Genome-wide association analyses in east Asians identify new susceptibility loci for colorectal cancer

Wei-Hua Jia1,16, Ben Zhang2,16, Keitaro Matsuo3, Aesun Shin4, Yong-Bing Xiang5, Sun Ha Jee6, Dong-Hyun Kim7, Zefang Ren1, Qiuyin Cai2, Jirong Long2, Jiajun Shi2, Wanqing Wen2, Gong Yang2, Ryan J Delahanty2, Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)8, Colon Cancer Family Registry (CCFR)8, Bu-Tian Ji9, Zhi-Zhong Pan1, Fumihiko Matsuda10, Yu-Tang Gao5, Jae Hwan Oh11, Yoon-Ok Ahn12, Eun Jung Park6, Hong-Lan Li5, Ji Won Park11, Jaeseong Jo6, Jin-Young Jeong7, Satoyo Hosono3, Graham Casey13, Ulrike Peters14,15, Xiao-Ou Shu2, Yi-Xin Zeng1,17 & Wei Zheng2,17

To identify new genetic factors for colorectal cancer (CRC), we conducted a genome-wide association study in east Asians. By analyzing genome-wide data in 2,098 cases and 5,749 controls, we selected 64 promising SNPs for replication in an independent set of samples, including up to 5,358 cases and 5,922 controls. We identified four SNPs with association P values of 8.58 × 10−7 to 3.77 × 10−10 in the combined analysis of all east Asian samples. Three of the four were replicated in a study conducted in 26,060 individuals of European descent, with combined P values of 1.22 × 10−10 for rs647161 (5q31.1), 6.64 × 10−9 for rs2423279 (20p12.3) and 3.06 × 10−9 for rs10774214 (12p13.32 near the CCND2 gene), derived from meta-analysis of data from both east Asian and European-ancestry populations. This study identified three new CRC susceptibility loci and provides additional insight into the genetics and biology of CRC.

CRC is one of the most commonly diagnosed malignancies in east Asia and many other parts of the world1. Genetic factors have an important role in the etiology of both sporadic and familial CRC2. However, less than 6% of CRC cases can be explained by rare, high-penetrance variations in CRC susceptibility genes, explain less than 15% of the heritability for this common malignancy10,11. Furthermore, with the exception of a small study conducted in Japan12, all other GWAS have been conducted in populations of European ancestry, which differ from other populations in certain features of genetic architecture. Many of the variants discovered in populations of European ancestry show only weak or no association with CRC in other ancestry groups13. Therefore, additional GWAS are needed, particularly in populations not of European ancestry, to fully uncover the genetic basis for CRC susceptibility.

In 2009, we initiated the Asia Colorectal Cancer Consortium (ACCC), a GWAS in east Asians, to search for previously unknown genetic risk factors for CRC. The discovery stage (stage 1) consisted of five GWAS conducted in China, Korea and Japan, including 2,293 CRC cases and 5,780 controls (Supplementary Table 1). Cases and controls were genotyped using several SNP arrays, including the Affymetrix Genome-Wide Human SNP Array 6.0 (906,602 SNPs), the Affymetrix Genome-Wide Human SNP Array 5.0 (443,104 SNPs), the Illumina Infinium HumanHap610 BeadChip (592,044 SNPs), the Illumina Human610-Quad BeadChip (620,901 SNPs) and the Illumina HumanOmniExpress BeadChip (729,462 SNPs) (Supplementary Table 1). After quality control exclusions as described previously14–17, 2,098 cases and 5,749 controls remained for this study (Supplementary Tables 1 and 2). Also excluded from the analyses were SNPs with call rate of <95%, genotype concordance rate of <95%

1State Key Laboratory of Oncology in South China, Cancer Center, Sun Yat-sen University, Guangzhou, China. 2Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee, USA. 3Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan. 4Molecular Epidemiology Branch, National Cancer Center, Goyang-si, Korea. 5Department of Epidemiology, Shanghai Cancer Institute, Shanghai, China. 6Institute for Health Promotion, Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea. 7Department of Social and Preventive Medicine, Hallym University College of Medicine, Okcheon-dong, Korea. 8A complete list of members is provided in the Acknowledgements. 9Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, Maryland, USA. 10Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan. 11Center for Colorectal Cancer, National Cancer Center, Goyang-si, Korea. 12Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea. 13Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA. 14Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. 15Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA. 16These authors contributed equally to this work. 17These authors jointly directed this work. Correspondence should be addressed to W.Z. (wei.zheng@vanderbilt.edu).

Received 29 March; accepted 29 November; published online 23 December 2012; doi:10.1038/ng.2505
between positive control samples, minor allele frequency (MAF) of <5% or P value for Hardy-Weinberg equilibrium of <1.0 × 10−5 in controls for each study. Imputation was conducted for each study following the MaCH algorithm18 using phased HapMap 2 Han Chinese in Beijing, China (CHB) and Japanese in Japan (JPT) samples as the reference. No apparent genetic admixture was detected, except for one sample from KCPS-II (Supplementary Fig. 1). Associations between CRC risk and each of the genotyped and imputed SNPs were evaluated using logistic regression within each study after adjusting for age, sex and the first ten principal components (stage 1) and study site. *P for heterogeneity across studies in GWAS and replication was calculated using Cochran’s Q test.

Of the 64 SNPs evaluated in stage 2, 7 showed association with CRC risk at P < 0.05 with a direction of association consistent with that observed in stage 1 (Table 1 and Supplementary Table 4). In the combined analysis of data from stages 1 and 2, P values for associations with two SNPs (rs647161 at 5q31.1, odds ratio (OR) = 1.17, P = 3.77 × 10−10, and rs10774214 at 12p13.32, OR = 1.17, P = 5.48 × 10−10) were lower than the conventional genome-wide significance level of 5.0 × 10−8, providing convincing evidence for an association of these SNPs with CRC risk (Table 1). An additional SNP, rs2423279, showed a significant association in stage 2 after Bonferroni correction (corrected P < 7.8 × 10−4 ) but did not reach the conventional GWAS significance level for association with CRC risk in the combined analysis of all samples (OR = 1.14, P = 2.29 × 10−7). The association between CRC risk and each of these three SNPs was consistent across most studies (Fig. 1). Results for the other four SNPs that replicated in stage 2 at P < 0.05 (rs1665650, rs2850966, rs1580743 and rs4503064) are also presented (Supplementary Table 4), including one SNP (rs1665650) with an association P value of 8.58 × 10−7 in the combined analysis of all data from both stages (Table 1).

We next evaluated these top four SNPs (Table 1) using data from GWAS in the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and the Colon Cancer Family Registry (CCFR), which together include 11,870 cases and 14,190 controls of European ancestry2,3. Three of the four SNPs were replicated in the GECCO and CCFR sample, although the strength of the associations was weaker than in east Asians (Table 2). These results provide independent support of our findings in the east Asian population. Meta-analyses of data from both east Asian and European-ancestry populations provided strong evidence for associations of CRC risk with these SNPs, with P values all below the genome-wide significance threshold of 5 × 10−8 (Table 2). The weaker associations observed in European-ancestry populations could be explained, in part, by differences in LD patterns at these loci for east Asians.
Asians and Europeans (Supplementary Fig. 4). It is possible that causal variants in these regions are tagged by different SNPs in these two populations or that there is allelic heterogeneity, in which different underlying causal variants exist in populations of Asian and European ancestry. The difference in LD structure between Asian and European descendants and possible allelic heterogeneity in these two populations might explain, in part, why these loci were not discovered in previous studies conducted in individuals of European ancestry. The fourth SNP evaluated in the GECCO and CCFR sample, rs1665650, however, was not replicated in individuals with European ancestry (OR = 0.96, P = 0.05).

Stratification analyses showed that the association of CRC risk with each of the three replicated SNPs was generally consistent in Chinese, Korean and Japanese individuals (Phet > 0.05), although the association with rs2423279 was not statistically significant in Japanese, perhaps owing to a small sample size (Supplementary Table 5). Associations of these three SNPs with CRC risk were similar for men and women (Phet > 0.05) (Supplementary Table 6).

The rs10774214 SNP is located just 15 kb upstream of CCND2, the gene encoding cyclin D2 (Fig. 2a), a member of the D-type cyclin family, which also includes cyclins D1 and D3. These cyclins have a critical role in cell cycle control (from G1 to S phase) through activation of cyclin-dependent kinases (CDKs), primarily CDK4 and CDK6 (ref. 22). CCND2 is closely related to CCND1, a well-established human oncogene22,23. Although CCND2 has been less well studied than CCND1, several studies, including The Cancer Genome Atlas (TCGA), have shown that CCND2 is overexpressed in a substantial proportion of human colorectal tumors22–25. Overexpression of this cyclin may be an independent predictor of survival in individuals with CRC24. Several other genes, including PARP11, FGFR3, FGFR6, C12orf5 and RAD51AP1, are also in close proximity to the SNP identified in our study, of which both C12orf5 (also known as TIGAR, encoding TP53-induced glycolysis and apoptosis regulator) and RAD51AP1 were found to be overexpressed in CRC tissue included in TCGA25. rs10774214 is in strong LD with several SNPs that are located in potential transcription factor–binding sites, as determined using the TRANSFAC database26. Additional research may be warranted regarding possible mechanisms by which this SNP is related to CRC risk.

The rs647161 SNP is located on chromosome 5q31.1, where a cluster of SNPs were associated with CRC risk (Fig. 2b). Of the genes in this region (including PITX1, CATSPER3, PCBD2, MIR4461 and H2AFY), PITX1 is the closest to rs647161 (approximately 129 kb upstream). The PITX1 gene (encoding paired-like homeodomain 1) has been described as a tumor suppressor gene and may be involved in the tumorigenesis of multiple human cancers27–31, including CRC27,32. PITX1 has been reported to suppress tumorigenicity by downregulating the RAS pathway, which is frequently altered in colorectal tumors27. Inhibition of PITX1 induces the RAS pathway and tumorigenicity, and restoring PITX1 in colon cancer cells inhibits tumorigenicity27. It has also been reported that PITX1 may activate TP53 (ref. 33) and regulate telomerase activity34. Consistent with its possible function as a tumor suppressor gene, PITX1 has been found to be downregulated in human cancer tissue samples and cell lines27–30,32. CRC tissue expressing wild-type KRAS showed significantly lower expression of PITX1 than tissue with mutant KRAS32. Most recently, low PITX1 expression was found to be associated with poor survival in individuals with CRC35. In addition, rs6596201, which is in moderate LD with rs647161 (r² = 0.90), is an expression quantitative trait locus (eQTL) (P = 2.42 × 10−28) for the PITX1 gene36. Several other genes at this locus, including C5orf24, H2AFY and NEUROG1, were also found to be highly expressed in colorectal tumors included in TCGA (P < 0.001)32. Additional studies are warranted to explore a possible role for these genes in the etiology of CRC.

Table 2 Association of CRC risk with the top three risk variants in European descendants and east Asian and European descendants combined

<table>
<thead>
<tr>
<th>SNP</th>
<th>Alleles</th>
<th>MAFb</th>
<th>European-ancestry populationsc</th>
<th>East Asian and European-ancestry populations combinedd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cases</td>
<td></td>
<td>cases</td>
<td>Controls</td>
</tr>
<tr>
<td>rs10774214</td>
<td>T/C</td>
<td>0.385</td>
<td>0.379</td>
<td>11,870</td>
</tr>
<tr>
<td>rs647161</td>
<td>A/C</td>
<td>0.680</td>
<td>0.667</td>
<td>11,870</td>
</tr>
<tr>
<td>rs2423279</td>
<td>C/T</td>
<td>0.263</td>
<td>0.252</td>
<td>11,870</td>
</tr>
</tbody>
</table>

*aAlleles (minor/major) for east Asians. bMAF in European-ancestry populations. cSummary statistics were generated using inverse variance–weighted fixed-effects meta-analysis.

Figure 1 Forest plots for the three SNPs showing evidence of an association with CRC risk. Per-allele ORs are presented, with the area of each box proportional to the inverse variance weight of the estimate. Horizontal lines represent 95% confidence intervals.
The rs2423279 SNP is located on chromosome 20p12.3, close to the HAO1 and PLCB1 genes (Fig. 2c). HAO1 encodes hydroxy-acid oxidase, which oxidizes 2-hydroxycacid. PLCB1 encodes phospholipase C-β1, which has an important role in the intracellular transduction of many extracellular signals. Overexpression of the PLCB1 gene has been observed in CRC tissue.22 Possible mechanisms by which these genes are involved in CRC carcinogenesis are unknown. The rs2423279 SNP is 1,408,069 bp downstream of rs961253, a SNP previously identified in a European GWAS as being associated with CRC risk.19 However, these two SNPs are not correlated in east Asians (r² = 0) or in Europeans (r² = 0). Adjustment for rs961253 did not change the results for rs2423279 (data not shown).

To our knowledge, this is the largest GWAS performed for CRC in east Asians, a population that differs from populations of European ancestry in CRC risk and certain aspects of genetic architecture. Results from our study, along with data from a large study conducted in a population of European ancestry, provide convincing evidence of associations with CRC risk for three new independent susceptibility loci at 5q31.1, 12p13.32 and 20p12.3. Results from this study provide new insights into the genetics and biology of CRC.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The authors wish to thank the study participants and research staff for their contributions and commitment to this project. R. Courtney for DNA preparation, J. He for data processing and analyses, and M.J. Daly for clerical support in manuscript preparation. This research was supported in part by US National Institutes of Health (NIH) grants R37CA070867, R01CA082729, R01CA124558, R01CA148667 and R01CA122364, as well as by Ingram Professorship and Research Reward funds from the Vanderbilt University School of Medicine. Participating studies (grant support) in the consortium are as follows: Shanghai Women’s Health Study (US NIH, R01CA082729), Shanghai Breast and Endometrial Cancer Studies (US NIH, R01CA064277 and R01CA092585; contributing only controls), Guangzhou Colorectal Cancer Study (National Key Scientific and Technological Project, 2011ZX09307-001-04, and the National Basic Research Program, 2011CB504303, contributing only controls; the Natural Science Foundation of China, 81072383, contributing only controls), Aichi Colorectal Cancer Study (Grant-in-Aid for Cancer Research, the 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, 17015008 and 221S0001), Korea–National Cancer Center Colorectal Cancer Study (Basic Science Research Program through the National Research Foundation of Korea, 2010-0010276; National Cancer Center Korea, 0910220), Korea-Seoul Colorectal Cancer Study (none reported) and KCPS-II colorectal Cancer Study (National R&D Program for Cancer Control, 0920330; Seoul R&D Program, 10526).

We wish to thank all participants, staff and investigators of GECCO and CCFR for making it possible to present the results in individuals of European ancestry for new CRC susceptibility loci identified in east Asians. Investigators (institution and location) from GECCO and CCFR who provided support for this project include (in alphabetical order) Aaron K. Aragaki (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA), John A. Baron (Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA), Sonja J. Berndt (Division of Cancer Epidemiology and Genetics, National Cancer Institute, US NIH, Bethesda, Maryland, USA), Stéphane Bezieau (Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France), Hermann Brenner, Katja Butterbach (Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany), Bette J. Caan (Division of Research, Kaiser Permanente Medical Care Program, Oakland, California, USA), Christopher S. Carlson (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and School of Public Health,}

Figure 2 Regional plots of association results and recombination rates for the three SNPs showing evidence of association with CRC risk. Genotyped and imputed data from GWAS samples are plotted on the basis of their chromosomal position in NCBI Human Genome Build 36.3. For each region, the SNP selected for stage 2 replication is denoted with a diamond, and the P value from the combined analysis of stage 1 and 2 data is provided. (a-c) Data are shown for rs10774214 (a), rs647161 (b) and rs2423279 (c).


**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2505.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

Genotyping and quality control procedures. Detailed descriptions of genotyping and quality control procedures as well as design of plates and control samples are given in the Supplementary Note. Briefly, in stage 1, 481 cases and 2,632 controls from Shanghai-1 were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 as described previously.14 The average concordance percentage of quality control samples was 99.7%, with a median value of 100% in Shanghai-1 (n = 1,522) and the Korea–National Cancer Center (Korea-NCC) Study (n = 2,721). Japanese participants were from two studies: the Aichi Study 1 (Aichi-1, n = 1,346) and the Aichi Study 2 (Aichi-2, n = 1,507). We also evaluated associations for the top 4 SNPs using data from 11,870 CRC cases and 14,190 controls of European ancestry included in GECCO and CC CFR, which included 14 studies from the United States, Europe, Canada and Australia.4,20,21 Approval was granted from the relevant institutional review boards at all study sites, and all included participants gave informed consent.

Statistical analyses. Dosage data for genotyped and imputed SNPs for participants in each stage 1 study were analyzed using the program mach2dat18 (see URLs). We coded 0, 1 or 2 copies of the effect allele as the dosage for genotyped SNPs, and, for imputed SNPs, we used the expected number of copies of the effect allele as the dosage score. This approach has been shown to give unbiased estimates in meta-analyses.42 Associations between SNPs and CRC risk were assessed using ORs and 95% CIs derived from logistic regression models. ORs were estimated on the basis of the log-additive model and adjusted for age, sex and the first ten principal components. PLINK version 1.07 (see URLs) also was used to analyze genotype data and yielded results virtually identical to those derived from dosage data using mach2dat.18 Meta-analyses were performed using the inverse-variance method, assuming a fixed-effects model, and calculations were implemented in the METAL package19 (see URLs).

Similar to stage 1, we used logistic regression models to derive ORs and 95% CIs for the 64 selected SNPs in stage 2, assuming a log-additive model with adjustment for age and sex. We performed joint analyses to generate summary results for combined samples from all studies, with additional adjustment for study site. We also conducted stratification analysis for the top four SNPs by population ancestry (Chinese, Korean or Japanese) and by sex. We used Cochran's Q statistic to test for heterogeneity44 and the $P$ statistic to quantify heterogeneity across studies as described elsewhere in detail.46 Analyses for stage 2, as well as combined stage 1 and 2 data, were conducted using SAS, version 9.2 (see URLs), with the use of two-tailed tests. $P$ values of $<5\times10^{-8}$ in the combined analysis is considered statistically significant. We used Haploviz version 4.2 (see URLs; ref. 47) to generate a genome-wide Manhattan plot for results from the stage 1 meta-analysis. Forest plots.
and quantile-quantile plots were drawn using R. We drew regional association plots using the website-based tool LocusZoom, version 1.1 (see URLs; ref. 48). LD plots were generated using Haploview47 and the UCSC Genome Browser (see URLs).


