ARTICLE

Novel Common Genetic Susceptibility Loci for Colorectal Cancer

Stephanie L. Schmit*, Christopher K. Edlund*, Fredrick R. Schumacher*, Jian Gong*, Tabitha A. Harrison, Jeroen R. Huyghe, Chenxu Qu, Marilena Melas, David J. Van Den Berg, Hansong Wang, Stephanie Tring, Sarah J. Plummer, Demetrius Albanes, M. Henar Alonso, Christopher I. Amos, Kristen Anton, Aaron K. Aragaki, Volker Arndt, Elizabeth L. Barry, Sonja I. Berndt, Stéphane Bezieau, Stephanie Bien, Amanda Bloomer, Juergen Boehm, Marie-Christine Boutron-Ruault, Hermann Brenner, Stefanie Brezina, Daniel D. Buchanan, Katja Butterbach, Bette J. Caan, Peter T. Campbell, Christopher S. Carlson, Jose E. Castelao, Andrew T. Chan, Jenny Chang-Claude, Stephen J. Chanock, Iona Cheng, Ya-Wen Cheng, Lee Soo Chin, James M. Church, Timothy Church, Gerhard A. Coetzee, Michelle Cotterchio, Marcia Cruz Correa, Keith R. Curtis, David Duggan, Ulrike Peters[†], Graham Casey[†], Li Hsu[†], David V. Conti[†], Stephen B. Gruber[†]

See the Notes section for the full list of authors' and affiliations.
Correspondence to: Stephen B. Gruber, MD, PhD, MPH, 1441 Eastlake Avenue, Norman Topping Tower (NTT) Room 4436, Los Angeles, CA 90033 (e-mail: sgruber@usc.edu).
*Equal contribution
†Jointly supervised this research

Abstract

Background: Previous genome-wide association studies (GWAS) have identified 42 loci ($P < 5 \times 10^{-8}$) associated with risk of colorectal cancer (CRC). Expanded consortium efforts facilitating the discovery of additional susceptibility loci may capture unexplained familial risk.
Methods: We conducted a GWAS in European descent CRC cases and control subjects using a discovery–replication design, followed by examination of novel findings in a multiethnic sample (cumulative n = 163 315). In the discovery stage (36 948 case subjects/30 864 control subjects), we identified genetic variants with a minor allele frequency of 1% or greater associated with risk of CRC using logistic regression followed by a fixed-effects inverse variance weighted meta-analysis. All novel independent variants reaching genome-wide statistical significance (two-sided $P < 5 \times 10^{-8}$) were tested for replication in separate European ancestry samples (12 952 case subjects/48 383 control subjects). Next, we examined the generalizability of discovered variants in East Asians, African Americans, and Hispanics (12 085 case subjects/22 083 control subjects). Finally, we examined the contributions of novel risk variants to familial relative risk and examined the prediction capabilities of a polygenic risk score. All statistical tests were two-sided.
Results: The discovery GWAS identified 11 variants associated with CRC at $P < 5 \times 10^{-8}$, of which nine (at 4q22.2/5p15.33/5p13.1/6p21.31/6p12.1/10q11.23/12q24.21/16q24.1/20q13.13) independently replicated at a P value of less than .05. Multiethnic follow-up supported the generalizability of discovery findings. These results demonstrated a 14.7% increase in familial relative risk explained by common risk alleles from 10.3% (95% confidence interval [CI] = 7.9% to 13.7%; known variants) to 11.9%

(95% CI = 9.2% to 15.5%; known and novel variants). A polygenic risk score identified 4.3% of the population at an odds ratio for developing CRC of at least 2.0.

Conclusions: This study provides insight into the architecture of common genetic variation contributing to CRC etiology and improves risk prediction for individualized screening.

Colorectal cancer (CRC) is a complex polygenetic disease, and heritability accounts for up to 35% of the variation in risk of developing CRC (1,2). Some of this heritability is attributable to rare high-penetrance alleles associated with cancer syndromes, now routinely incorporated into clinical care. In addition, genome-wide association studies (GWAS) have identified variation in numerous regulatory regions and other genomic loci that contribute quantifiable risks for CRC development. Specifically, GWAS have identified approximately 70 common genetic variants across 42 regions ($P < 5\times10^{-8}$) associated with risk of CRC, as larger study populations have been amassed and racial/ethnic representation has increased (3–11). Expanded consortium efforts facilitating the discovery of additional risk loci may capture unexplained familial risk.

Our prior collaborative work identified six novel CRC susceptibility loci based on a discovery sample of 18 299 case subjects and 19 656 control subjects of European ancestral heritage (12). Results from this GWAS contributed to the development of the Illumina Infinium OncoArray-500K BeadChip (OncoArray; San Diego, CA), a genotyping array designed to interrogate genomic variation associated with predisposition to five of the most common cancers (prostate, breast, colorectal, lung, and ovarian) (13). Here, we describe results from a new discovery-replication GWAS, including for the first time findings from the OncoArray Project. Then, we present a follow-up evaluation of genome-wide statistically significant ($P < 5\times10^{-8}$) risk alleles in individuals from diverse ethnic groups (East Asian, Hispanic, and African American) to investigate if the findings generalize to other populations. Our goal was to discover and replicate new CRC susceptibility loci by assembling the largest international study population to date (n = 163 315).

Methods

Study Overview

This investigation included genetic data from 53 observational studies and clinical trials (Supplementary Figure 1, Supplementary Table 1, available online). In the discovery stage, we combined genotype and epidemiologic data from individuals with European ancestry from all of our consortium efforts to date (CORECT, CCFR, and GECCO), including the new OncoArray Project (36 948 case subjects and 30 864 control subjects) (Supplementary Table 2, Supplementary Figures 2 and 3, available online). In the replication stage, we leveraged data from an independent set of European descent participants (12 952 case subjects and 48 383 control subjects) (Supplementary Table 3, available online). In the follow-up stage to assess generalizability of findings, we examined data from a multiethnic sample set (12 085 case subjects and 22 083 control subjects) that included East Asians from the OncoArray Project (Supplementary Table 4, Supplementary Figure 4, available online) and prior studies (14,15), African Americans (15,16), and Hispanics/Latinos (17). Details of the study populations, genotyping, quality control (QC), and imputation for each stage of this GWAS are described in the Supplementary Methods (available online). Participants

provided written informed consent, and the Institutional Review Boards at each center approved the study. For more specific information on consent and study approvals at each institution, see the Supplementary Methods (available online).

Statistical Analysis

Detailed descriptions of the statistical analysis for each study stage are described in the Supplementary Methods (available online). Briefly, we examined the association between allelic dosage for all autosomal variants with a minor allele frequency (MAF) of 0.01 or greater that passed stringent imputation quality control procedures and CRC status using logistic regression adjusted for appropriate study-specific covariates and principal components (PCs) that capture global ancestry. Summary statistics from European descent samples included in our prior consortium efforts (Discovery Part 1) (18) and the OncoArray Project (Discovery Part 2) were combined in a fixed-effect inverse variance–weighted meta-analysis. Consistency of odds ratios (ORs) across studies was assessed using Cochran's Q test of heterogeneity. The most statistically significantly associated variant in each novel genome-wide statistically significant locus (two-sided $P < 5\times10^{-8}$) from this discovery analysis was then examined for association with risk of CRC in the independent replication stage of European ancestry participants (Supplementary Methods, available online). Criteria for independent replication included a consistent direction of association and a P value of less than .05 based on a meta-analysis of study-specific logistic regression models. Finally, all variants reaching genome-wide statistical significance ($P < 5\times10^{-8}$) in the discovery stage and a P value of less than .05 in the replication stage were assessed for generalizability in the multiethnic follow-up stage of East Asians, African Americans, and Hispanics. All statistical tests were two-sided.

Polygenic Risk Scores and Familial Relative Risk Explained

Polygenic risk scores (PRS) in European descent replication phase participants were calculated using previously known susceptibility variants and novel independently replicated variants identified by this effort. PRS were categorized into percentile categories based on a weighted sum of risk allele counts among control subjects (<1%, 1%–10%, 10%–25%, 25%–75%, 75%–90%, 90%–99%, and >99%, with 25%–75% serving as the reference). Weights were applied based on bias-corrected logORs from our European descent discovery analysis. Logistic regression was used to examine CRC risk across PRS categories (after adjusting for age, sex, PCs, and PC*study) for known and known+novel variants, respectively. We also stratified the PRS at a clinically actionable threshold of an odds ratio of 2.0 or greater. To consider the applicability of our European-derived PRS to East Asian populations, we also examined the performance of this score in the East Asian case subjects and control subjects genotyped on the OncoArray. Next, the contributions to familial risk of the known+novel and the known-only variants were investigated.

ARTICLE

Sample inclusions and methods for bias correction, PRS, and family relative risk explained analyses are described in more detail in the Supplementary Methods (available online).

In Silico Functional Follow-up

We conducted eQTL analysis in colonic mucosa from healthy control subjects (n = 50) and normal mucosa adjacent to colon cancer (n = 100) in the Colonomics study (19) as well as transverse colon tissues (n = 169) from the Genotype-Tissue Expression (GTEx) project (Supplementary Methods, available online) (20). Briefly, in Colonomics, for each variant, Pearson partial correlation adjusted for tissue type (healthy or adjacent to tumor) was used to explore the association of single nucleotide polymorphism (SNP)/indel dosage data with gene expression for genes located within 2MB of the SNP of interest. For GTEx, the laboratory and analytic methods have previously been described in detail (20).

Additionally, candidate functional variants were identified using published methods (21). Briefly, index variants and SNPs (CEU, 1KGP, June 2014 release) in LD with each risk variant (we report $r^2 \geq .6$ except where noted as $r^2 \geq .2$) were aligned with chromatin immunoprecipitation and sequencing (ChIP-seq) tracks for histone methylation and acetylation marks associated with enhancers H3K4me1 and H3K27ac. For this study, we referenced Sigmoid Colon H3K27 acetylation from the Roadmap Epigenomics Consortium (22) as well as CRC cell lines SW480 and HCT-116 H3K4 monomethylation generated in our laboratory (G. Casey) and by the ENCODE project, respectively (23,24). To further characterize the novel CRC genetic risk loci, we performed in silico bioinformatic functional annotation of each region.

Results

Discovery GWAS (European Descent)

The discovery GWAS identified 11 common risk variants at 4q22.2, 5q15.33, 5p13.1, 6p21.31, 6p12.1, 10q11.23, 12q24.21, 13q13.2, 16q24.1, 20q11.22, and 20q13.13, all of which were independent of known risk loci (>500 kb away or $r^2 > .2$ with a previously known variant) and reached the accepted genome-wide statistical significance threshold ($P < 5 \times 10^{-8}$) (Table 1). Association results from the discovery stage also indicated that 62 (92.5%) of the 67 known autosomal risk variants (three out of 70 known risk variants were excluded due to MAF < 0.01, low-quality imputation, or location on chromosome X) replicated at a nominal level of statistical significance ($P < .05$) (Supplementary Table 5, available online). A quantile–quantile plot illustrates appropriate control for population stratification with a λ of 1.05 (sample size–adjusted $\lambda_{1000} = 1.002$) (Supplementary Figure 5, available online). A Manhattan plot illustrates the genomic location of novel loci in relation to previously published risk regions (Figure 1). Regional association plots in Supplementary Figure 6 depict the 11 risk variants in the context of their surrounding linkage disequilibrium (LD) structures and nearby genes. The MAFs of these 11 variants in 1KGP Europeans ranged from 0.097 to 0.495, and the odds ratios for association ranged from 0.90 to 1.08 (Table 1). Effect sizes adjusted for potential bias in estimation due to the winner's curse are summarized in Supplementary Table 6 and Supplementary Figure 7 (available online).

Table 1. Eleven novel low-penetrance risk variants identified from the discovery GWAS (European descent) with $P < 5 \times 10^{-8}$ and their results in an independent European replication set

Locus	EFF/REF allele	rsID:CHR:BP	FRQ_EFF (1KGP EUR)	Discovery (n_case = 36 948, n_control = 30 864)				Replication (n_case = 12 952, n_control = 48 383)			
				OR (95% CI)	P*	I^2, %	$P_{heterogeneity}$†	OR (95% CI)	P*	I^2, %	$P_{heterogeneity}$†
4q22.2	T/C	rs1370821:4:94943383	0.401	1.07 (1.04 to 1.09)	4.0×10^{-8}	0	.58	1.05 (1.02 to 1.08)	.003	42.1	.14
5p15.33	A/G	rs2735940:5:1296486	0.511	0.92 (0.89 to 0.94)	3.1×10^{-13}	0	.59	0.93 (0.90 to 0.96)	3.0×10^{-6}	0	.75
5p13.1	G/GT	rs58791712:5:40281797‡	0.745	0.91 (0.89 to 0.93)	7.3×10^{-14}	56.7	.13	0.90 (0.87 to 0.93)	1.1×10^{-9}	28.4	.24
6p21.31	T/C	rs6906359:6:35528378‡	0.097	0.90 (0.86 to 0.93)	3.4×10^{-8}	0	.65	0.93 (0.89 to 0.98)	.005	0	.55
6p12.1	T/C	rs62404968:6:55714314	0.248	0.92 (0.89 to 0.94)	8.6×10^{-10}	0	.32	0.94 (0.90 to 0.97)	3.8×10^{-4}	0	.96
10q11.23	T/C	rs10994860:10:52645424	0.202	0.92 (0.89 to 0.95)	3.5×10^{-8}	0	.35	0.96 (0.92 to 1.00)	.04	0	.43
12q24.21	CACAA/C	rs72013726:12:115890835‡	0.505	0.93 (0.90 to 0.95)	5.0×10^{-11}	0	.84	0.95 (0.92 to 0.98)	9.1×10^{-4}	0	.83
13q13.2	C/G	rs10161980:13:34093518	0.620	1.08 (1.05 to 1.10)	4.7×10^{-9}	0	.81	1.03 (0.99 to 1.06)	.13	21.6	.28
16q24.1	C/G	rs2696839:16:86340448	0.495	0.94 (0.92 to 0.96)	2.0×10^{-8}	75.6	.04	0.96 (0.93 to 0.99)	.009	25.5	.25
20q11.22	T/C	rs2295444:20:33173883	0.495	0.93 (0.91 to 0.95)	3.3×10^{-9}	0	.97	0.97 (0.94 to 1.00)	.08	0	.59
20q13.13	T/C	rs1810502:20:49057488	0.449	0.93 (0.91 to 0.96)	1.02×10^{-8}	0	.98	0.94 (0.91 to 0.97)	5.9×10^{-5}	11.8	.34

*P values were derived from a fixed-effects inverse variance-weighted meta-analysis. All tests were two-sided. 1KGP EUR = 1000 Genomes Europeans; BP = position; CHR = chromosome; CI = confidence interval; EFF = effect allele; FRQ = frequency; OR = odds ratio; REF = reference allele (reference category for the odds ratios).

†P values were derived from Cochran's Q test of heterogeneity. All tests were two-sided.

‡Proxies were used in the independent replication stage (r^2 values from 1KGP Phase 3 Release 5): rs12520534 (chr5:40281761), $r^2 = 1.0$; rs144037597 (chr6: 35528204), $r^2 = 1.0$; rs12822984 (chr12:115888504), $r^2 = 0.81$.
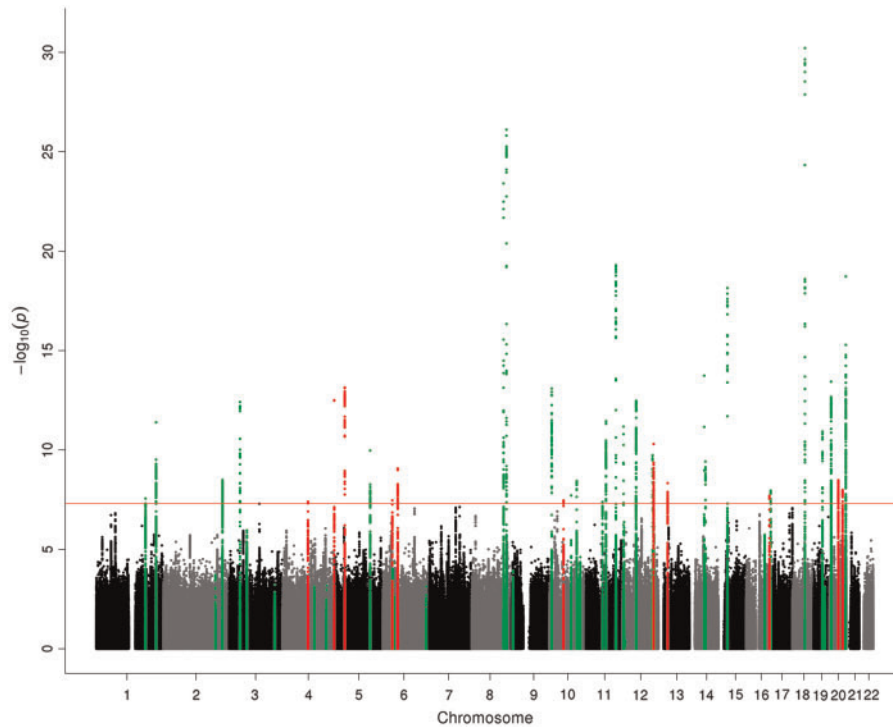
Figure 1. Manhattan plot summarizing the discovery genome-wide association study association results (n_{case} = 36 948, $n_{control}$ = 30 864). **Green** = known risk loci (within 500 kb or $r^2 > .2$ with an index variant); **red** = novel risk loci (outside 500 kb or $r^2 > .2$ with an index variant).

Replication (European Descent)

The association between each of the 11 candidate susceptibility variants identified in the discovery stage and risk of CRC in an independent sample revealed consistent directions of association and consistent effect sizes for all variants (Table 1). Also, odds ratios for association were statistically significant for nine of 11 variants. The remaining two loci that were identified in the discovery stage (rs10161980 and rs2295444) demonstrated supportive but not statistically significant evidence of replication, and thus require further validation in future studies. Notably, the two variants with statistical evidence of heterogeneity in the discovery stage meta-analysis replicated in this independent sample set (rs58791712 and rs2696839).

Multiethnic Follow-up

Subsequently, we examined the nine novel, replicated risk variants across three diverse ethnic populations. We examined the association between each variant and risk of CRC in East Asians (n = 21 630) (Supplementary Figure 4, available online), African Americans (n = 6597), and Hispanics (n = 5941). All nine variants demonstrated a consistent direction of association in follow-up studies, except for rs62404968 and rs10994860 in Hispanics (Table 2). Eight out of the nine variants (all but rs10994860) were associated with risk of CRC in at least one population at a nominal level of statistical significance (P < .05).

Polygenic Risk Score Analysis and Familial Relative Risk Explained

PRS analysis conducted in a subset of European descent replication phase participants revealed that the estimated odds of

developing CRC for individuals with scores in the top 1% as compared with the 25%–75% reference category was 2.18 (Supplementary Table 7, available online). Based on the 76 known and novel variants, 4.3% of the study population could be identified for targeted screening based on a clinically actionable threshold of an odds ratio of 2.0 or greater (Supplementary Table 7, available online) (25,26). This is in comparison with 1.4% of the study population that is identifiable based on previously known variants only (data not shown). The known+novel PRS performed similarly in East Asians, and the cutpoint to reach a clinically actionable odds ratio of at least 2.0 in this population was 99.1% (Supplementary Table 7, available online).

Overall, 76 variants explained 11.9% (95% confidence interval [CI] = 9.2% to 15.5%) of the known familial relative risk, as compared with 10.3% (95% CI = 7.9% to 13.7%) for the previously known variants only. This represents a 14.7% increase in familial relative risk explained. Estimation of the proportion of explained familial risk incorporated uncertainty in risk estimation for each variant and uncertainty in the specification of the familial relative risk.

eQTL Analysis

Analysis of *cis* gene expression data for the nine novel susceptibility variants revealed several noteworthy eQTLs in Colonomics and GTEx transverse colon samples (Supplementary Table 8, available online). For example, rs10994860 is a statistically significant eQTL for *ASAH2* (effect size = –0.61, P = 5.7E×10^{-5}). Further, in the Colonomics data set, rs6906359 is a statistically significant eQTL for several genes including *BRPF3*, showing overexpression for C/C as compared with T/T genotypes (partial r^2 = .09, P = 2.6×10^{-4}). The most statistically significant eQTLs in each region with at least one variant associated at the P < .05 level in the Colonomics

Table 2. Multiethnic follow-up of nine novel, independently replicated low-penetrance risk variants

Locus	EFF/REF allele	rsID:CHR:BP	FRQ_EFF (1KGP EAS)	FRQ_EFF (1KGP AMR)	FRQ_EFF (1KGP AFR)	East Asians (OncoArray, ACCC, US–Japan GWAS)				Hispanic/Latinos (HCCS, MEC, SIGMA)		African Americans (AA, CRC, GWAS)	
						OR (95% CI)	P*	I^2, %	$P_{heterogeneity}$†	OR (95% CI)	P*	OR (95% CI)	P*
4q22.2	T/C	rs1370821:4:94943383	0.331	0.274	0.065	1.03 (0.99 to 1.08)	.13	49.6	.14	1.17 (1.06 to 1.29)	.001	1.04 (0.92 to 1.17)	.54
5p15.33	A/G	rs2735940:5:1296486	0.478	0.432	0.521	0.93 (0.87 to 1.00)	.03	61.2	.11	0.99 (0.91 to 1.08)	.84	0.90 (0.83 to 0.98)	.01
5p13.1	G/GT	rs5791712:5:40281797	0.956	0.765	0.924	0.87 (0.75 to 1.02)	.09	0	.57	0.85 (0.77 to 0.94)	.001	NA	NA
6p21.31	T/C	rs6906359:6:35528378	0.069	0.138	0.141	0.99 (0.91 to 1.07)	.73	0	.45	0.82 (0.73 to 0.93)	.001	0.96 (0.84 to 1.08)	.47
6p12.1	T/C	rs62404968:6:55714314	0.061	0.133	0.072	0.97 (0.88 to 1.05)	.44	60	.08	1.03 (0.90 to 1.17)	.69	0.85 (0.74 to 0.97)	.02
10q11.23	T/C	rs10994860:10:52645424	0.047	0.110	0.222	0.97 (0.89 to 1.06)	.47	51.2	.13	1.00 (0.87 to 1.16)	.97	0.99 (0.90 to 1.09)	.87
12q24.21	CACAA/C	rs72013726:12:115890835	0.643	0.633	0.360	0.92 (0.87 to 0.98)	.007	53.9	.14	0.97 (0.89 to 1.06)	.53	NA	NA
16q24.1	C/G	rs2696839:16:86340448	0.253	0.334	0.293	0.93 (0.89 to 0.98)	.004	45	.18	0.90 (0.82 to 0.98)	.02	0.92 (0.84 to 1.00)	.06
20q13.13	T/C	rs1810502:20:49057488	0.612	0.507	0.545	0.94 (0.90 to 0.98)	.007	49.2	.16	0.92 (0.84 to 1.00)	.05	0.95 (0.88 to 1.03)	.24

*P values were derived from a fixed-effects inverse variance–weighted meta-analysis. All tests were two-sided. 1KGP = 1000 Genomes; AA = African American; ACCC = Asia Colorectal Cancer Consortium; AFR = African; AMR = Ad Mixed American; BP = position; CI = confidence interval; CRC = colorectal cancer; EFF = effect allele; CHR = chromosome; EAS = East Asian; FRQ = frequency; GWAS = genome-wide association study; HCCS = Hispanic Colorectal Cancer Study; MEC = Multiethnic Cohort; OR = odds ratio; REF = reference allele (reference category for the odds ratios); SIGMA = Slim Initiative in Genomic Medicine for the Americas.
†P values were derived from Cochran's Q test of heterogeneity. All tests were two-sided.

ARTICLE

data set are summarized in Supplementary Figure 8 (available online).

Discussion

This collaborative study included over 163 000 individuals for the identification and further evaluation of 9 replicable novel CRC genetic susceptibility loci. Nine low-penetrance risk loci represent approximately a 21% increase from those previously discovered to date (n = 42). Nine risk variants replicated in an independent sample of European ancestry participants, and eight of those generalized to at least one of three other racial/ethnic populations. Our findings contribute substantially to the known familial relative risk explained by low-penetrance susceptibility alleles, with a 14.7% increase from 10.3% (previously known only) to 11.9% (known + novel reported here) explained. Further, PRS analysis underscores the impact of common CRC risk alleles, particularly among individuals with the highest counts of risk variants. Our findings suggest that 4.3% of the population could be targeted for earlier and more frequent screening based on germline genetic profiling of all known common CRC susceptibility variants. This supports our previous findings that GWAS have the potential to inform appropriate tailoring of screening guidelines to population subgroups (27).

The consistent direction of association for all nine novel risk variants in East Asians and African Americans (all but two in Hispanics) underscores the generalizability of our findings from European ancestry individuals. However, the statistically significant association of some but not all variants with CRC risk across the additional ethnic subgroups supports the importance of expanded sample sizes in certain populations as well as ongoing multiethnic fine-mapping studies to identify the strongest signals and most likely putative functional variant(s) at particular loci in other ancestral populations.

Two of the nine risk alleles map to intragenic or coding regions. First, rs62404968 maps to 6p12.1 and lies within an intron of BMP5. BMP5 encodes bone morphogenetic protein 5, which is part of the transforming growth factor–beta (TGF-β) superfamily. Members of the BMP and TGF-β family have been implicated as risk genes for CRC in previous GWAS, including BMP2 and BMP4 on chromosomes 20 and 14, respectively (28). The associated SNP, rs62404968, or any of the 20 SNPs in LD, do not map to any predicted regulatory/enhancer regions based on histone marks, suggesting that further functional follow-up is needed to understand the functional mechanism likely acting on the strong candidate gene BMP5. Second, rs10994860 maps to 10q11.23 and lies within exon 1 of A1CF, representing a putative candidate functional SNP. APOBEC1 complementation factor (A1CF) is a critical component of the apolipoprotein B mRNA editing enzyme complex. There are two SNPs (rs71457593 and rs10994720) in LD with rs10994860 that both map to histone peaks also suggesting potential functionality.

The remaining seven risk alleles map to intergenic regions of the genome. SNP rs1370821 maps to 4q22.2, with the two nearest genes being ATOH1 and SMARCAD1 (approximately 85 kb away). ATOH1 encodes atonal homolog BHLH transcription factor 1, which belongs to the basic helix-loop-helix family of transcription factors. SMARCAD1 encodes matrix-associated actin-dependent regulator of chromatin, a member of the SNF subfamily of helicase proteins that plays an important role in heterochromatin reorganization following DNA replication. Although the associated SNP, rs1370821, does not map to any

candidate regulatory regions, two SNPs (rs2510787, rs2433324) in LD with rs1370821 lie within an intron of the gene encoding PDZ and LIM domain protein 5 (*PDLIM5*), and both map to histone marks. Also, rs1370821 warrants further functional characterization because of its proximity to *BMPR1B*, a gene where there is statistical evidence of an eQTL relationship by genotype in the Colonomics data set and where the gene family is related to polyposis and CRC susceptibility (17).

The indel rs58791712 (G/GT) maps to 5p13.1. The nearest genes, *PTGER4* and *LINC00603*, lie approximately 400 kb from the index variant. *PTGER4* encodes PGE2 receptor EP4 subtype and is one of four receptors identified for prostaglandin E2. This indel does not map to any histone marks, making it unlikely to be a functional variant. However, there are three SNPs (rs72748452, rs755989, and rs4957261) in LD with rs58791712 that overlap histone peaks.

The SNP rs2735940 maps to 5p15.33 and lies adjacent to the *TERT* gene. *TERT* encodes the telomerase catalytic subunit protein that helps to maintain telomere ends by addition of the telomere repeat TTAGGG. *TERT* has been identified previously as a candidate risk gene in several cancers including CRC (29–34). The SNP rs2735940 does not map to any histone marks. However, this SNP is in LD with three SNPs (rs380145, rs246995, and rs246994) that map to histone marks and lie within an intron of *CLPTM1L* (rs380145) or the predicted gene *BC034612* (rs246995 and rs246994).

The SNP rs6906359 maps to 6p21.31, and the closest gene is *FKBP5* approximately 12 kb away. *FKBP5* encodes FK506 binding protein 5, a member of the immunophilin protein family that plays a role in immunoregulation, protein folding, and trafficking. However, rs6906359 does not overlap any histone marks. Of the SNPs in LD with rs6906359 that overlap histone peaks, two SNPs (rs72894781 and rs72894784) map within an intron of *TEAD3*, one SNP (rs16878812) maps within an intron of *FKBP5*, and one SNP is intergenic (rs45493300).

The indel rs72013726 (CACAA/C) maps to 12q24.21. The nearest gene, *MED13L*, lies approximately 500 kb from rs72013726. *MED13L* encodes thyroid hormone receptor–associated protein 2 and is one of many proteins that function as a transcriptional coactivator for RNA polymerase II–transcribed genes. SNP rs72013726 maps to a histone peak, making it a potential functional SNP.

The SNP rs2696839 maps to 16q24.1 and lies 15 kb from the predicted gene *LOC146513*. Although this SNP does not map to any histone marks, all four SNPs (rs12932862, rs12149163, rs12149501, and rs2665316) in LD with rs2696839 do. Of note, there are several lncRNAs in this region.

The SNP rs1810502 maps to 20q13.13 near the gene *PTPN1*, approximately 70 kb away. *PTPN1* encodes protein-tyrosine phosphatase 1B, a member of the protein tyrosine phosphatase family. This SNP and 14 other SNPs in LD with rs1810502 map to histone marks, implying the possibility that any one of these 15 SNPs could be functionally relevant to CRC etiology.

Our study design has strengths and limitations. We conducted a rigorous two-stage study with discovery and independent replication in European descent participants. Further, a major strength is that we utilized data from the independent replication phase to conduct PRS and familial relative risk explained analyses. Of note, despite a 14.7% increase beyond prior knowledge, still less than 12% of familial relative risk is explained by GWAS-identified alleles, including our nine new loci. Thus, additional efforts are needed to fully explain the genetic architecture of this complex disease, potentially with gene–environment interactions. Space limitations preclude

detailed descriptions of eQTL analyses for each SNP. However, we found little or no evidence of the nine novel index SNPs in relation to gene expression for our speculatively implicated genes. Additional eQTL analyses in expanded normal colon tissue sample sets that examine the full landscape of SNPs in LD with the index SNP may help to elucidate the impact of germline susceptibility loci on gene expression. Future studies will be advantageous to identify rare and intermediate frequency susceptibility alleles through expanded sample size as well as increased racial/ethnic minority inclusion. Multiethnic samples will be useful for fine-mapping known and novel risk regions as well as for identifying population-specific variation. In summary, this GWAS provides insight into the etiologies of CRC and provides a basis for future fine-mapping, functional characterization, and risk modeling research.

Funding

Notes

Authors: Stephanie L. Schmit*, Christopher K. Edlund*, Fredrick R. Schumacher*, Jian Gong*, Tabitha A. Harrison, Jeroen R. Huyghe, Chenxu Qu, Marilena Melas, David J. Van Den Berg, Hansong Wang, Stephanie Tring, Sarah J. Plummer, Demetrius Albanes, M. Henar Alonso, Christopher I. Amos, Kristen Anton, Aaron K. Aragaki, Volker Arndt, Elizabeth L. Barry, Sonja I. Berndt, Stéphane Bezieau, Stephanie Bien, Amanda Bloomer, Juergen Boehm, Marie-Christine Boutron-Ruault, Hermann Brenner, Stefanie Brezina, Daniel D. Buchanan, Katja Butterbach, Bette J. Caan, Peter T. Campbell, Christopher S. Carlson, Jose E. Castelao, Andrew T. Chan, Jenny Chang-Claude, Stephen J. Chanock, Iona Cheng, Ya-Wen Cheng, Lee Soo Chin, James M. Church, Timothy Church, Gerhard A. Coetzee,

Michelle Cotterchio, Marcia Cruz Correa, Keith R. Curtis, David Duggan, Douglas F. Easton, Dallas English, Edith J. M. Feskens, Rocky Fischer, Liesel M. FitzGerald, Barbara K. Fortini, Lars G. Fritsche, Charles S. Fuchs, Manuela Gago-Dominguez, Manish Gala, Steven J. Gallinger, W. James Gauderman, Graham G. Giles, Edward L. Giovannucci, Stephanie M. Gogarten, Clicerio Gonzalez-Villalpando, Elena M. Gonzalez-Villalpando, William M. Grady, Joel K. Greenson, Andrea Gsur, Marc Gunter, Christopher A. Haiman, Jochen Hampe, Sophia Harlid, John F. Harju, Richard B. Hayes, Philipp Hofer, Michael Hoffmeister, John L. Hopper, Shu-Chen Huang, Jose Maria Huerta, Thomas J. Hudson, David J. Hunter, Gregory E. Idos, Motoki Iwasaki, Rebecca D. Jackson, Eric J. Jacobs, Sun Ha Jee, Mark A. Jenkins, Wei-Hua Jia, Shuo Jiao, Amit D. Joshi, Laurence N. Kolonel, Suminori Kono, Charles Kooperberg, Vittorio Krogh, Tilman Kuehn, Sébastien Küry, Andrea LaCroix, Cecelia A. Laurie, Flavio Lejbkowicz, Mathieu Lemire, Heinz-Josef Lenz, David Levine, Christopher I. Li, Li Li, Wolfgang Lieb, Yi Lin, Noralane M. Lindor, Yun-Ru Liu, Fotios Loupakis, Yingchang Lu, Frank Luh, Jing Ma, Christoph Mancao, Frank J. Manion, Sanford D. Markowitz, Vicente Martin, Koichi Matsuda, Keitaro Matsuo, Kevin J. McDonnell, Caroline E. McNeil, Roger Milne, Antonio J. Molina, Bhramar Mukherjee, Neil Murphy, Polly A. Newcomb, Kenneth Offit, Hanane Omichessan, Domenico Palli, Jesus P. Paredes Cotoré, Julyann Pérez-Mayoral, Paul D. Pharoah, John D. Potter, Conghui Qu, Leon Raskin, Gad Rennert, Hedy S. Rennert, Bridget M. Riggs, Clemens Schafmayer, Robert E. Schoen, Thomas A. Sellers, Daniela Seminara, Gianluca Severi, Wei Shi, David Shibata, Xiao-Ou Shu, Erin M. Siegel, Martha L. Slattery, Melissa Southey, Zsofia K. Stadler, Mariana C. Stern, Sebastian Stintzing, Darin Taverna, Stephen N. Thibodeau, Duncan C. Thomas, Antonia Trichopoulou, Shoichiro Tsugane, Cornelia M. Ulrich, Franzel J. B. van Duijnhoven, Bethany van Guelpan, Joseph Vijai, Jarmo Virtamo, Stephanie J. Weinstein, Emily White, Aung Ko Win, Alicja Wolk, Michael Woods, Anna H. Wu, Kana Wu, Yong-Bing Xiang, Yun Yen, Brent W. Zanke, Yi-Xin Zeng, Ben Zhang, Niha Zubair, Sun-Seog Kweon, Jane C. Figueiredo, Wei Zheng, Loic Le Marchand, Annika Lindblom, Victor Moreno, Ulrike Peters†, Graham Casey†, Li Hsu†, David V. Conti†, Stephen B. Gruber†

*Equal contribution

†Jointly supervised this research

Affiliations of authors: Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL (SLS, AB, BMR, TAS, EMS); Department of Preventive Medicine, USC Norris Comprehensive Cancer Center (SLS, CKE, CQ, DJVDB, ST, WJG, CAH, SCH, GEI, HJL, KJM, CEM, MM, MCS, DCT, AHW, DVC, SBG), Department of Medicine (HJL, SBG), and Department of Preventive Medicine (JCF), Keck School of Medicine, University of Southern California, Los Angeles, CA; Department of Epidemiology and Biostatistics (FRS) and Department of Family Medicine and Community Health, Mary Ann Swetland Center for Environmental Health, Case Comprehensive Cancer Center (LL), Case Western Reserve University, Cleveland, OH; Public Health Sciences Division (JG, TAH, JRH, AKA, SB, CSC, KRC, SJ, CK, AL, YL, PAN, JDP, CQ, EW, NZ, UP, LH) and Translational Research Program (CIL), Fred Hutchinson Cancer Research Center, Seattle, WA; Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI (HW, LLM); Center for Public Health Genomics, University of Virginia, Charlottesville, VA (SJP, GC); Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD (DA, SIB, SJC, SJW); Catalan Institute of Oncology, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain (MHA, VMo);

CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain (MHA, VMa, AJM, VMo); University of Barcelona, Barcelona, Spain (MHA, VMo); Department of Biomedical Data Science (CIA, KA) and Department of Epidemiology (ELB), Geisel School of Medicine, Dartmouth College, Hanover, NH; Division of Clinical Epidemiology and Aging Research (VA, HB, KB, MH), German Cancer Consortium (HB), Unit of Genetic Epidemiology, Division of Cancer Epidemiology (JCC), and Division of Cancer Epidemiology (TK), German Cancer Research Center (DKFZ), Heidelberg, Germany; Centre Hospitalier Universitaire Hotel-Dieu, Nantes, France (SB); Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU), Nantes, France (SB, SK); Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, UT (JB, CMU); CESP (U1018 INSERM), Facultés de Médecine Université Paris-Sud, UVSQ, Université Paris-Saclay, Villejuif, France (MCBR, HO, GS); Gustave Roussy, Villejuif, France (MCBR, HO); Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany (HB); Medical University of Vienna, Department of Medicine I, Institute of Cancer Research, Vienna, Austria (SB, AG, PH); Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia (DDB, DE, GGG, MAJ, RM, AKW); Colorectal Oncogenomics Group, Department of Pathology (DDB) and Genetic Epidemiology Laboratory, Department of Pathology (MS), University of Melbourne, Melbourne, Victoria, Australia; Genetic Medicine and Familial Cancer Centre, The Royal Melbourne Hospital, Parkville, Victoria, Australia (DDB); Division of Research, Kaiser Permanente Medical Care Program of Northern California, Oakland, CA (BJC); Epidemiology Research Program, American Cancer Society, Atlanta, GA (PTC, EJJ); Genetic Oncology Unit, Instituto de Investigación Sanitaria Galicia Sur (IISGS), Complejo Hospitalario Universitario de Vigo (CHUVI), SERGAS, Vigo (Pontevedra) Spain (JEC); Division of Gastroenterology (ATC, MG) and Clinical and Translational Epidemiology Unit (ATC, MG, ADJ), Massachusetts General Hospital, Boston, MA; Harvard Medical School, Boston, MA (JCC, ELG); University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (JCC); Cancer Prevention Institute of California, Fremont, CA (IC); Ph.D. Program of Cancer Research and Drug Discovery (YWC, YY), Joint Biobank, Office of Human Research (YRL), and School of Medicine (FL), Taipei Medical University, Taipei, Taiwan; Cancer Science Institute of Singapore, National University of Singapore, Singapore (LSC); Department of Colorectal Surgery, Cleveland Clinic, Cleveland, OH (JMC); Division of Environmental Health Sciences, University of Minnesota, Minneapolis, MN (TC); Van Andel Research Institute, Grand Rapids, MI (GAC); Prevention and Cancer Control, Cancer Care Ontario, Toronto, ON, Canada (MC); Puerto Rico Cancer Center (MCC) and Division of Cancer Biology (JPM), University of Puerto Rico, San Juan, Puerto Rico; Genetic Basis of Human Disease Division, Translational Genomics Research Institute, Phoenix, AZ (DD); Centre for Cancer Genetic Epidemiology (DFE) and Centre for Cancer Genetic Epidemiology (PDP), Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia (DE, LMF, GGG, RM); Division of Human Nutrition, Wageningen University and Research, Wageningen, the Netherlands (EJMF, FJBdV); University of Michigan Comprehensive Cancer Center, Ann Arbor, MI (RF, JFH, FJM, BM); Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia

(LMF); W.M. Keck Science Department, Claremont Colleges, Claremont, CA (BKF); Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI (LGF); K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Sør-Trøndelag, Norway (LGF); Department of Medical Oncology, Dana-Farber Cancer Institute, Brookline, MA (CSF); Department of Medicine (CSF) and Channing Division of Network Medicine (ELG), Brigham and Women's Institute, Brookline, MA; Genomic Medicine Group, Galician Foundation of Genomic Medicine, Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saude (SERGAS), Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago De Compostela, Spain (MGD); Moores Cancer Center, University of California San Diego, La Jolla, CA (MGD); Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital, Toronto, Ontario, Canada (SJG); Department of Biostatistics (SMG, CAL, DL), Department of Medicine, Division of Gastroenterology, School of Medicine (WMG), and Department of Epidemiology, School of Public Health (UP), University of Washington, Seattle, WA; Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Cuernavaca, Morelos, Mexico (CGV); Centro de Estudios en Diabetes AC, Mexico City, Mexico (EMGV); Department of Pathology, University of Michigan, Ann Arbor, MI (JKG); Nutrition and Metabolism Section, IARC, Lyon, CEDEX 08, France (MG, NM); Medical Department 1, University Hospital Dresden, TU Dresden, Dresden, Germany (JH); Department of Radiation Sciences, Oncology, Umea University, Umea, Sweden (SH, BvG); Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, NY (RBH); Centre for MEGA Epidemiology, The University of Melbourne, Carlton, Victoria, Australia (JLH); Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Murcia, Spain (JMH); AbbVie, Redwood City, CA (TJH); Ontario Institute for Cancer Research, Toronto, Ontario, Canada (TJH, ML); Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard School of Public Health, Boston, MA (DJH, ADJ); Division of Epidemiology, Center for Public Health Sciences (MI), National Cancer Center (ST), Tokyo, Japan; Department of Medicine, Ohio State University, Columbus, OH (RDJ); Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, South Korea (SHJ); State Key Laboratory of Oncology in South China, Cancer Center, Sun Yatsen University, Guangzhou, China (WHJ, YXZ); Office of Public Health Studies, University of Hawaii Manoa, Honolulu, HI (LNK); Department of Preventive Medicine, Kyushu University, Fukuoka, Japan (SK); Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy (VK); Clalit Health Services National Israeli Cancer Control Center, Haifa, Israel (FL, GR, HSR); Department of Community Medicine and Epidemiology, Carmel Medical Center, Haifa, Israel (FL, GR, HSR); Institute of Epidemiology, PopGen Biobank, Christian-Albrechts-University Kiel, Kiel, Germany (WL); Department of Health Science Research, Mayo Clinic, Scottsdale, AZ (NML); Unit of Oncology, Department of Clinical and Experimental Oncology, Instituto Oncologico Veneto, IRCCS Padua, Italy (FL); Division of Epidemiology, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN (YL, LR, XOS, WZ); Sino-American Cancer Foundation, Temple City, CA (FL); Harvard School of Public Health, Boston, MA (JM); Genentech, Inc., Basel, Switzerland (CM); Departments of Medicine and Genetics, Case Comprehensive Cancer Center, Case Western Reserve University, and University Hospitals of Cleveland, Cleveland, OH (SDM); Biomedicine Institute (IBIOMED), University of León, León, Spain (VMa); Research Group on Gene-Environment Interactions and Health, University of León, León, Spain (AJM); Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan (KoM); Department of Epidemiology, Nagoya University Graduate School of Medicine, Showa-ku, Nagoya, Japan (KeM); Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Chikusa-Ku Nagoya, Japan (KeM); Clinical Genetics Service (KO), Department of Medicine (ZKS, JVo), Memorial Sloan Kettering Cancer Center, New York, NY; Cancer Risk Factors and Life-Style Epidemiology Unit, Cancer Research and Prevention Institute-ISPO, Florence, Italy (DP); Department of Surgery, Complejo Hospitalario Universitario de Santiago (CHUS), Servicio Galego de Saúde (SERGAS), Santiago De Compostela, Spain (JPPC); Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN (LR, XOS, WZ); Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel (GR); Department of Visceral and Thoracic Surgery, University Hospital Schleswig-Holstein, Kiel Campus, Kiel, Germany (CS); Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA (RES); Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD (DS); Human Genetics Foundation (HuGeF), Torino, Italy (GS); Department of Surgery, Children's Hospital Los Angeles, Los Angeles, CA (WS); Department of Surgery, University of Tennessee Health Science Center, Memphis, TN (DS); Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT (MLS); Department of Medicine, Weill Cornell Medical Center, New York, NY (ZKS); Department of Hematology and Oncology University of Munich (LMU), Munich, Germany (SS); Phoenix College, Phoenix, AZ (DT); Mayo Clinic, Rochester, MN (SNT); Hellenic Health Foundation, Athens, Greece (AT); Center for Public Health Sciences, National Cancer Center, Tokyo, Japan (ST);Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland (JVa); Institute of Environmental Medicine (AWW) and Department of Molecular Medicine and Surgery (AL), Karolinska Institutet Solna, Stockholm, Sweden; Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada (MW); Department of Nutrition, Harvard School of Public Health, Boston, MA (KW); State Key Laboratory of Oncogene and Related Genes and Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China (YBX); Department of Medical Oncology and Therapeutic Research, City of Hope National Medical Center, Duarte, CA (YY); Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada (BWZ); The University of Ottawa, Ottawa, Ontario, Canada (BWZ); Division of Noncommunicable Disease Epidemiology and Southwest Hospital Clinical Research Center, Third Military Medical University, Chongqing, China (BZ); Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, South Korea (SSK); South Korea Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, South Korea (SSK); Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA (JCF); Department of Clinical Genetics, Karolinska University Hospital Solna, Stockholm, Sweden (AL).

ARTICLE

ARTICLE

Christoph Mancao is an employee of Genentech and holds shares/stocks from Roche/Genentech. The other authors have no competing interests to declare.

References

1. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343(2):78–85.
2. Burt R. Inheritance of colorectal cancer. *Drug Discov Today Dis Mech*. 2007;4(4): 293–300.
3. Peters U, Bien S, Zubair N. Genetic architecture of colorectal cancer. *Gut*. 2015; 64(10):1623–1636.
4. Lemire M, Qu C, Loo LW, et al. A genome-wide association study for colorectal cancer identifies a risk locus in 14q23.1. *Hum Genet*. 2015;134(11–12): 1249–1262.
5. Orlando G, Law PJ, Palin K, et al. Variation at 2q35 (PNKD and TMBIM1) influences colorectal cancer risk and identifies a pleiotropic effect with inflammatory bowel disease. *Hum Mol Genet*. 2016;25(11):2349–2359.
6. Zeng C, Matsuda K, Jia WH, et al. Identification of susceptibility loci and genes for colorectal cancer risk. *Gastroenterology*. 2016;150(7):1633–1645.
7. Gruber SB, Moreno V, Rozek LS, et al. Genetic variation in 8q24 associated with risk of colorectal cancer. *Cancer Biol Ther*. 2007;6(7):1143–1147.
8. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet*. 2007;39(11):1315–1317.
9. Carvajal-Carmona LG, Cazier JB, Jones AM, et al. Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: Refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. *Hum Mol Genet*. 2011;20(14): 2879–2988.
10. Houlston RS, Cheadle J, Dobbins SE, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat Genet*. 2010;42(11):973–977.
11. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet*. 2008;40(12):1426–1435.
12. Schumacher FR, Schmit SL, Jiao S, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun*. 2015;6: 7138.
13. Amos CI, Dennis J, Wang Z, et al. The OncoArray Consortium: A network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomarkers Prev*. 2017;26(1):126–135.
14. Zhang B, Jia WH, Matsuda K, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet*. 2014; 46(6):533–542.
15. Wang H, Burnett T, Kono S, et al. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. *Nat Commun*. 2014;5:4613.
16. Wang H, Haiman CA, Burnett T, et al. Fine-mapping of genome-wide association study-identified risk loci for colorectal cancer in African Americans. *Hum Mol Genet*. 2013;22(24):5048–5055.
17. Schmit SL, Schumacher FR, Edlund CK, et al. Genome-wide association study of colorectal cancer in Hispanics. *Carcinogenesis*. 2016;37(6):547–556.
18. Schumacher FR, Schmit SL, Jiao S, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun*. 2015;6:7138.
19. Closa A, Cordero D, Sanz-Pamplona R, et al. Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis. *Carcinogenesis*. 2014;35(9):2039–2046.
20. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 2015;348(6235): 648–660.
21. Fortini BK, Tring S, Plummer SJ, et al. Multiple functional risk variants in a SMAD7 enhancer implicate a colorectal cancer risk haplotype. *PLoS One*. 2014; 9(11):e111914.
22. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28(10):1045–1048.
23. O'Geen H, Echipare L, Farnham PJ. Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol Biol*. 2011;791: 265–286.
24. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*. 2004;306(5696):636–640.
25. Tung N, Domchek SM, Stadler Z, et al. Counselling framework for moderate-penetrance cancer-susceptibility mutations. *Nat Rev Clin Oncol*. 2016;13(9): 581–588.
26. National Comprehensive Cancer Network. Genetic/familial high-risk assessment: Colorectal (version 2.2016). https://www.nccn.org/professionals/physician_gls/f_guidelines.asp. Accessed October 27, 2016.
27. Hsu L, Jeon J, Brenner H, et al. A model to determine colorectal cancer risk using common genetic susceptibility loci. *Gastroenterology*. 2015;148(7): 1330–1339, e14.
28. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet*. 2008;40(12):1426–1435.
29. Karami S, Han Y, Pande M, et al. Telomere structure and maintenance gene variants and risk of five cancer types. *Int J Cancer*. 2016;139(12):2655–2670.
30. Wang Z, Zhu B, Zhang M, et al. Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet*. 2014;23(24):6616–6633.
31. Walsh KM, Codd V, Smirnov IV, et al. Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk. *Nat Genet*. 2014;46(7):731–735.
32. Bojesen SE, Pooley KA, Johnatty SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet*. 2013;45(4):371–384, 384e1–384e2.
33. Haiman CA, Chen GK, Vachon CM, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet*. 2011;43(12):1210–1214.
34. Kinnersley B, Migliorini G, Broderick P, et al. The TERT variant rs2736100 is associated with colorectal cancer risk. *Br J Cancer*. 2012;107(6):1001–1008.
35. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*. 2007;39(5):645–649.
36. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet*. 2009;41(9):986–990.
37. Petersen GM, Amundadottir L, Fuchs CS, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat Genet*. 2010;42(3):224–228.
38. Landi MT, Chatterjee N, Yu K, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet*. 2009;85(5):679–691.

ARTICLE