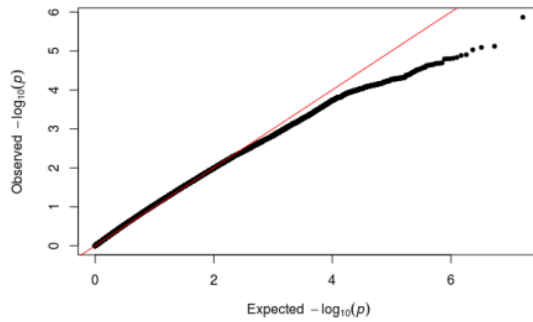


Fig. 6 Observed versus expected distribution of p-values for the colorectal cancer genome-wide scan.

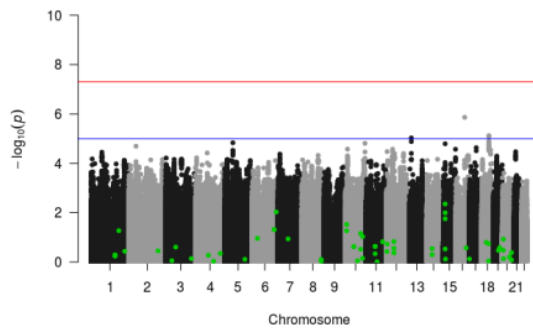


Fig. 7 Manhattan plot for colorectal cancer genome-wide scan. Several SNPs were flagged with p-values $< 1E-5$. Green dots indicate variants flagged in previous genetic association studies.

Table 4 Results from colorectal cancer genome-wide scan. Genetic variants with a marginal p-value < 1E-5.

Rsid	Gene	Chr	Pos	Ref	Alt	MaF	OR	SE	P
rs201447553	Intergenic	13	32081199	G	GT	0.218	10.746	0.536	9.36E-06
-	Intergenic	16	59740698	T	TA	0.423	17.539	0.593	1.36E-06
rs17663205	Utr3:MRO	18	48324484	G	C	0.169	0.089	0.540	7.59E-06
rs56387261	Intron:MRO	18	48325957	C	T	0.170	0.088	0.545	8.12E-06

Chr: Chromosome

Pos: Chromosome Position (build 37)

Ref/Alt: Reference and alternate allele

MaF: Minor allele frequency

OR: Odds ratio

SE: Standard error

References

1. L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA: a cancer journal for clinicians*, vol. 65, no. 2, pp. 87–108, 2015.
2. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: a cancer journal for clinicians*, vol. 66, no. 1, pp. 7–30, 2016.
3. B. Pardamean, J. W. Baurley, C. I. Pardamean, and J. C. Figueiredo, "Changing colorectal cancer trends in asians," *International journal of colorectal disease*, vol. 31, no. 8, p. 1537, 2016.
4. M. A. Pourhoseingholi, "Increased burden of colorectal cancer in asia," *World journal of gastrointestinal oncology*, vol. 4, no. 4, p. 68, 2012.
5. M. P. W. journal of gastrointestinal Oncology and undefined 2012, "Increased burden of colorectal cancer in Asia," *ncbi.nlm.nih.gov*.
6. C. J. Ng, C. H. Teo, N. Abdullah, W. P. Tan, and H. M. Tan, "Relationships between cancer pattern, country income and geographical region in Asia.," *BMC cancer*, vol. 15, p. 613, sep 2015.
7. Globocan, "Estimated Cancer Incidence, Mortality, and Prevalence Worldwide in 2012."
8. U. Peters, S. Bien, and N. Zubair, "Genetic architecture of colorectal cancer," *Gut*, vol. 64, no. 10, pp. 1623–1636, 2015.
9. C. A. Haiman and D. O. Stram, "Exploring genetic susceptibility to cancer in diverse populations," *Current opinion in genetics & development*, vol. 20, no. 3, pp. 330–335, 2010.
10. W.-H. Jia, B. Zhang, K. Matsuo, A. Shin, Y.-B. Xiang, S. H. Jee, D.-H. Kim, Z. Ren, Q. Cai, J. Long, *et al.*, "Genome-wide association analyses in east asians identify new susceptibility loci for colorectal cancer," *Nature genetics*, vol. 45, no. 2, p. 191, 2013.
11. B. Zhang, W.-H. Jia, K. Matsuda, S.-S. Kweon, K. Matsuo, Y.-B. Xiang, A. Shin, S. H. Jee, D.-H. Kim, Q. Cai, *et al.*, "Large-scale genetic study in east asians identifies six new loci associated with colorectal cancer risk," *Nature genetics*, vol. 46, no. 6, p. 533, 2014.
12. J. W. Baurley, C. K. Edlund, C. I. Pardamean, D. V. Conti, and A. W. Bergen, "Smoke-screen: A targeted genotyping array for addiction research," *BMC Genomics*, vol. 17, p. 145, dec 2016.
13. S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, *et al.*, "Next-generation genotype imputation service and methods," *Nature genetics*, vol. 48, no. 10, p. 1284, 2016.
14. . G. P. Consortium *et al.*, "A global reference for human genetic variation," *Nature*, vol. 526, no. 7571, p. 68, 2015.
15. P. Loh, "Eagle v2.4 user manual."
16. A. Raj, M. Stephens, and J. K. Pritchard, "faststructure: variational inference of population structure in large snp data sets," *Genetics*, vol. 197, no. 2, pp. 573–589, 2014.
17. S. L. Schmit, F. R. Schumacher, C. K. Edlund, D. V. Conti, U. Ihenacho, P. Wan, D. Van Den Berg, G. Casey, B. K. Fortini, H. J. Lenz, T. Tusié-Luna, C. A. Aguilar-Salinas, H. Moreno-Macías, A. Huerta-Chagoya, M. L. Ordóñez-Sánchez, R. Rodríguez-Guillén, I. Cruz-Bautista, M. Rodríguez-Torres, L. L. Muñoz-Hernández, O. Arellano-Campos, D. Gómez, U. Alvirde, C. González-Villalpando, M. E. González-Villalpando, L. L. Marchand, C. A. Haiman, and J. C. Figueiredo, "Genome-wide association study of colorectal cancer in Hispanics," *Carcinogenesis*, vol. 37, pp. 547–556, jun 2016.
18. A. E. Raftery, "Approximate bayes factors and accounting for model uncertainty in generalised linear models," *Biometrika*, vol. 83, no. 2, pp. 251–266, 1996.
19. A. Armagan, D. B. Dunson, and J. Lee, "Generalized double pareto shrinkage," *Statistica Sinica*, vol. 23, no. 1, p. 119, 2013.
20. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.
21. J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, p. 1, 2010.

- 378 22. N. G. Polson and J. G. Scott, "Data augmentation for non-gaussian regression models
379 using variance-mean mixtures," *Biometrika*, vol. 100, no. 2, pp. 459–471, 2013.
- 380 23. S. Konishi and G. Kitagawa, "Bayesian information criteria," *Information Criteria and
381 Statistical Modeling*, pp. 211–237, 2008.
- 382 24. R Core Team, *R: A Language and Environment for Statistical Computing*. R Founda-
383 tion for Statistical Computing, Vienna, Austria, 2016.
- 384 25. S. Widjaja and H. Yo, "RM-049Colorectal cancer in Indonesia - a centre report," *Annals
385 of Oncology*, vol. 27, no. suppl 2, pp. ii97.2–ii97, 2016.
- 386 26. A. I. Phipps, N. M. Lindor, M. A. Jenkins, J. A. Baron, A. K. Win, S. Gallinger,
387 R. Gryfe, and P. A. Newcomb, "Colon and rectal cancer survival by tumor location and
388 microsatellite instability: the Colon Cancer Family Registry," *Diseases of the colon and
389 rectum*, vol. 56, pp. 937–44, aug 2013.
- 390 27. K. Hemminki, I. Santi, M. Weires, H. Thomsen, J. Sundquist, and J. L. Bermejo, "Tumor
391 location and patient characteristics of colon and rectal adenocarcinomas in relation to
392 survival and TNM classes," *BMC Cancer*, vol. 10, p. 688, dec 2010.
- 393 28. R. Cui, Y. Okada, S. G. Jang, J. L. Ku, J. G. Park, Y. Kamatani, N. Hosono, T. Tsun-
394 oda, V. Kumar, C. Tanikawa, N. Kamatani, R. Yamada, M. Kubo, Y. Nakamura, and
395 K. Matsuda, "Common variant in 6q26-q27 is associated with distal colon cancer in an
396 asian population," *Gut*, vol. 60, pp. 799–805, June 2011.
- 397 29. L. Zhu, M. Du, D. Gu, L. Ma, H. Chu, N. Tong, J. Chen, Z. Zhang, and M. Wang,
398 "Genetic variant rs7758229 in 6q26-q27 is not associated with colorectal cancer risk in
399 a chinese population," *PLoS One*, vol. 8, p. e59256, Mar. 2013.
- 400 30. C.-M. Hsu, P.-M. Lin, J.-G. Chang, H.-C. Lin, S.-H. Li, S.-F. Lin, and M.-Y. Yang,
401 "Upregulated SLC22A3 has a potential for improving survival of patients with head
402 and neck squamous cell carcinoma receiving cisplatin treatment," *Oncotarget*, vol. 8,
403 pp. 74348–74358, Sept. 2017.
- 404 31. C. Grisanzio, L. Werner, D. Takeda, B. C. Awoyemi, M. M. Pomerantz, H. Yamada,
405 P. Sooriakumaran, B. D. Robinson, R. Leung, A. C. Schinzel, I. Mills, H. Ross-Adams,
406 D. E. Neal, M. Kido, T. Yamamoto, G. Petrozziello, E. C. Stack, R. Lis, P. W. Kantoff,
407 M. Loda, O. Sartor, S. Egawa, A. K. Tewari, W. C. Hahn, and M. L. Freedman, "Genetic
408 and functional analyses implicate the NUDT11, HNF1B, and SLC22A3 genes in prostate
409 cancer pathogenesis," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, pp. 11252–11257, July
410 2012.
- 411 32. Q. Li and Y. Shu, "Role of solute carriers in response to anticancer drugs," *Mol Cell
412 Ther.* vol. 2, p. 15, May 2014.
- 413 33. S. Yokoo, S. Masuda, A. Yonezawa, T. Terada, T. Katsura, and K.-I. Inui, "Signifi-
414 cance of organic cation transporter 3 (SLC22A3) expression for the cytotoxic effect of
415 oxaliplatin in colorectal cancer," *Drug Metab. Dispos.*, vol. 36, pp. 2299–2306, Nov.
416 2008.
- 417 34. N. Whiffin, F. J. Hosking, S. M. Farrington, C. Palles, S. E. Dobbins, L. Zgaga, A. Lloyd,
418 B. Kinnarsley, M. Gorman, A. Tenesa, P. Broderick, Y. Wang, E. Barclay, C. Hayward,
419 L. Martin, D. D. Buchanan, A. K. Win, J. Hopper, M. Jenkins, N. M. Lindor, P. A.
420 Newcomb, S. Gallinger, D. Conti, F. Schumacher, G. Casey, T. Liu, Swedish Low-
421 Risk Colorectal Cancer Study Group, H. Campbell, A. Lindblom, R. S. Houlston, I. P.
422 Tomlinson, and M. G. Dunlop, "Identification of susceptibility loci for colorectal cancer
423 in a genome-wide meta-analysis," *Hum. Mol. Genet.*, vol. 23, pp. 4729–4737, Sept. 2014.
- 424 35. C. Tanikawa, Y. Kamatani, A. Takahashi, Y. Momozawa, K. Leveque, S. Nagayama,
425 K. Mimori, M. Mori, H. Ishii, J. Inazawa, J. Yasuda, A. Tsuboi, A. Shimizu, M. Sasaki,
426 T. Yamaji, N. Sawada, M. Iwasaki, S. Tsugane, M. Naito, K. Wakai, T. Koyama,
427 T. Takezaki, K. Yuji, Y. Murakami, Y. Nakamura, M. Kubo, and K. Matsuda, "GWAS
428 identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12," *Carcinogenesis*,
429 vol. 39, pp. 652–660, May 2018.
- 430 36. S. L. Schmit, C. K. Edlund, F. R. Schumacher, J. Gong, T. A. Harrison, J. R. Huyghe,
431 C. Qu, M. Melas, D. J. Van Den Berg, H. Wang, S. Tring, S. J. Plummer, D. Albanes,
432 M. H. Alonso, C. I. Amos, K. Anton, A. K. Aragaki, V. Arndt, E. L. Barry, S. I.
433 Berndt, S. Bezieau, S. Bien, A. Bloomer, J. Boehm, M.-C. Boutron-Ruault, H. Brenner,
434 S. Brezina, D. D. Buchanan, K. Butterbach, B. J. Caan, P. T. Campbell, C. S. Carlson,
435 J. E. Castelao, A. T. Chan, J. Chang-Claude, S. J. Chanock, I. Cheng, Y.-W. Cheng,

- 436 L. S. Chin, J. M. Church, T. Church, G. A. Coetzee, M. Cotterchio, M. Cruz Correa,
437 K. R. Curtis, D. Duggan, D. F. Easton, D. English, E. J. M. Feskens, R. Fischer, L. M.
438 FitzGerald, B. K. Fortini, L. G. Fritsche, C. S. Fuchs, M. Gago-Dominguez, M. Gala,
439 S. J. Gallinger, W. J. Gauderman, G. G. Giles, E. L. Giovannucci, S. M. Gogarten,
440 C. Gonzalez-Villalpando, E. M. Gonzalez-Villalpando, W. M. Grady, J. K. Greenson,
441 A. Gsur, M. Gunter, C. A. Haiman, J. Hampe, S. Harlid, J. F. Harju, R. B. Hayes,
442 P. Hofer, M. Hoffmeister, J. L. Hopper, S.-C. Huang, J. M. Huerta, T. J. Hudson, D. J.
443 Hunter, G. E. Idos, M. Iwasaki, R. D. Jackson, E. J. Jacobs, S. H. Jee, M. A. Jenkins,
444 W.-H. Jia, S. Jiao, A. D. Joshi, L. N. Kolonel, S. Kono, C. Kooperberg, V. Krogh,
445 T. Kuehn, S. Küry, A. LaCroix, C. A. Laurie, F. Lejbkovicz, M. Lemire, H.-J. Lenz,
446 D. Levine, C. I. Li, L. Li, W. Lieb, Y. Lin, N. M. Lindor, Y.-R. Liu, F. Loupakis, Y. Lu,
447 F. Luh, J. Ma, C. Mancao, F. J. Manion, S. D. Markowitz, V. Martin, K. Matsuda,
448 K. Matsuo, K. J. McDonnell, C. E. McNeil, R. Milne, A. J. Molina, B. Mukherjee,
449 N. Murphy, P. A. Newcomb, K. Offit, H. Omichessan, D. Palli, J. P. P. Cotoré, J. Pérez-
450 Mayoral, P. D. Pharoah, J. D. Potter, C. Qu, L. Raskin, G. Rennert, H. S. Rennert,
451 B. M. Riggs, C. Schafmayer, R. E. Schoen, T. A. Sellers, D. Seminara, G. Severi, W. Shi,
452 D. Shibata, X.-O. Shu, E. M. Siegel, M. L. Slattery, M. Southey, Z. K. Stadler, M. C.
453 Stern, S. Stintzing, D. Taverna, S. N. Thibodeau, D. C. Thomas, A. Trichopoulou,
454 S. Tsugane, C. M. Ulrich, F. J. B. van Duijnhoven, B. van Guelpan, J. Vijai, J. Virtamo,
455 S. J. Weinstein, E. White, A. K. Win, A. Wolk, M. Woods, A. H. Wu, K. Wu, Y.-B.
456 Xiang, Y. Yen, B. W. Zanke, Y.-X. Zeng, B. Zhang, N. Zubair, S.-S. Kweon, J. C.
457 Figueiredo, W. Zheng, L. L. Marchand, A. Lindblom, V. Moreno, U. Peters, G. Casey,
458 L. Hsu, D. V. Conti, and S. B. Gruber, "Novel common genetic susceptibility loci for
459 colorectal cancer," *J. Natl. Cancer Inst.*, June 2018.
- 460 37. F. R. Schumacher, S. L. Schmit, S. Jiao, C. K. Edlund, H. Wang, B. Zhang, L. Hsu,
461 S.-C. Huang, C. P. Fischer, J. F. Harju, G. E. Idos, F. Lejbkovicz, F. J. Manion,
462 K. McDonnell, C. E. McNeil, M. Melas, H. S. Rennert, W. Shi, D. C. Thomas, D. J.
463 Van Den Berg, C. M. Hutter, A. K. Aragaki, K. Butterbach, B. J. Caan, C. S. Carlson,
464 S. J. Chanock, K. R. Curtis, C. S. Fuchs, M. Gala, E. L. Giovannucci, S. M. Gogarten,
465 R. B. Hayes, B. Henderson, D. J. Hunter, R. D. Jackson, L. N. Kolonel, C. Kooperberg,
466 S. Küry, A. LaCroix, C. C. Laurie, C. A. Laurie, M. Lemire, D. Levine, J. Ma, K. W.
467 Makar, C. Qu, D. Taverna, C. M. Ulrich, K. Wu, S. Kono, D. W. West, S. I. Berndt,
468 S. Bezieau, H. Brenner, P. T. Campbell, A. T. Chan, J. Chang-Claude, G. A. Coetzee,
469 D. V. Conti, D. Duggan, J. C. Figueiredo, B. K. Fortini, S. J. Gallinger, W. J. Gauder-
470 man, G. Giles, R. Green, R. Haile, T. A. Harrison, M. Hoffmeister, J. L. Hopper, T. J.
471 Hudson, E. Jacobs, M. Iwasaki, S. H. Jee, M. Jenkins, W.-H. Jia, A. Joshi, L. Li, N. M.
472 Lindor, K. Matsuo, V. Moreno, B. Mukherjee, P. A. Newcomb, J. D. Potter, L. Raskin,
473 G. Rennert, S. Rosse, G. Severi, R. E. Schoen, D. Seminara, X.-O. Shu, M. L. Slat-
474 tery, S. Tsugane, E. White, Y.-B. Xiang, B. W. Zanke, W. Zheng, L. Le Marchand,
475 G. Casey, S. B. Gruber, and U. Peters, "Genome-wide association study of colorectal
476 cancer identifies six new susceptibility loci," *Nat. Commun.*, vol. 6, p. 7138, July 2015.
- 477 38. J. B. Sneddon, H. H. Zhen, K. Montgomery, M. van de Rijn, A. D. Tward, R. West,
478 H. Gladstone, H. Y. Chang, G. S. Morganroth, A. E. Oro, and P. O. Brown, "Bone
479 morphogenetic protein antagonist gremlin 1 is widely expressed by cancer-associated
480 stromal cells and can promote tumor cell proliferation," *Proc. Natl. Acad. Sci. U. S.*
481 *A.*, vol. 103, pp. 14842–14847, Oct. 2006.
- 482 39. H. Stabile, S. Mitola, E. Moroni, M. Belleri, S. Nicoli, D. Coltrini, F. Peri, A. Pessi,
483 L. Orsatti, F. Talamo, V. Castronovo, D. Waltregny, F. Cotelli, D. Ribatti, and
484 M. Presta, "Bone morphogenic protein antagonist drg/gremlin is a novel proangiogenic
485 factor," *Blood*, vol. 109, pp. 1834–1840, Mar. 2007.
- 486 40. J. Ziai, E. Matloff, J. Choi, N. Kombo, M. Materin, and A. E. Bale, "Defining the poly-
487 posis/colorectal cancer phenotype associated with the ashkenazi GREM1 duplication:
488 counselling and management recommendations," *Genet. Res.*, vol. 98, p. e5, Mar. 2016.
- 489 41. H. Davis, S. Irshad, M. Bansal, H. Rafferty, T. Boitsova, C. Bardella, E. Jaeger, A. Lewis,
490 L. Freeman-Mills, F. C. Giner, P. Rodenas-Cuadrado, S. Mallappa, S. Clark, H. Thomas,
491 R. Jeffery, R. Poulson, M. Rodriguez-Justo, M. Novelli, R. Chetty, A. Silver, O. J. San-
492 som, F. R. Greten, L. M. Wang, J. E. East, I. Tomlinson, and S. J. Leedham, "Aberrant

- 493 epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem
494 cell niche," *Nat. Med.*, vol. 21, pp. 62–70, Jan. 2015.
- 495 42. F. O. Desmet, D. Hamroun, M. Lalande, G. Collod-Bèroud, M. Claustres, and
496 C. Bèroud, "Human Splicing Finder: An online bioinformatics tool to predict splicing
497 signals," *Nucleic Acids Research*, vol. 37, no. 9, 2009.
- 498 43. Y. Yang, P. Junjie, C. Sanjun, and Y. Ma, "Long non-coding RNAs in Colorectal Cancer
499 : Progression and Future Directions," *Journal of Cancer*, vol. 8, 2017.
- 500 44. W. Guo, Q. Wang, Y. Zhan, X. Chen, Q. Yu, J. Zhang, Y. Wang, X. J. Xu, and
501 L. Zhu, "Transcriptome sequencing uncovers a three-long noncoding RNA signature in
502 predicting breast cancer survival," *Scientific Reports*, vol. 6, 2016.
- 503 45. E. S. Wan, W. Qiu, V. J. Carey, J. Morrow, H. Bacherman, M. G. Foreman, J. E.
504 Hokanson, R. P. Bowler, J. D. Crapo, and D. L. DeMeo, "Smoking-associated site-
505 specific differential methylation in buccal mucosa in the COPD Gene study," *American
506 Journal of Respiratory Cell and Molecular Biology*, vol. 53, pp. 246–254, aug 2015.
- 507 46. T.-H. Liu, F. Zheng, M.-Y. Cai, L. Guo, H.-X. Lin, J.-W. Chen, Y.-J. Liao, H.-F. Kung,
508 Y.-X. Zeng, and D. Xie, "The putative tumor activator ARHGEF3 promotes nasopharyngeal
509 carcinoma cell pathogenesis by inhibiting cellular apoptosis," *Oncotarget*, vol. 7,
510 pp. 25836–25848, May 2016.
- 511 47. Y. Shi, G.-B. Chen, X.-X. Huang, C.-X. Xiao, H.-H. Wang, Y.-S. Li, J.-F. Zhang, S. Li,
512 Y. Xia, J.-L. Ren, and B. Guleng, "Dragon (repulsive guidance molecule b, RGMb) is a
513 novel gene that promotes colorectal cancer growth," *Oncotarget*, vol. 6, pp. 20540–20554,
514 Aug. 2015.
- 515 48. I. Suryapranata and R. Kusuma unpublished, N.D.
- 516 49. U. Peters, S. Jiao, F. R. Schumacher, C. M. Hutter, A. K. Aragaki, J. A. Baron, S. I.
517 Berndt, S. Bézieau, H. Brenner, K. Butterbach, B. J. Caan, P. T. Campbell, C. S.
518 Carlson, G. Casey, A. T. Chan, J. Chang-Claude, S. J. Chanock, L. S. Chen, G. A.
519 Coetzee, S. G. Coetzee, D. V. Conti, K. R. Curtis, D. Duggan, T. Edwards, C. S.
520 Fuchs, S. Gallinger, E. L. Giovannucci, S. M. Gogarten, S. B. Gruber, R. W. Haile,
521 T. A. Harrison, R. B. Hayes, B. E. Henderson, M. Hoffmeister, J. L. Hopper, T. J.
522 Hudson, D. J. Hunter, R. D. Jackson, S. H. Jee, M. A. Jenkins, W. H. Jia, L. N. Kolonel,
523 C. Kooperberg, S. Küry, A. Z. Lacroix, C. C. Laurie, C. A. Laurie, L. Le Marchand,
524 M. Lemire, D. Levine, N. M. Lindor, Y. Liu, J. Ma, K. W. Makar, K. Matsuo, P. A.
525 Newcomb, J. D. Potter, R. L. Prentice, C. Qu, T. Rohan, S. A. Rosse, R. E. Schoen,
526 D. Seminara, M. Shrubsole, X. O. Shu, M. L. Slaterry, D. Taverna, S. N. Thibodeau,
527 C. M. Ulrich, E. White, Y. Xiang, B. W. Zanke, Y. X. Zeng, B. Zhang, W. Zheng, and
528 L. Hsu, "Identification of genetic susceptibility loci for colorectal tumors in a genome-
529 wide meta-analysis," *Gastroenterology*, vol. 144, pp. 799–807.e24, apr 2013.
- 530 50. N. Whiffin, F. J. Hosking, S. M. Farrington, C. Palles, S. E. Dobbins, L. Zgaga, A. Lloyd,
531 B. Kinnerley, M. Gorman, A. Tenesa, P. Broderick, Y. Wang, E. Barclay, C. Hay-
532 ward, L. Martin, D. D. Buchanan, A. K. Win, J. Hopper, M. Jenkins, N. M. Lindor,
533 P. A. Newcomb, S. Gallinger, D. Conti, F. Schumacher, G. Casey, T. Liu, H. Campbell,
534 A. Lindblom, R. S. Houlston, I. P. Tomlinson, and M. G. Dunlop, "Identification of sus-
535 ceptibility loci for colorectal cancer in a genome-wide meta-analysis," *Human Molecular
536 Genetics*, vol. 23, pp. 4729–4737, sep 2014.
- 537 51. R. S. Houlston, J. Cheadle, S. E. Dobbins, A. Tenesa, A. M. Jones, K. Howarth, S. L.
538 Spain, P. Broderick, E. Domingo, S. Farrington, J. G. Prendergast, A. M. Pittman,
539 E. Theodoratou, C. G. Smith, B. Olver, A. Walther, R. A. Barnetson, M. Churchman,
540 E. E. Jaeger, S. Penegar, E. Barclay, L. Martin, M. Gorman, R. Mager, E. Johnstone,
541 R. Midgley, I. Niittymäki, S. Tuupanen, J. Colley, S. Idziaszyk, H. J. Thomas, A. M.
542 Lucassen, D. G. R. Evans, E. R. Maher, T. Maughan, A. Dimas, E. Dermitzakis, J. B.
543 Cazier, L. A. Aaltonen, P. Pharoah, D. J. Kerr, L. G. Carvajal-Carmona, H. Campbell,
544 M. G. Dunlop, and I. P. Tomlinson, "Meta-analysis of three genome-wide association
545 studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and
546 20q13.33," *Nature Genetics*, vol. 42, pp. 973–977, nov 2010.
- 547 52. F. R. Schumacher, S. L. Schmit, S. Jiao, C. K. Edlund, H. Wang, B. Zhang, L. Hsu,
548 S. C. Huang, C. P. Fischer, J. F. Harju, G. E. Idos, F. Lejbkiewicz, F. J. Manion,
549 K. McDonnell, C. E. McNeil, M. Melas, H. S. Rennert, W. Shi, D. C. Thomas, D. J.
550 Van Den Berg, C. M. Hutter, A. K. Aragaki, K. Butterbach, B. J. Caan, C. S. Carlson,

- 551 S. J. Chanock, K. R. Curtis, C. S. Fuchs, M. Gala, E. L. Giocannucci, S. M. Gogarten,
552 R. B. Hayes, B. Henderson, D. J. Hunter, R. D. Jackson, L. N. Kolonel, C. Kooperberg,
553 S. Kury, A. Lacroix, C. C. Laurie, C. A. Laurie, M. Lemire, D. Levine, J. Ma, K. W.
554 Makar, C. Qu, D. Taverna, C. M. Ulrich, K. Wu, S. Kono, D. W. West, S. I. Berndt,
555 S. Bezieau, H. Brenner, P. T. Campbell, A. T. Chan, J. Chang-Claude, G. A. Coetzee,
556 D. V. Conti, D. Duggan, J. C. Figueiredo, B. K. Fortini, S. J. Gallinger, W. J. Gauder-
557 man, G. Giles, R. Green, R. Haile, T. A. Harrison, M. Hoffmeister, J. L. Hopper, T. J.
558 Hudson, E. Jacobs, M. Iwasaki, S. H. Jee, M. Jenkins, W. H. Jia, A. Joshi, L. Li, N. M.
559 Lindor, K. Matsuo, V. Moreno, B. Mukherjee, P. A. Newcomb, J. D. Potter, L. Raskin,
560 G. Rennert, S. Rosse, G. Severi, R. E. Schoen, D. Seminara, X. O. Shu, M. L. Slat-
561 tery, S. Tsugane, E. White, Y. B. Xiang, B. W. Zanke, W. Zheng, L. Le Marchand,
562 G. Casey, S. B. Gruber, and U. Peters, "Genome-wide association study of colorectal
563 cancer identifies six new susceptibility loci," *Nature Communications*, vol. 6, p. 7138,
564 dec 2015.
- 565 53. L. M. Real, A. Ruiz, J. Gayán, A. González-Pérez, M. E. Sáez, R. Ramírez-Lorca, F. J.
566 Morón, J. Velasco, R. Marginet-Flinch, E. Musulén, J. M. Carrasco, C. Moreno-Rey,
567 E. Vázquez, M. Chaves-Conde, J. A. Moreno-Nogueira, M. Hidalgo-Pascual, E. Ferrero-
568 Herrero, S. Castellví-Bel, A. Castells, C. Fernandez-Rozadilla, C. Ruiz-Ponte, A. Car-
569 racedo, B. González, S. Alonso, and M. Perucho, "A colorectal cancer susceptibility
570 new variant at 4q26 in the Spanish population identified by genome-wide association
571 analysis," *PLoS ONE*, vol. 9, p. e101178, jun 2014.
- 572 54. M. G. Dunlop, S. E. Dobbins, S. M. Farrington, A. M. Jones, C. Palles, N. Whiffin,
573 A. Tenesa, S. Spain, P. Broderick, L. Y. Ooi, E. Domingo, C. Smillie, M. Henrion,
574 M. Frampton, L. Martin, G. Grimes, M. Gorman, C. Semple, Y. P. Ma, E. Barclay,
575 J. Prendergast, J. B. Cazier, B. Olver, S. Penegar, S. Lubbe, I. Chander, L. G. Carvajal-
576 Carmona, S. Ballereau, A. Lloyd, J. Vijayakrishnan, L. Zgaga, I. Rudan, E. Theodora-
577 tou, H. Thomas, E. Maher, G. Evans, L. Walker, D. Halliday, A. Lucassen, J. Paterson,
578 S. Hodgson, T. Homfray, L. Side, L. Izatt, A. Donaldson, S. Tomkins, P. Morrison,
579 C. Brewer, A. Henderson, R. Davidson, V. Murday, J. Cook, N. Haites, T. Bishop,
580 E. Sheridan, A. Green, C. Marks, S. Carpenter, M. Broughton, L. Greenhalge, M. Suri,
581 J. M. Starr, I. Deary, I. Kirac, D. Kovacevia, L. A. Aaltonen, L. Renkonen-Sinisalo,
582 J. P. Mecklin, K. Matsuda, Y. Nakamura, Y. Okada, S. Gallinger, D. J. Duggan,
583 D. Conti, P. Newcomb, J. Hopper, M. A. Jenkins, F. Schumacher, G. Casey, D. Easton,
584 M. Shah, P. Pharoah, A. Lindblom, T. Liu, D. Edler, C. Lenander, J. Dalén, F. Hjern,
585 N. Lundqvist, U. Lindfors, L. Pahlman, K. Smedh, A. Törnqvist, J. Holm, M. Janson,
586 M. Andersson, S. Ekelund, L. Olsson, C. G. Smith, H. West, J. P. Cheadle, G. MacDon-
587 ald, L. M. Samuel, A. Ahmad, P. Corrie, D. Jodrell, C. Palmer, C. Wilson, J. O'Hagan,
588 D. Smith, R. McDermott, J. Walshe, J. Cassidy, A. McDonald, N. Mohammed, J. White,
589 H. Yosef, O. Breathnach, L. Grogan, R. Thomas, M. Eatock, P. Henry, R. Houston,
590 P. Johnston, R. Wilson, I. Geh, F. Danwata, A. Hindley, S. Susnerwala, C. Bradley,
591 A. Conn, A. Raine, C. Twelves, S. Falk, K. Hopkins, S. Tahir, A. Dhadha, A. Mar-
592 aveyas, J. Sgouros, M. Teo, R. Ahmad, S. Cleator, A. Creak, C. Lowdell, P. Riddle,
593 K. Benstead, D. Farrugia, N. Reed, S. Shepherd, E. Levine, S. Mullamitha, M. Saun-
594 ders, J. Valle, G. Wilson, A. Jones, A. Weaver, P. I. Clark, B. Haylock, M. I. Iqbal,
595 A. S. Myint, S. Beesley, T. Seivitt, J. Nicoll, F. Daniel, V. Ford, T. Talbot, M. Butt,
596 A. Hamid, P. MacK, R. Roy, R. Osborne, F. McKinna, H. Alsab, D. Basu, P. Murray,
597 B. Sizer, F. A. Azam, R. Neupane, A. Waterston, J. Glaholm, C. Blesing, S. Lowndes,
598 A. Medisetti, A. Gaya, M. Leslie, N. Maisey, P. Ross, G. Dunn, O. Al-Salih, H. S. Wasan,
599 L. T. Tan, J. Dent, U. Hofmann, J. K. Joffe, E. Sherwin, R. Soomal, A. Chakrabarti,
600 S. Joseph, J. Van Der Voet, N. J. Wadd, D. Wilson, S. Anjarwalia, J. Hall, R. Hughes,
601 A. Polychronis, J. H. Scarffe, M. Hill, R. D. James, R. Shah, J. Summers, A. Hartley,
602 D. Carney, J. McCaffrey, B. Bystricky, S. O'Reilly, R. Gupta, T. Al-Mishlab, F. Gidden,
603 R. O'Hara, J. Stewart, R. Ashford, R. Glynn-Jones, M. Harrison, S. Mawdsley, H. Bar-
604 low, M. Tighe, J. Walthers, J. Neal, C. Rees, J. Bridgewater, S. Karp, U. McGovern,
605 P. J. Atherton, H. El-Deeb, C. MacMillan, K. Patel, E. M. Bessell, P. D. Dickinson,
606 V. Potter, C. Jephcott, K. McAdam, J. Wrigley, S. Muthuramalingam, A. O'Callaghan,
607 L. Melcher, C. Braconi, J. I. Geh, D. Palmer, P. Narayana, N. Steven, A. Gaya, S. Rud-
608 man, P. Chakrabarti, K. Kelly, C. MacGregor, D. Whillis, A. Freebairn, J. Gilder-

- 609 sleve, S. Sharif, G. Astras, T. Hickish, D. Beech, R. Ellis, R. Kulkarni, K. Shankland,
610 R. Begent, A. Mayer, T. Meyer, S. Strauss, V. Hall, S. Raj, I. Chau, D. Cunningham,
611 A. Birtle, A. Biswas, M. Wise, S. Cummins, S. Essapen, G. Middleton, C. Topham,
612 R. Langley, A. Webb, M. Wilkins, T. J. Iveson, C. Askill, J. Wagstaff, A. Azzabi,
613 A. Bateman, J. Prejbisz, D. Tsang, N. Ali, A. Jones, P. O'Neill, C. Cottrill, D. Propper,
614 F. J. Lofts, J. Kennedy, D. A. Anthony, R. Cooper, A. Crellin, A. Melcher, M. Seymour,
615 C. Baughan, E. Alexander, J. Crown, D. Fennelly, F. Adab, S. Girdharan, I. Pedley,
616 K. Wright, P. Bliss, G. Cogill, N. Lo, E. Toy, D. Hochhauser, J. Ledermann, A. Brew-
617 ster, T. Maughan, D. Mort, S. Mukherjee, W. Dobrowsky, P. Calvert, G. Leonard,
618 H. Ford, A. M. Moody, S. Goriah, M. Wilkins, S. Clive, L. Dawson, C. McLean, H. A.
619 Phillips, K. Gopi, M. Tomlinson, S. Clenton, D. Furniss, J. Hornbuckle, S. Pledge,
620 J. Wadsley, M. Abbas, E. Marshall, C. Harper-Wynne, A. Barnes, S. Kumar, V. Vi-
621 gneswaran, S. Gollins, M. Genton, G. Sparrow, C. Bale, C. Fuller, A. Mullard, N. Stuart,
622 R. Williams, M. Keane, T. Maughen, R. Adams, A. Madi, E. Hodgkinson, P. Rogers,
623 M. Pope, R. Kaplan, A. Meade, M. Parmar, S. Kenny, D. Fisher, L. Harper, J. Mitchell,
624 L. Nichols, B. Sydes, L. Clement, E. Kay, C. Courtney, M. Gallagher, C. Murphy,
625 L. Thompson, S. Beall, S. Hassan, R. Gracie, G. Griffiths, M. Mason, C. Parker,
626 R. Rudd, P. Johnson, J. Whelan, J. Northover, J. Brown, M. Aapro, R. Stout, R. Midg-
627 ley, D. J. Kerr, H. A. Campbell, I. P. Tomlinson, and R. S. Houlston, "Common variation
628 near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk," *Nature Ge-
629 netics*, vol. 44, pp. 770–776, jul 2012.
- 630 55. R. Cui, Y. Okada, S. G. Jang, J. L. Ku, J. G. Park, Y. Kamatani, N. Hosono, T. Tsun-
631 oda, V. Kumar, C. Tanikawa, N. Kamatani, R. Yamada, M. Kubo, Y. Nakamura, and
632 K. Matsuda, "Common variant in 6q26-q27 is associated with distal colon cancer in an
633 Asian population," *Gut*, vol. 60, pp. 799–805, jun 2011.
- 634 56. B. W. Zanke, C. M. Greenwood, J. Rangrej, R. Kustra, A. Tenesa, S. M. Farring-
635 ton, J. Prendergast, S. Olschwan, T. Chiang, E. Crowdy, V. Ferretti, P. Laflamme,
636 S. Sundararajan, S. Roumy, J. F. Olivier, F. Robidoux, R. Sladek, A. Montpetit,
637 P. Campbell, S. Bezieau, A. M. O'Shea, G. Zogopoulos, M. Cotterchio, P. Newcomb,
638 J. McLaughlin, B. Younghusband, R. Green, J. Green, M. E. Porteous, H. Campbell,
639 H. Blanche, M. Sahbatou, E. Tubacher, C. Bonaiti-Pellié, B. Buecher, E. Riboli, S. Kury,
640 S. J. Chanock, J. Potter, G. Thomas, S. Gallinger, T. J. Hudson, and M. G. Dunlop,
641 "Genome-wide association scan identifies a colorectal cancer susceptibility locus on chro-
642 mosome 8q24," *Nature Genetics*, vol. 39, pp. 989–994, aug 2007.
- 643 57. S. B. Gruber, V. Moreno, L. S. Rozek, H. S. Rennerts, F. Lejbkowitz, J. D. Bonner,
644 J. K. Greenon, T. J. Giordano, E. R. Fearson, and G. Rennert, "Genetic variation
645 in 8q24 associated with risk of colorectal cancer," *Cancer biology & therapy*, vol. 6,
646 pp. 1143–7, jul 2007.
- 647 58. C. A. Haiman, L. Le Marchand, J. Yamamoto, D. O. Stram, X. Sheng, L. N. Kolonel,
648 A. H. Wu, D. Reich, and B. E. Henderson, "A common genetic risk factor for colorectal
649 and prostate cancer," *Nature Genetics*, vol. 39, pp. 954–956, aug 2007.
- 650 59. I. P. Tomlinson, E. Webb, L. Carvajal-Carmona, P. Broderick, K. Howarth, A. M.
651 Pittman, S. Spain, S. Lubbe, A. Walther, K. Sullivan, E. Jaeger, S. Fielding, A. Rowan,
652 J. Vijaykrishnan, E. Domingo, I. Chandler, Z. Kemp, M. Qureshi, S. M. Farrington,
653 A. Tenesa, J. G. Prendergast, R. A. Barnetson, S. Penegar, E. Barclay, W. Wood,
654 L. Martin, M. Gorman, H. Thomas, J. Peto, D. T. Bishop, R. Gray, E. R. Maher,
655 A. Lucassen, D. Kerr, D. G. R. Evans, C. Schafmayer, S. Buch, H. Völzke, J. Hampe,
656 S. Schreiber, U. John, T. Koessler, P. Pharoah, T. Van Wezel, H. Morreau, J. T. Wijnen,
657 J. L. Hopper, M. C. Southey, G. G. Giles, G. Severi, S. Castellví-Bel, C. Ruiz-Ponte,
658 A. Carracedo, A. Castells, A. Försti, K. Hemminki, P. Vodicka, A. Naccarati, L. Lipton,
659 J. W. Ho, K. K. Cheng, P. C. Sham, J. Luk, J. A. Agúndez, J. M. Ladero, M. De La
660 Hoya, T. Caldés, I. Niittymäki, S. Tuupanen, A. Karhu, L. Aaltonen, J. B. Cazier,
661 H. Campbell, M. G. Dunlop, and R. S. Houlston, "A genome-wide association study
662 identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3," *Nature
663 Genetics*, vol. 40, pp. 623–630, may 2008.
- 664 60. C. M. Hutter, M. L. Slatery, D. J. Duggan, J. Muehling, K. Curtin, L. Hsu, S. A.
665 Beresford, A. Rajkovic, G. E. Sarto, J. R. Marshall, N. Hammad, R. Wallace, K. W.
666 Makar, R. L. Prentice, B. J. Caan, J. D. Potter, and U. Peters, "Characterization of the

- 667 association between 8q24 and colon cancer: Gene-environment exploration and meta-
668 analysis," *BMC Cancer*, vol. 10, p. 670, dec 2010.
- 669 61. A. Tenesa, S. M. Farrington, J. G. Prendergast, M. E. Porteous, M. Walker, N. Haq,
670 R. A. Barnetson, E. Theodoratou, R. Cetnarskyj, N. Cartwright, C. Semple, A. J.
671 Clark, F. J. Reid, L. A. Smith, K. Kavoussanakis, T. Koessler, P. D. Pharoah, S. Buch,
672 C. Schafmayer, J. Tepel, S. Schreiber, H. Völzke, C. O. Schmidt, J. Hampe, J. Chang-
673 Claude, M. Hoffmeister, H. Brenner, S. Wilkening, F. Canzian, G. Capella, V. Moreno,
674 I. J. Deary, J. M. Starr, I. P. Tomlinson, Z. Kemp, K. Howarth, L. Carvajal-Carmona,
675 E. Webb, P. Broderick, J. Vijayakrishnan, R. S. Houlston, G. Rennert, D. Ballinger,
676 L. Rozek, S. B. Gruber, K. Matsuda, T. Kidokoro, Y. Nakamura, B. W. Zanke, C. M.
677 Greenwood, J. Rangrej, R. Kustra, A. Montpetit, T. J. Hudson, S. Gallinger, H. Camp-
678 bell, and M. G. Dunlop, "Genome-wide association scan identifies a colorectal cancer
679 susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21," *Nature Ge-*
680 *netics*, vol. 40, pp. 631–637, may 2008.
- 681 62. H. Wang, C. A. Haiman, T. Burnett, B. K. Fortini, L. N. Kolonel, B. E. Henderson,
682 L. B. Signorello, W. J. Blot, T. O. Keku, S. I. Berndt, P. A. Newcomb, M. Pande, C. I.
683 Amos, D. W. West, G. Casey, R. S. Sandler, R. Haile, D. O. Stram, and L. Le Marchand,
684 "Fine-mapping of genome-wide association study-identified risk loci for colorectal cancer
685 in African Americans," *Human Molecular Genetics*, vol. 22, pp. 5048–5055, dec 2013.
- 686 63. E. Jaeger, E. Webb, K. Howarth, L. Carvajal-Carmona, A. Rowan, P. Broderick,
687 A. Walther, S. Spain, A. Pittman, Z. Kemp, K. Sullivan, K. Heinemann, S. Lubbe,
688 E. Domingo, E. Barclay, L. Martin, M. Gorman, I. Chandler, J. Vijayakrishnan,
689 W. Wood, E. Papaemmanuil, S. Penegar, M. Qureshi, S. Farrington, A. Tenesa, J. B.
690 Cazier, D. Kerr, R. Gray, J. Peto, M. Dunlop, H. Campbell, H. Thomas, R. Houlston,
691 and I. Tomlinson, "Common genetic variants at the CRAC1 (HMPS) locus on chromo-
692 some 15q13.3 influence colorectal cancer risk," *Nature Genetics*, vol. 40, pp. 26–28, jan
693 2008.
- 694 64. P. Broderick, L. Carvajal-Carmona, A. M. Pittman, E. Webb, K. Howarth, A. Rowan,
695 S. Lubbe, S. Spain, K. Sullivan, S. Fielding, E. Jaeger, J. Vijayakrishnan, Z. Kemp,
696 M. Gorman, I. Chandler, E. Papaemmanuil, S. Penegar, W. Wood, G. Sellick,
697 M. Qureshi, A. Teixeira, E. Domingo, E. Barclay, L. Martin, O. Sieber, D. Kerr, R. Gray,
698 J. Peto, J. B. Cazier, I. Tomlinson, and R. S. Houlston, "A genome-wide association
699 study shows that common alleles of SMAD7 influence colorectal cancer risk," *Nature*
700 *Genetics*, vol. 39, pp. 1315–1317, nov 2007.

Genetic risk...

ORIGINALITY REPORT

% **18**
SIMILARITY INDEX

% **1**
INTERNET SOURCES

% **1**
PUBLICATIONS

% **17**
STUDENT PAPERS

PRIMARY SOURCES

1 Submitted to University of Florida
Student Paper % **17**

2 tesisenxarxa.net
Internet Source <% **1**

3 www.hssnet.co.jp
Internet Source <% **1**

4 hal.archives-ouvertes.fr
Internet Source <% **1**

5 www.nature.com
Internet Source <% **1**

6 Zhang, Ben, Wei-Hua Jia, Keitaro Matsuo, Aesun Shin, Yong-Bing Xiang, Koichi Matsuda, Sun Ha Jee, Dong-Hyun Kim, Peh Yean Cheah, Zefang Ren, Qiuyin Cai, Jirong Long, Jiajun Shi, Wanqing Wen, Gong Yang, Bu-Tian Ji, Zhi-Zhong Pan, Fumihiko Matsuda, Yu-Tang Gao, Jae Hwan Oh, Yoon-Ok Ahn, Michiaki Kubo, Lai Fun Thean, Eun Jung Park, Hong-Lan Li, Ji Won Park, Jaeseong Jo, Jin-Young Jeong, <% **1**

Satoyo Hosono, Yusuke Nakamura, Xiao-Ou Shu, Yi-Xin Zeng, and Wei Zheng. "Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians : GWAS Identifies a New SMAD7 Risk Variants for CRC", *International Journal of Cancer*, 2014.

Publication

7

www.pharexmedics.com

Internet Source

<% 1

8

Du, Mengmeng, Shuo Jiao, Stephanie A. Bien, Manish Gala, Goncalo Abecasis, Stephane Bezieau, Hermann Brenner, Katja Butterbach, Bette J. Caan, Christopher S. Carlson, Graham Casey, Jenny Chang-Claude, David V. Conti, Keith R. Curtis, David Duggan, Steven Gallinger, Robert W. Haile, Tabitha A. Harrison, Richard B. Hayes, Michael Hoffmeister, John L. Hopper, Thomas J. Hudson, Mark A. Jenkins, Sébastien Küry, Loic Le Marchand, Suzanne M. Leal, Polly A. Newcomb, Deborah A. Nickerson, John D. Potter, Robert E. Schoen, Fredrick R. Schumacher, Daniela Seminara, Martha L. Slattery, Li Hsu, Andrew T. Chan, Emily White, Sonja I. Berndt, and Ulrike Peters. "Fine-Mapping of Common Genetic Variants Associated with Colorectal Tumor Risk Identified Potential Functional Variants", *PLoS ONE*,

<% 1

2016.

Publication

9

radboudimaging.nl

Internet Source

<% 1

10

Yang, Baiyu, Aaron P. Thrift, Jane C. Figueiredo, Mark A. Jenkins, Fredrick R. Schumacher, David V. Conti, Yi Lin, Aung Ko Win, Paul J. Limburg, Sonja I. Berndt, Hermann Brenner, Andrew T. Chan, Jenny Chang-Claude, Michael Hoffmeister, Thomas J. Hudson, Loïc Le Marchand, Polly A. Newcomb, Martha L. Slattery, Emily White, Ulrike Peters, Graham Casey, and Peter T. Campbell. "Common variants in the obesity-associated genes FTO and MC4R are not associated with risk of colorectal cancer", *Cancer Epidemiology*, 2016.

Publication

<% 1

11

Yusuke Takahashi, Keishi Sugimachi, Ken Yamamoto, Atsushi Niida et al. "Japanese genome-wide association study identifies a significant colorectal cancer susceptibility locus at chromosome 10p14", *Cancer Science*, 2017

Publication

<% 1

12

Jianbo Tian, Jiang Chang, Jing Gong, Jiao Lou et al. "Systematic Functional Interrogation of Genes in GWAS Loci Identified ATF1 as a Key

<% 1

Driver in Colorectal Cancer Modulated by a Promoter-Enhancer Interaction", The American Journal of Human Genetics, 2019

Publication

13

rd.springer.com

Internet Source

<% 1

14

GongXin Yu. "An Exon-Based Comparative Variant Analysis Pipeline to Study the Scale and Role of Frameshift and Nonsense Mutation in the Human-Chimpanzee Divergence", Comparative and Functional Genomics, 2009

Publication

<% 1

15

Li Hsu, Jihyoun Jeon, Hermann Brenner, Stephen B. Gruber et al. "A Model to Determine Colorectal Cancer Risk Using Common Genetic Susceptibility Loci", Gastroenterology, 2015

Publication

<% 1

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES

< 5 WORDS