Identification of Novel Loci and New Risk Variant in Known Loci for Colorectal Cancer Risk in East Asians

A full list of authors and affiliations appears at the end of the article.

Abstract

**Background**—Risk variants identified so far for colorectal cancer explain only a small proportion of familial risk of this cancer, particularly in Asians.

**Methods**—We performed a genome-wide association study (GWAS) of colorectal cancer in East Asians, including 23,572 colorectal cancer cases and 48,700 controls. To identify novel risk loci, we selected 60 promising risk variants for replication using data from 58,131 colorectal cancer cases and 67,347 controls of European descent. To identify additional risk variants in known colorectal cancer loci, we performed conditional analyses in East Asians.

**Results**—An indel variant, rs67052019 at 1p13.3, was found to be associated with colorectal cancer risk at \( P = 3.9 \times 10^{-8} \) in Asians (OR per allele deletion = 1.13, 95% confidence interval = 1.08–1.18). This association was replicated in European descendants using a variant (rs2938616) in complete linkage disequilibrium with rs67052019 \( (P = 7.7 \times 10^{-3}) \). Of the remaining 59 variants, 12 showed an association at \( P < 0.05 \) in the European-ancestry study, including rs11108175 and rs9634162 at \( P < 5 \times 10^{-8} \) and two variants with an association near the genome-wide significance level \( (rs60911071, P = 5.8 \times 10^{-8}; rs62558833, P = 7.5 \times 10^{-8}) \) in the combined...
analyses of Asian- and European-ancestry data. In addition, using data from East Asians, we identified 13 new risk variants at 11 loci reported from previous GWAS.

**Conclusions**—In this large GWAS, we identified three novel risk loci and two highly suggestive loci for colorectal cancer risk and provided evidence for potential roles of multiple genes and pathways in the etiology of colorectal cancer. In addition, we showed that additional risk variants exist in many colorectal cancer risk loci identified previously.

**Impact**—Our study provides novel data to improve the understanding of the genetic basis for colorectal cancer risk.

**Introduction**

Colorectal cancer is the third most commonly diagnosed cancer in men and the second in women, with 1.65 million new cases and almost 835,000 deaths in 2015 (1). Inherited genetic susceptibility contributes significantly to the etiology of colorectal cancer (2). Rare high-penetrance germline mutations in colorectal cancer predisposition genes (APC, MUTYH, MLH1, MSH2, MSH6, PMS2, PTEN, STK11, GREM1, BMPR1A, SMAD4, POLE, POLD1, NTHL1, and TP53; refs. 3–5) are estimate to account for less than 10% of colorectal cancer cases in the general population (3). Over the past decade, genome-wide association studies (GWAS) have identified about 100 independent loci associated with colorectal cancer risk (4–13). These common genetic risk variants, however, explain only a small proportion of the familial relative risk of colorectal cancers (10, 11, 13), indicating that additional susceptibility variants remain to be identified.

Asians differ significantly from European descendants in genetic architectures. Genetic studies in Asians may provide an opportunity to explore the genetic architecture of colorectal cancer including identification of novel variants. We established the Asia Colorectal Cancer Consortium (ACCC) in 2010 to identify new genetic risk factors for colorectal cancer. Over the past 10 years, we have identified about 30 novel colorectal cancer risk loci (6–8, 11, 14). To further increase the statistical power of uncovering novel susceptibility loci for colorectal cancer, we utilized data from studies of 58,131 cases and 67,347 controls of European ancestry (10) to replicate promising risk variants identified in GWAS of 23,572 cases and 48,700 controls recruited from 15 studies conducted in three East Asian Countries (China, Japan, and Korea). Furthermore, we performed conditional analyses to identify potential independent signals at each of the colorectal cancer risk loci identified in previous studies in Asians.

**Materials and Methods**

**Overview of study population and study design**

We recently reported results from a large GWAS conducted in the Asia Colorectal Cancer Consortium (ACCC) that identified 13 novel genetic loci for colorectal cancer risk (11). In this study, we increased the sample size further by including one additional study—the Korean-National Cancer Center CRC Study 2 (Korea-NCC2: 622 cases and 832 controls). Included in the current analysis were 23,572 colorectal cancer cases and 48,700 controls from 15 studies conducted in China, Japan, and South Korea (Supplementary Table S1;
Supplementary Data). To increase the statistical power, we used data from European descendants to replicate promising findings from the ACCC.

On the basis of the results from the meta-analysis of all Asian studies, we selected 60 promising variants \( (P < 5 \times 10^{-3}, \text{Supplementary Table S2}) \) that are 500 kb away from any established colorectal cancer risk loci \( (10, 11) \) at the time of study design for replication using data from 58,131 cases and 67,347 controls of European descent \( (10) \). These cases and controls were derived from three colorectal cancer study consortia: the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), the Colorectal Transdisciplinary (CORECT) Study, and the Colon Cancer Family Registry (CCFR). The details of each included study have been reported previously \( (9, 10) \). For this analysis, the Haplotype Reference Consortium panel without indel variants was used as reference for imputation.

All study protocols were approved by the relevant Institutional Review Boards \( (10, 11) \) and informed consents were obtained from study participants. Research was conducted in accordance with the Belmont report.

**Genotyping and imputation**

Details of genotyping, genotype calling, quality control and imputation for the ACCC have been reported previously \( \text{Supplementary Data; refs. 6–8, 11, 14} \). The genotypes for samples from six studies \( (\text{Shanghai-4, Aichi-2, Korea-NCC, Korea-NCC2, Korea-Seoul, and HCES2-CRC; Supplementary Table S1}) \) genotyped with Illumina MEGA-Expanded Array \( (\text{Illumina Inc.)}) \) were updated on the basis of the manually reclustered genotype cluster files \( (\text{for 123K SNPs}) \). Little evidence of population stratification was found in these studies \( (6, 8, 11), \) based on principal component (PC) analysis, using EIGENSTRAT \( (\text{Supplementary Fig. S1; ref. 15}) \). To increase the genome coverage and facilitate the meta-analysis, the 1000 Genomes Project phase III mixed reference haplotypes \( (\text{version 5}) \) were used to impute untyped genotype data with Michigan Imputation Server \( (\text{minimac3 for imputation and SHAPEIT for prephasing}) \).

We evaluated the quality of imputation using whole-genome sequencing data for the five variants \( (\text{rs67052019, rs60911071, rs62558833, rs11108175, and rs9634162; Supplementary Table S3}) \) that were highly significantly associated with colorectal cancer risk. Data for these five variants were extracted from whole-genome sequence datasets for 290 Shanghai colorectal cancer samples and compared with the genotypes derived from the imputation-based approach. The whole genome sequence was performed using the BGISEQ-500 sequencing platform with paired-end reads in length with \( 2 \times 100 \) bp \( (\text{mean read depth ~50M}) \). The sequencing reads for each sample were mapped to the human reference genome \( (\text{hg38}) \) using the Burrows-Wheeler Aligner BWA program \( (\text{version 0.75}) \). The aligned reads were processed using the Genome Analysis ToolKit \( (\text{GATKv3.7}) \). Variant calling was performed individually for each sample with the GATK HaplotypeCaller tool and all samples together with GenotypeGVCFs to create a complete list of SNPs and indel VCFs. The Variant Quality Score Recalibration \( (\text{VQSR}) \) was then applied to filter variants of low quality.
Statistical analysis

We used the score test implemented in Rvtest (16) to associate genotype dosages with colorectal cancer risk after adjusting for age, sex, and the first five PCs of each individual study. SNPs with a low imputation quality ($R^2 < 0.3$) or a low MAF (<0.1% in the combined samples) were excluded from the downstream analysis. Summary statistics from each of 15 ACCC case–control studies were meta-analyzed using METAL (17) with the inverse variance-weighted fixed effect model. Associations with a $P < 5 \times 10^{-8}$ in the Asian studies alone or in combination with European studies were regarded as genome-wide significant. Each independent locus was defined as ±500 kb on either side of the most significant SNP that reached a genome-wide significant threshold ($P < 5 \times 10^{-8}$). The Cochran Q test (18) was used to evaluate the heterogeneity across studies and subgroups. We did not observe any apparent inflation in association statistics from the ACCC (Supplementary Fig. S2; refs. 6, 8, 11).

We performed approximate conditional analyses based on meta-analysis summary statistics to identify additional independent association signals at each locus using the GCTA-COJO method (19). The linkage disequilibrium (LD) matrix used in the analyses were based on 6,684 unrelated East Asian samples (interindividual genetic relationships < 0.025; ref. 11). Considering relatively small sample sizes of colorectal cancer studies in populations of East Asian ancestry (10, 11), variants that are conditionally independent of the index SNPs within a region and reached an empirical locus-wide significance of $P < 5 \times 10^{-5}$ were considered to be distinct association signals.

In silico functional characterization of novel loci

To identify potential functional variants and target genes at novel risk loci identified in this study, we annotated all variants that are in LD ($r^2 \geq 0.8$, EAS in the 1000 Genomes Project; 131 variants in total; Supplementary Table S4) with each of identified GWAS index variants within 500 kb using ANNOVAR (20). We used PolyPhen2 (21) and SIFT (22) to assess potential functional impact of coding variants. We further used regulatory information from the Roadmap Epigenomic project (epigenomic profiling) and the ENCODE project (regulatory protein binding and regulatory motifs) to characterize these 131 selected variants with the web-based HaploReg v4 (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) as described previously (11). We tested for enrichment of colorectal cancer index variants with functional domains using the software of Genomic Regulatory Elements and Gwas Overlap algoRithm (GREGOR;ref. 23). This method tests for an increase in the number of colorectal cancer-associated index variants (94 independent GWAS index variants identified in populations of East Asian ancestry at $P < 5 \times 10^{-8}$ or replicated in populations of East Asian ancestry at $P < 5 \times 10^{-2}$ for those variants originally identified in populations of European ancestry; Supplementary Table S5), or their LD proxies ($r^2 0.7$, EAS in the 1000 Genomes Project), overlapping with the regulatory features (peaks from H3K4me1 and H3K27ac as enhancers, peaks from H3K4me3 and H3K9ac as promoters, and open chromatin as measured by DNase hypersensitivity) more often than expected by chance by comparing to permuted control sets in which the variants are matched on variant frequency, number of LD proxies, and distance to the nearest gene. A saddle-point approximation was used to estimate the $P$ value based on the distribution of permuted statistics (23).
cis-expression quantitative trait loci analysis

We conducted cis-expression quantitative trait loci (cis-eQTLs) analyses for each of identified novel risk variants. The RNA was extracted from tumor-adjacent normal tissues obtained from 133 East Asian patients with colorectal cancer (7, 8, 11). We profiled gene expression using RNA sequencing with total mapped reads > 14M for each sample (11). We quantified gene expression levels using FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values. For cis-eQTLs analyses, we defined as a region ±1 mb within each risk variant. These patients were genotyped using the Illumina MEGA array as described before (11). We used linear regression models to associate gene expression levels with SNP genotypes with adjustment for sex and the top two PCs. We also evaluated associations of novel risk variants with gene expressions in 246 transverse colon tissues included in the Genotype-Tissue Expression (GTEx) database (24).

Results

In the meta-analysis of all Asian data (23,572 colorectal cancer cases and 48,700 controls; Supplementary Table S1), we identified an indel variant at 1p13.3 (rs67052019) with an OR of 1.13 (95% CI = 1.08–1.18) for colorectal cancer risk per deletion copy at the genome-wide significance level (P < 5 × 10−8; Table 1; Fig. 1; Supplementary Figs. S3 and S4). A variant (rs2938616) in complete LD with rs67052019 (D’EAS or EUR = 1.00, r2EAS or EUR = 1.00) was associated with colorectal cancer risk in populations of European descent with OR of the deletion linked G allele = 1.03 (95% CI = 1.01–1.04) and P = 7.7 x 10−3 (Table 1), and this variant was also associated with colorectal cancer risk in populations of Asian descent [OR of the deletion linked G allele = 1.10 (95% CI = 1.05–1.15) and P = 1.1 x 10−5; Table 1].

Of the remaining 59 variants for replication in populations of European descent (58,131 colorectal cancer cases and 67,347 controls; Supplementary Table S2), 12 variants were replicated in the same direction as observed in East Asian populations at P < 0.05. Three variants (rs4308634, rs11108175, and rs9634162) were identified to be associated with colorectal cancer risk at the genome-wide significance level (P < 5 × 10−8) in the combined meta-analysis results of East-Asians and Europeans (Table 1; Fig. 1; Supplementary Figs. S3 and S4). However, the variant of rs4308634 [OR for the G allele = 1.04 (95% CI = 1.03–1.06) and P = 6.5 × 10−9 in the combined meta-analyses] at the 7p12.3 was in high LD (r2EAS = 0.68) with a nearby variants (rs10951878) recently identified in populations of European ancestry (13). In addition, two variants (rs60911071 and rs62558833) were associated with colorectal cancer risk near the genome-wide significance level (P = 5.8 × 10−8 and 7.5 × 10−8, respectively; Table 1; Fig. 1; Supplementary Fig. S4). Little heterogeneity of the effect size was observed between these two populations for these variants (P ≥ 0.01, Table 1). Stratification analyses of these newly identified risk variants (rs67052019, rs60911071, rs62558833, rs11108175, and rs9634162) by tumor site (colon or rectum) did not identify any significant heterogeneity (P > 0.05; Supplementary Table S6).

The average imputation quality for these five highly significant variants (rs67052019, rs60911071, rs62558833, rs11108175, and rs9634162) for colorectal cancer risk ranged from 0.78 to 0.99 in Asian cohorts and from 0.98 to 0.99 in European cohorts.
(Supplementary Table S3). The overall genotype concordance rates for these five common variants (minor allele frequency ≥0.18; Table 1) between imputed data and whole-genome sequencing data were high (>0.90) based on 290 Shanghai colorectal cancer samples, indicating that the imputation quality was excellent for these SNPs (Supplementary Table S3).

Evaluation of local secondary signals

Independent secondary signals were reported in our previous studies for four colorectal cancer risk loci in East-Asian populations (EIF3H at 8q23.3, NKX2–3 at 10q24.2, VTI1A/TCF7L2 at 10q25.2 and SMAD7 at 18q21.1; refs. 7, 8, 11, 14, 25). In this study, we observed 13 independent secondary signals at 11 additional loci in East Asians at the locus-wide significance level (P < 5.0 × 10^{-5}; SLCO2A1 at 3q22.2, PITX1 at 5q31.1, NOTCH4/HLA-DRB1/HLA-DRB5 at 6p21.32, GATA3 at 10p14, PRICKLE1 at 12q12, LRP1/ARHGAP9 at 12q13.3, MYL2/SH2B3 at 12q24.11, WWOX/MAF at 16q23.2, NXN at 17p13.3, MYHAS/TMEM238L at 17p12 and BMP2 at 20p12.3; Table 2).

Functional characterization of risk loci and cis-eQTL analyses

We used functional genomics data to annotate each of the identified five novel variants (Table 1) as well as their correlated variants (r^2_EAS ≥ 0.80). Aligning these risk variants with histone methylation/acetylation marks and DNase hypersensitivity sites (26) revealed that variants at three loci (1p13.3, 8p21.2, and 12q22) overlapped with the promoter/enhancer histone marks or DNase hypersensitivity sites in gastrointestinal tissues (Supplementary Table S4). This suggests that these variants may be involved in regulating gene expressions in gastrointestinal tissues. We further tested for tissue regulatory element enrichments of 94 colorectal cancer-associated variants identified or replicated in populations of East Asian ancestry (Supplementary Table S5 and Materials and Methods) using GREGOR (23). We found that colorectal cancer-associated variants were strongly associated with regulatory elements in gastrointestinal tissues (fetal stomach, fetal small intestine, fetal large intestine, small intestine, sigmoid colon, colonic mucosa, rectal mucosa, and stomach mucosa; >2.0 × enrichment; Supplementary Tables S7 and S8). Interestingly, monocytes were also enriched as indicated by DNase I hypersensitive sites (P = 8.1 × 10^{-13}, 2.3 × enrichment; Supplementary Tables S7 and S8). Similar results were observed on the basis of colorectal cancer-associated variants identified separately in populations of East Asian ancestry (fetal stomach, fetal small intestine, fetal large intestine, sigmoid colon, rectal mucosa, stomach mucosa, and monocytes; >2.0 × enrichment and P < 7.2 × 10^{-6}) or populations of European ancestry (fetal stomach, fetal small intestine, fetal large intestine, sigmoid colon, colonic mucosa, and rectal mucosa; >2.0 × enrichment and P < 6.7 × 10^{-10}).

We performed cis-eQTL analyses using transcriptome data from tumor-adjacent normal colon tissues from 133 patients with colorectal cancer of East Asian ancestry (Supplementary Table S9) and transverse colon tissues from 246 individuals predominantly of European ancestry in the GTEx (Supplementary Table S10). Significant correlations at P < 0.05 were found for 3 and 6 SNP–gene expression pairs in the East Asian and GTEx datasets, respectively. The colorectal cancer risk (deletion) allele of rs67052019 was associated with reduced expression of UBL4B and GPR61, but increased expression of...
**KCNC4-AS1** (consistent between the East Asian and GTEx datasets) and **TMEM167B**. The colorectal cancer risk T allele of rs62558833 was associated with increased expression of **SMU1**, **DCAF12**, and **NUDT2**. The CRC risk A allele of rs11108175 was associated with increased expression of **VEZT**.

**Discussion**

In this meta-analysis of 23,572 colorectal cancer cases and 48,700 controls in East-Asians and follow-up replication analyses of 58,131 colorectal cancer cases and 67,347 controls in individuals of European descent, we identified three novel risk loci and two highly suggestive loci for colorectal cancer risk. In addition, we identified 13 secondary signals at 11 known colorectal cancer risk loci in East Asians. Using functional genomics data, we showed that three of the newly identified risk variants, or their highly correlated variants, are located in regulatory regions of the genome. It indicates that these variants potentially regulate the expression of nearby genes in gastrointestinal tissues. In addition, monocytes were implicated in colorectal cancer carcinogenesis for the first time based on tissue regulatory element enrichment analyses. Our cis-eQTL analyses provide additional supports for a possible role of several risk variants identified in our study in regulating expression of cancer-related genes. Our study provides novel information toward the understanding of the genetic and biological basis of colorectal cancer.

At the 1p13.3 locus, the deletion allele of rs67052019 was associated with increased colorectal cancer risk. The variant rs67052019 was 58.8 Kb upstream of the **EPS8L3** gene. The function of the encoded protein by **EPS8L3** is unknown. Interestingly, the deletion allele of rs67052019 was associated with increased **KCNC4-AS1** (**KCNC4 antisense RNA**) expression in both Asian and GTEx datasets. The encode protein by **KCNC4** is the voltage gated Kv3.4 potassium channel protein involved in regulating mammalian cellcycle (27, 28). Therroles of Kv channels in cancer development and progression have been well established, and they are not only involved in cell proliferation and tumor growth, but also in cell migration and metastasis (29, 30).

At the 8p21.2 locus, rs60911071 is 35 kb downstream of **STC1**. Another variant (rs2928679, $r^2_{EAS \text{ or } EUR} = 0.0$, between rs60911071 and rs2928679) in this region was reported to be associated with prostate cancer risk in populations of European ancestry (31). The encoded protein by **STC1** is a secreted glycoprotein, and is expressed ubiquitously, including in the gastrointestinal tract. STC-1 is reported to mediate the metastatic effect of platelet-derived growth factor signaling in colorectal cancer–associated fibroblasts (32). High expressions of **STC1** are correlated with poor postoperative survival in patients with colorectal cancer (33).

At the 9p13.3 locus, rs62558833 is an intronic variant of **UBAP2**. The encoded protein by **UBAP2** functions in the ubiquitination pathway. It inhibits the invasion of hepatocellular carcinoma cell by ubiquitinating and degrading Annexin A2 (34). The colorectal cancer risk T allele of rs62558833 was associated with an increased expression of **SMU1** (Supplementary Table S9), **DCAF12**, and **NUDT2** (Supplementary Table S10). The encoded protein by **SMU1** is suggested to be involved in genome stability maintenance by negatively regulating DNA synthesis (35). The encoded protein by **DCAF12** is required for...
developmental apoptosis (36). The encoded protein by NUDT2 is suggested to be a tumor-promoting factor, and high expressions of NUDT2 are associated with poor prognosis and an increased risk of breast cancer recurrence (37, 38).

At the 12q22 locus, rs11108175 is a downstream variant of NTN4. Another nearby variant of rs17356907 ($r^2_{EAS}$ or $r^2_{EUR} = 0.0$ between rs11108175 and rs17356907) was reported to be associated with breast cancer risk in European populations (39). NTN4 encodes a member of the netrin family of proteins that functions in various biological processes, including axonal guidance, tumorigenesis, and angiogenesis (40). NTN4 overexpression is observed to suppress primary and metastatic colorectal tumor progression through inhibiting tumor growth and angiogenesis (41–43). The colorectal cancer risk A allele of rs11108175 was nominally associated with an increased VEZT expression (Supplementary Table S10). The encode protein by VEZT is a ubiquitous transmembrane protein that is localized to adherens junctions in epithelial cells (44). How the aberrant expression of VEZT affects colorectal cancer risk warrants further investigations.

At the 12q24.21 locus, rs9634162 is 9.9 kb downstream of the TBX3 gene. Two nearby variants, rs59336 ($P = 3.7 \times 10^{-7}$ on colorectal cancer risk, $r^2_{EAS} = 0.58$ and $r^2_{EUR} = 0.66$ between rs9634162 and rs59336) (45) and rs1427760 ($P = 2.5 \times 10^{-7}$ on colorectal cancer risk, $r^2_{EAS} = 0.78$ and $r^2_{EUR} = 0.89$ between rs9634162 and rs1427760; ref. 10) were reported to be associated with colorectal cancer risk in populations of European ancestry. We established this locus as one of the bona fide colorectal cancer risk loci at genome-wide significance. The protein encoded by TBX3 belongs to the evolutionarily conserved T-box family of transcription factors that play critical roles in early embryonic development. Overexpression of Tbx3 is associated with multiple cancers potentially by modulating cell proliferation and survival, tumor formation and metastasis, and drug resistance (46). Emerging evidence suggests that Tbx3 is not only important for stem cell self-renewal, but also is extensively involved in cancer stemness (46, 47).

Independent secondary signals were observed at colorectal cancer risk loci in both East-Asian populations (7, 8, 11, 14, 25) and European populations (10, 13, 48–51). We observed 13 independent secondary signals at 11 additional colorectal cancer loci in East Asians. Conditional joint analysis could refine association signals and uncover additional GWAS loci for colorectal cancer. The loci of MYL2/SH2B3 at 12q24.11 and LRP1/ARHGAP9 at 12q13.3 that were reported in European populations for colorectal cancer risk (10, 52) only reached the genome-wide significance in the conditional joint analysis in East Asians. The missense variants of rs78894077 in SH2B3 at 12q24.11 only exist in East Asian populations (Table 2) and is predicted to be highly “deleterious” in both PolyPhen2 (21) and SIFT (22). This strongly implicates that SH2B3 is the underlying causal gene for colorectal cancer risk at this locus. The low-frequency intronic variant of rs368674461 in MYHAS at 17p12 also only exists in East Asian populations (Table 2); however, the function of MYHAS is currently unknown.

In summary, we identified three novel variants and two highly suggestive loci for colorectal cancer risk in this large GWAS of colorectal cancer. Combining information from functional annotations, cis-eQTL analyses and literature review, we propose the putative candidate
genes for these loci: KCNC4-AS1 at 1p13.3, STC1 at 8p21.2, NUDT2 at 9p13.3, NTN4 at 12q22, and TBX3 at 12q24.21. Some of the putative target genes suggested by the results from our studies are located in established pathways for colorectal tumorigenesis, such as the maintaining of colon stem cells (TBX3). Multiple independent signals exist at many colorectal cancer loci; therefore, more familial relative risk of colorectal cancer could be explained at a locus when multiple independent signals were considered. However, extensive fine-mapping and functional follow-up studies are needed to identify the causal variants and target genes at each of the identified regions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Yingchang Lu¹, Sun-Seog Kweon²,³, Qiuyin Cai¹, Chizu Tanikawa⁴, Xiao-Ou Shu¹, Wei-Hua Jia⁵, Yong-Bing Xiang⁶, Jeroen R. Huyghe⁷, Tabitha A. Harrison⁷, Jeongseon Kim⁸, Aesun Shin⁹, Dong-Hyun Kim¹⁰, Keitaro Matsuo¹¹,¹², Sun Ha Jee¹³, Xingyi Guo¹, Wanning Wen¹, Jiajun Shi¹, Bingshan Li¹⁴, Nan Wang¹⁵, Min-Ho Shin¹⁶, Hong-Lan Li¹⁶, Zefang Ren¹⁷, Jae Hwan Oh¹⁸, Isao Oze¹⁰, Yoon-Ok Ahn¹⁹, Keum Ji Jung²⁰, Jing Gao²⁰, Yu-Tang Gao²⁰, Zhi-Zhong Pan⁵, Yoichiro Kamatani²¹,²², Andrew T. Chan²³,²⁴,²⁵,²⁶, Andrea Gsur²⁷, Jeroen R. Huyghe²⁸, Loic Le Marchand²⁹, Li Li³⁰, Annika Lindblom³¹,³², Victor Moreno³³,³⁴,³⁵, Polly A. Newcomb³⁶,³⁷,³⁸, Kenneth Offit³⁷,³⁸, Paul D.P. Pharoah³⁹, Franzel J.B. van Duijnhoven⁴⁰, Bethany Van Guelpen⁴¹, Pavel Vodicca⁴²,⁴³,⁴⁴, Stephanie J. Weinstein⁴⁵, Alicja Wolk⁴⁶, Anna H. Wu⁴⁷, Li Hsu⁷,⁴⁸, Yi-Xin Zeng⁵, Jirong Long¹, Ulrike Peters⁷,⁴⁹, Koichi Matsuda⁵⁰, Wei Zheng¹

Affiliations

¹Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee. ²Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, South Korea. ³Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, South Korea. ⁴Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. ⁵State Key Laboratory of Oncology in South China, Cancer Center, Sun Yat-sen University, Guangzhou, China. ⁶State Key Laboratory of Oncogene and Related Genes & Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. ⁷Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington. ⁸Department of Cancer Biomedical Science, Graduate School of Cancer Science and Policy, National Cancer Center, Gyeonggi-do, South Korea. ⁹Department of Preventive Medicine, Seoul National University College of Medicine, Seoul National University Cancer Research Institute, Seoul, Korea. ¹⁰Department of Social and Preventive Medicine, Hallym University College of Medicine, Okcheon-dong, Korea. ¹¹Division of Molecular and
Clinical Epidemiology, Aichi Cancer Center Research Institute, Nagoya, Japan.  
12Department of Epidemiology, Nagoya University Graduate School of Medicine, Nagoya, Japan.  
13Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea.  
14Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, Tennessee.  
15General Surgery Department, Tangdu Hospital, Fourth Military Medical University, Xi’an, Shaanxi, China.  
16Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, South Korea.  
17School of Public Health, Sun Yat-sen University, Guangzhou, China.  
18Center for Colorectal Cancer, National Cancer Center Hospital, National Cancer Center, Gyeonggi-do, South Korea.  
19Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, South Korea.  
20Institute for Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea.  
21Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan.  
22Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.  
23Division of Gastroenterology, and Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts.  
24Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts.  
26Department of Epidemiology, and Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts.  
27Institute of Cancer Research, Department of Medicine I, Medical University Vienna, Vienna, Austria.  
28Department of Medicine I, University Hospital Dresden, Technische Universität Dresden (TU Dresden), Dresden, Germany.  
29University of Hawaii Cancer Center, Honolulu, Hawaii.  
30Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio.  
31Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden.  
32Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden.  
33Cancer Prevention and Control Program, Catalan Institute of Oncology-IDIBELL, L'Hospital de Llobregat, Barcelona, Spain.  
34CIBER Epidemiología y Salud Publica (CIBERESP), Madrid, Spain.  
35Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain.  
36School of Public Health, University of Washington, Seattle, Washington.  
37Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York.  
38Department of Medicine, Weill Cornell Medical College, New York, New York.  
39Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.  
40Division of Human Nutrition, Wageningen University and Research, Wageningen, the Netherlands.  
41Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden.  
42Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic.  
43Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic.  
44Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen, Czech
Acknowledgments

The authors thank all study participants and research staff of all parent studies for their contributions and commitment to this project. The authors thank Vanderbilt staff members Ms. Jing He for data processing and analyses and Mr. Marshal Younger for editing and preparing the manuscript. The work at Vanderbilt University Medical Center was supported by U.S. NIH grants R01CA188214, R37CA070867, UM1CA182910, R01CA124558, R01CA158473, and R01CA148667, as well as Anne Potter Wilson Chair funds from the Vanderbilt University School of Medicine. Sample preparation and genotyping assays at Vanderbilt University were conducted at the Survey and Biospecimen Shared Resources and Vanderbilt Microarray Shared Resource, which are supported in part by the Vanderbilt-Ingram Cancer Center (P30CA068485). Imputation and statistical analyses were performed on servers maintained by the Advanced Computing Center for Research and Education at Vanderbilt University (Nashville, TN). Studies (listed with grant support) participating in the Asia Colorectal Cancer Consortium include the Shanghai Women’s Health Study (US NIH, R37CA070867, UM1CA182910), the Shanghai Men’s Health Study (US NIH, R01CA082729, UM1CA173640), the Shanghai Breast and Endometrial Cancer Studies (US NIH, R01CA064277 and R01CA092585; contributing only controls), the Shanghai Colorectal Cancer Study 3 (US NIH, R37CA070867, R01CA188214 and Anne Potter Wilson Chair funds), the Guangzhou Colorectal Cancer Study (National Key Scientific and Technological Project, 2011ZX09303-001-04; the National Basic Research Program, 2011CB504303, contributing only controls), the Natural Science Foundation of China, 81072383, contributing only controls), the Hwasun Cancer Epidemiology Study—Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital Biomedical Research Institute, HCR18007), the Japan BioBank Colorectal Cancer Study (grant from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government), the Aichi Colorectal Cancer Study (Grant-in-Aid for Cancer Research, grant for the Third Term Comprehensive Control Research for Cancer and Grants-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, 17015018 and 22150001), the Korea-NCC (National Cancer Center) Colorectal Cancer Study (Basic Science Research Program through the National Research Foundation of Korea, 2010-0010276 and 2013R1A1A2A10008260; National Cancer Center Korea, 0910220), and the KCPS-II Colorectal Cancer Study (National R&D Program for Cancer Control, 1631020; Seoul R&D Program, 10526). Funding information for the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and its participating studies is provided in this section. GECCO: NCI, NIH, U.S. Department of Health and Human Services (U10 CA146930, U01 CA137088, R01 CA059045, U01 CA164930, R21 CA191312, R01CA191407). Genotyping/ Sequencing services were provided by the Center for Inherited Disease Research (CIDR; X01-HG008596 and X-01-HG007585). CIDR is fully funded through a federal contract from the NIH to The Johns Hopkins University, contract number HHSN268201200008I. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. ASTERISK: a Hospital Clinical Research Program (PFHC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Francaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Gene&étique and the Ligue Regionale Contre le Cancer (LRC). The ATBC Study is supported by the Intramural Research Program of the U.S. NCI, NIH, and by U.S. Public Health Service contract HHSN262201500005C from the NCI, Department of Health and Human Services. CLUE II: This research was funded by the American Institute for Cancer Research and the Maryland Cigarette Restitution Fund at Johns Hopkins, and the NCI (P30 CA06973, to W.G. Nelson). COLO2&3: NIH (R01 CA060987). ColoCare: This work was supported by the NIH (grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueirépdo/Colditz), 2P30CA015704–40 (Gilliland), R01 CA207371 (Ulrich/Li)), the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative. The Colon Cancer Family Registry (CFR) Illumina GWAS was supported by funding from the NCI, NIH (grant numbers U01 CA122839, R01 CA143247, to G. Casey). The Colon CFR participant recruitment and collection of data and biospecimens used in this study were supported by the NCI, NIH (grant number U01 CA167551). The content of this manuscript does not necessarily reflect the views or policies of the NCI or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government, any cancer registry, or the CCFR. COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Fonds, including funds from grant 2014/1179 as part of
Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. MECC: NIH (R37 CA54281, P01 CA033619, and R01 CA063464). MECC: This work was supported by the NIH, U.S. Department of Health and Human Services (R01 CA81488, to S.B. Gruber and G. Rennert).

MSKCC: The work at Sloan Kettering in New York was supported by the Robert and Kate Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. Moffitt: This work was supported by funding from the NIH (grant numbers R01 CA189184, P30 CA076292), Florida Department of Health Bankhead-Coley Grant 09BN-13, and the University of South Florida Oehler Foundation. Moffitt contributions were supported in part by the Total Cancer Care Initiative, Collaborative Data Services Core, and Tissue Core at the H. Lee Moffitt Cancer Center & Research Institute, a NCI-designated Comprehensive Cancer Center (grant number P30 CA076292).

NCCS I & II: We acknowledge funding support for this project from the NIH, R01 CA66635 and P30 DK034987. NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the NIH, U.S. Department of Health and Human Services (U01 CA74783); and National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute. NSHDS: Swedish Research Council; the Swedish Cancer Society; Region Västerbotten; the Lion’s Cancer Research Foundation, the Faculty of Medicine and Insamlingstiftelsen, all at Umeå University; and the Margareta Dammborg Memorial Fund. OFCCR: NIH, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); see CCFR section above. Additional funding toward genetic analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research, and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation. OSUMC: OCCPI funding was provided by Pelotonia, and HNPPC funding was provided by the NCI (CA16058 and CA67941). PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by NIH, Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 00438. PMH: NIH (R01 CA076366, to P.A. Newcomb). SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/ A10651); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge: SELECT: The Selenium and Vitamin E Cancer Prevention Trial (SELECT) was supported by the NCI of the NIH under award numbers UM1CA182883 and U10CA37429. SMS: This work was supported by the National Cancer Institute (grant P01 CA074184, to J.D. Potter and P.A. Newcomb; grants R01 CA097325, R03 CA153323, and K05 CA152715, to P.A. Newcomb), and the National Center for Advancing Translational Sciences at the NIH (grant KL2 TR000421, to A.N. Burnett-Hartman) The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the Swedish Research Council, K2015–55X–22674–01–4, K2008–55X–20157–03-3, K2006–72X–20157–01–2, and the Stockholm County Council (ALF project). Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council/Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute's Distinguished Professor Award to A. Wolk. UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614VITAL: NIH (K05 CA154337). WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, NIH, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN26820110002C, HHSN268201100003C, HHSN268201100004C, and HHSN267120100004C. ASTERISK: We are very grateful to Dr. Bruno Buechner without whom this project would not have existed. CLUE II: We thank Judith Hoffman-Bolton, Senior Research Program Coordinator, for her contributions to the conduct of CLUE. CORSA: We are grateful to Doris Meji and Monika Hunjadi for laboratory assistance. CPS-II: The authors would like to acknowledge the contribution to this study from the centralized cancer registries supported through the CDC and Prevention National Program of Cancer Registries, and cancer registries supported by the National Cancer Institute Surveillance, Epidemiology, and End Results Program. DACHS: Ute Handte-Daub, Uyt Benscheid, Muhabbet Celik, and Ursula Eiber for excellent technical assistance. EDRN: We acknowledge the following contributors to the development of the resource: University of Pittsburgh School of Medicine, Department of Gastroenterology, Hepatology and Nutrition: Lynda Dzubinski; University of Pittsburgh School of Medicine, Department of Pathology: Michelle Bisciglia; and University of Pittsburgh School of Medicine, Department of Biomedical Informatics. EPICOLON: We acknowledge the Spanish National DNA Bank, Biobank of Hospital Clinic–IDIBAPS and Biobanco Vasco for the availability of the samples. The work was carried out in part at the Esther Day Open Center, Barcelona. Harvard cohorts (HPFS, NHS, PHS): We would like to thank the participants and staff of the HPFS, NHS, and PHS for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WV. The authors assume full responsibility for analyses and interpretation of these data. LCCS: We recognize the contributions of Jennifer Barrett, Robin Waxman, Gillian Smith,and Emma Northwood in conducting this study. NSHDS: We thank the Biobank Research Unit at Umea University, Biobanken Norr at Region Västerbotten, and the cohort participants. WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full list of WHI investigators can be found at http://www.whi.org/researchers/Documents %20%20From%20a%20Page/WHI%20Investigator%20Short%20List.pdf. Collectively, the authors thank all those who helped make this research possible, including patients, healthy control persons, physicians, staff, technicians,
investigators, students, participating clinics and hospitals, state registries, and study teams for their participation in this study.

References


Figure 1.
Regional association plots of five colorectal cancer risk loci in Asian descendants (A, rs67052019; B, rs60911071; C, rs62558833; D, rs11108175; and E, rs9634162). Each dot represents the P value (on a log_{10} scale) of an SNP with colorectal cancer risk based on the meta-analysis results in East Asians only and is presented according to its genomic position (NCBI Build 37). The most significantly associated SNP in the combined meta-analyses is represented by purple. The color of all other SNPs indicates the level of LD with the lead SNP (estimated by ASN r^2 from the 1000 Genome Project data). Recombination rates were also estimated from 1000 Genomes Project data, and gene annotations within the 2 Mb regions that are centered on the newly identified risk variants were obtained from the UCSC Genome Browser.
Table 1.

Novel loci associated with colorectal cancer risk identified in this study.

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNP</th>
<th>Position (hg19)</th>
<th>Nearby genes</th>
<th>Alleles</th>
<th>East Asian descendants (up to 23,572 cases and 48,700 controls)</th>
<th>European descendants (up to 58,131 cases and 67,347 controls)</th>
<th>Combined meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RAF</td>
<td>OR (95% CI) P Phet</td>
<td>OR (95% CI) P Phet</td>
<td>OR (95% CI) P Phet</td>
</tr>
<tr>
<td>1p13.3</td>
<td>rs67052019</td>
<td>110,365,461</td>
<td>EPS8L3</td>
<td>Del/Ref</td>
<td>0.18 1.13 (1.08–1.18) 3.9 × 10⁻⁸</td>
<td>– NA – – – – – – – 1.10 (1.05–1.15) 1.1 × 10⁻⁵</td>
<td>0.18 0.28 3.9 × 10⁻⁸</td>
</tr>
<tr>
<td>12q22</td>
<td>rs11108175</td>
<td>96,050,887</td>
<td>NTN4</td>
<td>A/G</td>
<td>0.80 1.08 (1.05–1.12) 5.7 × 10⁻⁶</td>
<td>0.59 1.05 (1.03–1.07) 2.6 × 10⁻⁷</td>
<td>0.59 1.05 (1.03–1.07) 2.6 × 10⁻⁷</td>
</tr>
<tr>
<td>12q24.21</td>
<td>rs9634162</td>
<td>115,098,094</td>
<td>TBX3</td>
<td>A/G</td>
<td>0.48 1.07 (1.03–1.10) 2.4 × 10⁻⁵</td>
<td>0.51 1.04 (1.03–1.06) 7.5 × 10⁻⁷</td>
<td>0.51 1.04 (1.03–1.06) 7.5 × 10⁻⁷</td>
</tr>
<tr>
<td>8p21.2</td>
<td>rs60911071</td>
<td>23,664,632</td>
<td>STC1</td>
<td>G/C</td>
<td>0.64 1.07 (1.04–1.10) 1.2 × 10⁻⁶</td>
<td>0.95 1.05 (1.01–1.10) 3.2 × 10⁻²</td>
<td>0.95 1.05 (1.01–1.10) 3.2 × 10⁻²</td>
</tr>
<tr>
<td>9p13.3</td>
<td>rs62558833</td>
<td>34,039,002</td>
<td>UBAP2</td>
<td>T/C</td>
<td>0.55 1.08 (1.05–1.11) 3.6 × 10⁻⁷</td>
<td>0.29 1.03 (1.01–1.05) 1.7 × 10⁻³</td>
<td>0.29 1.03 (1.01–1.05) 1.7 × 10⁻³</td>
</tr>
</tbody>
</table>

Abbreviations: alleles, risk allele/reference allele; CI, confidence interval; Phet: P value derived from the heterogeneity test; RAF, risk allele frequency.

Notes:

a Del: deletion for ACAGAGATGTAGGGGC.

b Variant of rs2938616 (D' EAS or EUR = 1.00, r² EAS or EUR = 1.00 with rs67052019) was associated with colorectal cancer risk in populations of European ancestry [OR for the deletion linked G allele of rs2938616 = 1.03 (95% CI = 1.01–1.04) and P = 7.7 × 10⁻³] and in populations of East Asian ancestry [OR of the deletion linked G allele of rs2938616 = 1.10 (95% CI = 1.05–1.15) and P = 1.1 × 10⁻⁵].

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2020 October 19.
Secondary independent association signals ($P < 5 \times 10^{-5}$) identified within 500 kb of previously reported risk loci for colorectal cancer: results from the ACCC.

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNP</th>
<th>CHR</th>
<th>Pos</th>
<th>Nearest gene</th>
<th>EA</th>
<th>Single-SNP meta-analysis</th>
<th>Joint analysis</th>
<th>Multiple-independent-variant loci New/known</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Freq</td>
<td>OR (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>3q22.2</td>
<td>rs4854776</td>
<td>3</td>
<td>133,687,864</td>
<td>SLC02A1</td>
<td>A</td>
<td>0.78</td>
<td>0.94 (0.91–0.97)</td>
<td>2.43 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>rs61510274</td>
<td>3</td>
<td>133,740,515</td>
<td></td>
<td>C</td>
<td>0.41</td>
<td>0.90 (0.88–0.93)</td>
<td>2.68 × 10^{-14}</td>
</tr>
<tr>
<td></td>
<td>rs76491686</td>
<td>3</td>
<td>133,815,683</td>
<td></td>
<td>C</td>
<td>0.65</td>
<td>1.07 (1.05–1.10)</td>
<td>1.60 × 10^{-7}</td>
</tr>
<tr>
<td>5q31.1</td>
<td>rs7722513</td>
<td>5</td>
<td>134,464,066</td>
<td>PITX1</td>
<td>C</td>
<td>0.25</td>
<td>1.15 (1.11–1.18)</td>
<td>9.14 × 10^{-20}</td>
</tr>
<tr>
<td></td>
<td>rs3917784</td>
<td>5</td>
<td>134,497,599</td>
<td></td>
<td>A</td>
<td>0.10</td>
<td>1.10 (1.05–1.15)</td>
<td>6.56 × 10^{-5}</td>
</tr>
<tr>
<td>6p21.32</td>
<td>rs3800041</td>
<td>6</td>
<td>32,191,339</td>
<td>NOTCH4/HLA-DRB1/HLA-DRB5</td>
<td>T</td>
<td>0.14</td>
<td>1.15 (1.10–1.21)</td>
<td>3.61 × 10^{-9}</td>
</tr>
<tr>
<td></td>
<td>rs4713534</td>
<td>6</td>
<td>32,445,926</td>
<td></td>
<td>T</td>
<td>0.07</td>
<td>1.14 (1.06–1.23)</td>
<td>2.54 × 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>rs569582972</td>
<td>6</td>
<td>32,558,756</td>
<td></td>
<td>T</td>
<td>0.05</td>
<td>1.21 (1.10–1.33)</td>
<td>7.74 × 10^{-5}</td>
</tr>
<tr>
<td>8q23.3</td>
<td>rs649654</td>
<td>8</td>
<td>117,632,965</td>
<td>EIF3H</td>
<td>C</td>
<td>0.50</td>
<td>1.12 (1.09–1.15)</td>
<td>1.72 × 10^{-17}</td>
</tr>
<tr>
<td></td>
<td>rs12541711</td>
<td>8</td>
<td>117,715,457</td>
<td></td>
<td>A</td>
<td>0.03</td>
<td>1.26 (1.15–1.37)</td>
<td>7.67 × 10^{-7}</td>
</tr>
<tr>
<td>10p14</td>
<td>rs533919</td>
<td>10</td>
<td>8,266,151</td>
<td>GATA3</td>
<td>A</td>
<td>0.08</td>
<td>0.90 (0.86–0.95)</td>
<td>1.91 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>rs7894531</td>
<td>10</td>
<td>8,734,761</td>
<td></td>
<td>A</td>
<td>0.36</td>
<td>0.86 (0.84–0.89)</td>
<td>2.52 × 10^{-28}</td>
</tr>
<tr>
<td>10q24.2</td>
<td>rs6854283</td>
<td>10</td>
<td>101,290,301</td>
<td>NKK2-3, SLC25A28</td>
<td>T</td>
<td>0.44</td>
<td>0.92 (0.90–0.94)</td>
<td>6.02 × 10^{-11}</td>
</tr>
<tr>
<td></td>
<td>rs61871279</td>
<td>10</td>
<td>101,343,705</td>
<td></td>
<td>T</td>
<td>0.17</td>
<td>1.08 (1.05–1.12)</td>
<td>5.37 × 10^{-6}</td>
</tr>
<tr>
<td>10q25.2</td>
<td>rs4554811</td>
<td>10</td>
<td>114,278,734</td>
<td>VTI1A/TCF7L2</td>
<td>A</td>
<td>0.73</td>
<td>0.90 (0.87–0.92)</td>
<td>8.08 × 10^{-14}</td>
</tr>
<tr>
<td></td>
<td>rs11196172</td>
<td>10</td>
<td>114,726,843</td>
<td></td>
<td>A</td>
<td>0.69</td>
<td>1.12 (1.09–1.15)</td>
<td>1.74 × 10^{-15}</td>
</tr>
<tr>
<td>12q12</td>
<td>rs117912059</td>
<td>12</td>
<td>43,006,211</td>
<td>PRICKLE1</td>
<td>T</td>
<td>0.98</td>
<td>1.23 (1.12–1.36)</td>
<td>3.76 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>rs2730985</td>
<td>12</td>
<td>43,130,624</td>
<td></td>
<td>A</td>
<td>0.37</td>
<td>0.92 (0.90–0.95)</td>
<td>1.24 × 10^{-9}</td>
</tr>
<tr>
<td>12q13.3</td>
<td>rs7398375</td>
<td>12</td>
<td>57,540,848</td>
<td>LRP1/ARHGAP9</td>
<td>C</td>
<td>0.43</td>
<td>1.08 (1.05–1.12)</td>
<td>1.27 × 10^{-7}</td>
</tr>
<tr>
<td></td>
<td>rs7948748</td>
<td>12</td>
<td>57,873,498</td>
<td></td>
<td>T</td>
<td>0.89</td>
<td>1.08 (1.04–1.13)</td>
<td>5.30 × 10^{-4}</td>
</tr>
<tr>
<td>12q24.11</td>
<td>rs17550549</td>
<td>12</td>
<td>111,357,471</td>
<td>MYL2SH2B3</td>
<td>T</td>
<td>0.20</td>
<td>0.91 (0.88–0.95)</td>
<td>1.61 × 10^{-7}</td>
</tr>
<tr>
<td></td>
<td>rs78894077</td>
<td>12</td>
<td>111,856,673</td>
<td></td>
<td>T</td>
<td>0.06</td>
<td>0.86 (0.80–0.92)</td>
<td>2.22 × 10^{-5}</td>
</tr>
<tr>
<td>16q23.2</td>
<td>rs140851213</td>
<td>16</td>
<td>79,754,433</td>
<td>WWOX, MAF</td>
<td>T</td>
<td>0.28</td>
<td>0.92 (0.88–0.96)</td>
<td>3.57 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>rs4341754</td>
<td>16</td>
<td>80,039,621</td>
<td></td>
<td>C</td>
<td>0.43</td>
<td>0.91 (0.89–0.94)</td>
<td>5.12 × 10^{-12}</td>
</tr>
<tr>
<td>Locus</td>
<td>SNP</td>
<td>CHR</td>
<td>Pos</td>
<td>Nearest gene</td>
<td>EA</td>
<td>Single-SNP meta-analysis</td>
<td>Joint analysis</td>
<td>Multiple-independent-variant loci New/known</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>-----</td>
<td>------</td>
<td>--------------</td>
<td>----</td>
<td>--------------------------</td>
<td>---------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OR (95% CI)</td>
<td>P</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>17p13.3</td>
<td>rs9915645</td>
<td>17</td>
<td>812,534</td>
<td>NXN</td>
<td>T</td>
<td>0.48</td>
<td>0.92 (0.90–0.95)</td>
<td>4.87 × 10^{-10}</td>
</tr>
<tr>
<td></td>
<td>rs11651883</td>
<td>17</td>
<td>835,502</td>
<td></td>
<td>T</td>
<td>0.40</td>
<td>1.06 (1.03–1.09)</td>
<td>1.47 × 10^{-5}</td>
</tr>
<tr>
<td>17p12</td>
<td>rs368674461</td>
<td>17</td>
<td>10,485,457</td>
<td>MYHAS/TMEM238L</td>
<td>T</td>
<td>0.01</td>
<td>0.62 (0.49–0.79)</td>
<td>7.85 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>rs1078643</td>
<td>17</td>
<td>10,707,241</td>
<td></td>
<td>A</td>
<td>0.77</td>
<td>1.13 (1.09–1.17)</td>
<td>2.05 × 10^{-13}</td>
</tr>
<tr>
<td>20p12.3</td>
<td>rs6117209</td>
<td>20</td>
<td>6,302,114</td>
<td>BMP2</td>
<td>T</td>
<td>0.80</td>
<td>1.08 (1.04–1.12)</td>
<td>1.84 × 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>rs6085662</td>
<td>20</td>
<td>6,698,372</td>
<td></td>
<td>C</td>
<td>0.21</td>
<td>1.10 (1.07–1.14)</td>
<td>1.69 × 10^{-9}</td>
</tr>
</tbody>
</table>

Abbreviations: CHR, chromosome; CI, confidence interval; EA, effective allele; Freq, frequency of the effective allele; Pos, position (hg19); r, linkage disequilibrium correlation between an SNP and the next adjacent SNP at a locus.

a LD correlation between an SNP and the next adjacent SNP at a locus.

b Known: loci known to contain independent secondary signals for CRC risk in East Asians; new: loci not reported before to contain independent secondary signals for CRC risk in East Asians.