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 × 10⁻⁸ 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 world.¹ Genetic factors have an important role in the etiology of both sporadic and familial CRC.² However, less than 6% of CRC cases can be explained by rare, high-penetrance variants in the CRC susceptibility genes identified to date, such as the APC, SMAD4, AXIN2, BMPR1A, POLD1, STK11, MUTYH and DNA mismatch repair genes.² Over the past two decades, many candidate gene studies have evaluated common genetic risk factors for CRC; only a few of these have been replicated in subsequent studies.³ Recent genome-wide association studies (GWAS) have identified approximately 15 common genetic susceptibility loci for CRC⁴⁻¹². However, these newly identified genetic factors, along with known 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 groups.¹³ 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% and DNA mismatch repair genes in the combined analysis (Supplementary Table 1).

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 $<1.0 \times 10^{-5}$ in controls for each study. Imputation was conducted for each study following the MaCH algorithm \(^{18}\) using phased HapMap 2 Han Chinese in Beijing, China (CHB) and Japanese in Tokyo, 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. €χ^2\text{test.}$

Results for the other four SNPs that showed association in stage 2 after Bonferroni correction (corrected P $< 7.8 \times 10^{-4}$) 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 $\times 10^{-7}$). 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 $\times 10^{-7}$ 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 ancestry\(^ {2,20,21}\). 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 three SNPs, with P values all below the genome-wide significance threshold of 5 $\times 10^{-8}$ (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 compared to east Asians. Meta-analyses of data from both east Asian and European-ancestry populations provided strong evidence for associations of CRC risk with three SNPs, with P values all below the genome-wide significance threshold of 5 $\times 10^{-8}$.

### Table 1 Association of CRC risk with the top four risk variants identified in east Asian samples

| SNP      | Alleles\(^a\) | Chr. | Gene\(^b\) | Location (bp)\(^c\) | Stage | Sample size | MAF   | Sample size | MAF   | OR (95% CI)\(^d\) | P_{\text{rand}} | P_e | $\chi^2$ |
|----------|---------------|------|------------|---------------------|-------|-------------|-------|-------------|-------|-----------------|---------------|-----|---------|
| rs10774214 | T/C           | 12p13.32 | CCND2      | 4238613             | GWAS  | 2,098       | 0.373 | 5,749       | 0.348 | 1.20 (1.09–1.32) | 2.03 $\times 10^{-4}$ | 0.615 | 0%      |
|          |               |       |            |                     | Replication | 5,197     | 0.381 | 5,797       | 0.355 | 1.16 (1.09–1.23) | 5.80 $\times 10^{-7}$ |       |         |
|          |               |       |            |                     | Overall   | 7,295     | 0.379 | 11,546      | 0.352 | 1.17 (1.11–1.23) | 5.48 $\times 10^{-10}$ |       |         |
| rs647161  | A/C           | 5q31.1 | PITX1      | 134526991           | GWAS  | 2,098       | 0.353 | 5,749       | 0.308 | 1.22 (1.12–1.33) | 3.29 $\times 10^{-6}$ | 0.444 | 0%      |
|          |               |       |            |                     | Replication | 5,217     | 0.344 | 5,815       | 0.319 | 1.14 (1.07–1.21) | 1.15 $\times 10^{-5}$ |       |         |
|          |               |       |            |                     | Overall   | 7,315     | 0.347 | 11,564      | 0.313 | 1.17 (1.11–1.22) | 3.77 $\times 10^{-10}$ |       |         |
| rs2423279 | C/T           | 20p12.3 | HAO1       | 7760350             | GWAS  | 2,098       | 0.339 | 5,749       | 0.307 | 1.16 (1.07–1.26) | 4.96 $\times 10^{-4}$ | 0.331 | 12%     |
|          |               |       |            |                     | Replication | 5,227     | 0.315 | 5,811       | 0.297 | 1.13 (1.06–1.19) | 1.22 $\times 10^{-4}$ |       |         |
|          |               |       |            |                     | Overall   | 7,325     | 0.322 | 11,560      | 0.302 | 1.14 (1.08–1.19) | 2.29 $\times 10^{-7}$ |       |         |
| rs1665650 | T/C           | 10q26.12 | HSPA12A    | 118477090           | GWAS  | 2,098       | 0.346 | 5,749       | 0.310 | 1.20 (1.10–1.31) | 3.88 $\times 10^{-5}$ | 0.0018 |         |
|          |               |       |            |                     | Replication | 5,192     | 0.328 | 5,808       | 0.320 | 1.10 (1.04–1.17) | 5.80 $\times 10^{-7}$ |       |         |
|          |               |       |            |                     | Overall   | 7,290     | 0.333 | 11,557      | 0.315 | 1.13 (1.08–1.19) | 8.58 $\times 10^{-7}$ | 0.404 | 4%      |

\(^a\)Minor/major allele for east Asians; \(OR\), odds ratio; \(CI\), confidence interval.

\(^b\)Closest gene.

\(^c\)Location based on NCBI Human Genome Build 36.3.

\(^d\)Adjusted for age, sex, the first ten principal components (stage 1) and study site. €χ^2\text{test.}$
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 for each of the three replicated SNPs was generally consistent in Chinese, Korean and Japanese individuals (P_{het} > 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 (P_{het} > 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 CRC27–31, 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 colorectal cancer cells inhibits tumorigenicity27. It also has 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 rs10774214 (r^2 = 0.25), is an expression quantitative trait locus (eQTL) (P = 2.42 \times 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)25. 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

| SNP       | Alleles | Cases | Controls | Cases | Controls | OR (95% CI) | P_{meta} | Cases | Controls | OR (95% CI) | P_{meta} |
|-----------|---------|-------|----------|-------|----------|-------------|----------|-------|----------|-------------|----------|
| rs10774214| T/C     | 0.385 | 0.379    | 11.870| 14.190   | 1.04 (1.00–1.09) | 0.040   | 19.165| 25.736   | 1.09 (1.06–1.13) | 3.06 \times 10^{-8} |
| rs647161  | A/C     | 0.680 | 0.667    | 11.870| 14.190   | 1.07 (1.02–1.11) | 0.002   | 19.185| 25.754   | 1.11 (1.08–1.15) | 1.22 \times 10^{-10} |
| rs2423279 | C/T     | 0.263 | 0.252    | 11.870| 14.190   | 1.07 (1.03–1.12) | 0.001   | 19.195| 25.750   | 1.10 (1.06–1.14) | 6.64 \times 10^{-9} |

*Alleles (minor/major) for east Asians. MAF in European-ancestry populations. Summary 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 hydroxyacid oxidase, which oxidizes 2-hydroxyacids. 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. 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. However, these two SNPs are not correlated in east Asians ($r^2 = 0$) or in Europeans ($r^2 = 0$). 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, 12p13.22 and 20p12.3. Results from this study provide new insights into the genetics and biology of CRC.

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, 17015018 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), Christopher S. Carlson (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA) and 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, 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, 17015018 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. Investors (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), Christopher S. Carlson (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA) and 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, 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, 17015018 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. Investors (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), Christopher S. Carlson (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA) and 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, 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, 17015018 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. Investors (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), Christopher S. Carlson (Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA).
We also thank B. Buecher of ASTERISK; U. Handte-Daub, M. Celik, R. Hetter-Jensen, U. Benschied and U. Eliber of DACHS; P. Soule, H. Rau, I. Devivo, D. Hunter, Q. Guo, L. Zhu and H. Zhang of HPFS, NHS and PHS; C. Berg and P. Prorok of PCLO; T. Riley of Information Management Services Inc.; B. O’Brien of Westat Inc.; B. Kopp and W. Shao of SAIC-Frederick; investigators from the Women’s Health Initiative (WHI); see URLs) and the GECCO Coordinating Center. Participating studies were supported by: the GECCO and VIVACE meta-analysis are as follows: GECCO (US NIH, U01 CA137088 and R01 CA059045), DALS (US NIH, R01 CA048989), Colo263 (US NIH, R01 CA069097), DACHS (German Federal Ministry of Education and Research, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 1171-1, 01KH004 and 01ER09184), HPS (US NIH, PI U01 CA0058075, U1U1017552, R01 CA137178 and P50 CA127003), MEC (US NIH, R37 CA54281, P01 CA033619 and R01 CA063464), NHS (US NIH, R01 CA137178, P50 CA127003 and P01 CA087969), OFCCR (US NIH, U01 CA074738), PMH (US NIH, R01 CA073636), PHS (US NIH, CA042182), VITAL (US NIH, K05 CA154337), WHI (US NIH, HSNS268201000046C, HSNS268201100001C, HSNS268201100002C, HSNS268201100003C, HSNS268201100004C, HSNS2721011000004C and 268200764316C) and PCLO (US NIH, Z01 CP 011200, U1U1004464 and U01 HG004138). CCFR is supported by the National Cancer Institute NIH, under RFA CA095-95, through cooperative agreements with members of the Colon Cancer Family Registry and principal investigators of the Australasian Colorectal Cancer Family Registry (U01 CA907735), the Familial Colorectal Neoplasia Collaborative Group (U01 CA074799; USC), the Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (U01 CA074800), the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074800). The CWAS work was supported by a National Cancer Institute grant (U01CA122839). OFCCR was supported by a GL2 grant from the Ontario Research Fund, the Canadian Institutes of Health Research and the Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society Research Institute. B.Z. is a recipient of Senior Investigator Awards from the Ontario Institute for Cancer Research, through support from the Ontario Ministry of Economic Development and Innovation. ASTERISK was funded by a Regional Hospital Clinical Research Program (PHRC) and supported by the Regional Council of Pays de la Loire, Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), Association Ané de Bretagne Génétique et Ligue Régionale Contre le Cancer (LRC). PCLO data sets were accessed with approval through dbGaP (Cancer Genetic Markers of Susceptibility (CGEMS) prostate cancer scan, pmid00020711.1 and GWAS of Lung Cancer and Smoking, pmid000093.v2.p2).

AUTHOR CONTRIBUTIONS
W.Z. conceived and directed ACCC as well as the Shanghai-Vanderbilt Colorectal Cancer Genetics Project. W.-H.J., Y.-X.Z., K.M., A.S., Y.-B.X., S.H.J., D.-H.K., U.P.
Genetic predisposition to colorectal cancer.

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.

1. Jermain, A. et al. Global cancer statistics. CA Cancer J. Clin. 61, 69–90 (2011).
2. de la Chapelle, A. Genetic predisposition to colorectal cancer. Nat. Rev. Cancer 4, 769–780 (2004).
3. Dong, L.M. et al. Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. J. Am. Med. Assoc. 299, 2423–2436 (2008).
4. Zarke, B.W. et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat. Genet. 39, 989–994 (2007).
5. Tomilinson, J. et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat. Genet. 39, 984–988 (2007).
6. Broderick, P. et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat. Genet. 39, 1315–1317 (2007).
7. Jaeger, E. et al. Common genetic variants at the CRACI (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat. Genet. 40, 26–28 (2008).
8. Tenesa, A. et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat. Genet. 40, 631–637 (2008).
9. Tomilinson, J.P. et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat. Genet. 40, 623–630 (2008).
10. Houlston, R.S. et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat. Genet. 40, 1426–1435 (2008).
11. Houlston, R.S. 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. 42, 973–977 (2010).
12. Cui, R. et al. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. Gut 60, 799–805 (2011).
13. He, J. et al. Generalizability and epidemiologic characterization of eleven colorectal cancer GWAS hits in multiple populations. Cancer Epidemiol. Biomarkers Prev. 20, 70–81 (2011).
14. Zheng, W. et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat. Genet. 41, 324–328 (2009).
15. Bei, J.X. et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nat. Genet. 42, 599–603 (2010).
16. Jee, S.H. et al. Adiponectin concentrations: a genome-wide association study. Am. J. Hum. Genet. 87, 545–552 (2010).
17. Nakata, I. et al. Association between the SERPING1 gene and age-related macular degeneration and polypoidal choroidal vasculopathy in Japanese. PLoS ONE 6, e19108 (2011).
18. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. 34, 816–834 (2010).
19. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
20. Peters, U. et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum. Genet. 131, 217–234 (2012).
21. Figureiredo, J.C. et al. Genotype-environment interactions in microsatellite stable microsatellite instability-low colorectal cancer: results from a genome-wide association study. Cancer Epidemiol. Biomarkers Prev. 20, 758–766 (2011).
22. Mosgrove, E.A., Caldon, C.E., Barraclough, J., Stone, A. & Sutherland, R.L. Cyclin D as a therapeutic target in cancer. Nat. Rev. Cancer 11, 558–572 (2011).
23. Meremtsishein, A. et al. Expression of D-type cyclins in colon cancer and in cell lines from colon carcinomas. Br. J. Cancer 93, 338–345 (2005).
24. Sarkar, R. et al. Expression of cyclin D2 is an independent predictor of the development of hepatic metastasis in colorectal cancer. Colorectal Dis. 12, 316–323 (2010).
25. Gundem, G. et al. InTGen: integration and data mining of multidimensional oncogenic data. Nat. Methods 7, 92–93 (2010).
26. Matys, V. et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 34, D108–D110 (2006).
27. Kolfschoten, I.G. et al. A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenesis. Cell 121, 849–858 (2005).
28. Chen, Y. et al. Decreased PITX1 homeobox gene expression in human lung cancer. Lung Cancer 55, 287–294 (2007).
29. Chen, Y.N., Chen, H., Xu, Y., Zhang, X. & Luo, Y. Expression of putative homeobox 1 gene in human gastric carcinogenesis and its clinicopathological significance. Int. J. Cancer 120, 775–783 (2007).
30. Lord, R.V. et al. Increased CDX2 and decreased PITX1 homeobox gene expression in Barrett’s esophagus and Barrett’s-associated adenocarcinoma. Surgery 138, 924–931 (2005).
31. Nagel, S. et al. Activation of paired-homeobox gene PITX1 by del(5)(q31) in T-cell acute lymphoblastic leukemia. Leuk. Lymphoma 52, 1348–1359 (2011).
32. Watanabe, T. et al. Differential gene expression signatures between colorectal cancers with and without KRAS mutations: crosstalk between the KRAS pathway and other signalling pathways. Eur. J. Cancer 47, 1946–1954 (2011).
33. Liu, D.X. & Lobie, P.E. Transcriptional activation of p53 by PITX1. Cell Death Differ. 14, 1893–1907 (2007).
34. Qi, D.L. et al. Identification of PITX1 as a TERT suppressor gene located on human chromosome 5. Mol. Cell Biol. 31, 1624–1636 (2011).
35. Knösel, T. et al. Loss of desmocollin 1-3 and homeobox genes PITX1 and CDX2 are associated with tumor progression and survival in colorectal carcinoma. Int. J. Colorectal Dis. 27, 1391–1399 (2012).
36. Zeller, T. et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. PLoS ONE 5, e10693 (2010).
ONLINE METHODS

Study populations. After quality control filtering, 7,456 cases and 11,671 controls from 10 studies were included in the consortium (Supplementary Table 2). Detailed descriptions of participating studies and demographic characteristics of study participants are provided in the Supplementary Note. Briefly, the consortium included 10,730 Chinese participants, 5,544 Korean participants and 2,853 Japanese participants. Chinese participants were from five studies: the Shanghai Study 1 (Shanghai-1, n = 3,102), the Shanghai Study 2 (Shanghai-2, n = 485), the Guangzhou Study 1 (Guangzhou-1, n = 1,613), the Guangzhou Study 2 (Guangzhou-2, n = 2,892) and the Guangzhou Study 3 (Guangzhou-3, n = 2,638). Korean participants were from three studies: the Korean Cancer Prevention Study II (KCPS-II, n = 1,301), the Seoul Study (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 CCFR, which included 14 studies from the United States, Europe, Canada and Australia 20,21. Approval was granted from the relevant institutional review boards at all study sites, and all included participants gave informed consent.

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 (refs. 14, 37, 38). Stage 1 genotyping for 296 cases and 257 controls in Shanghai-2 was performed using Illumina HumanOmniExpress BeadChips. The same method was used to genotype cases from the Guangzhou-1 (n = 694) and Aichi-1 (n = 497) studies in stage 1. The positive quality control samples in these studies had an average concordance percentage of 99.41% and a median value of 99.97%. Cases and controls in KCPS-II were genotyped using the Affymetrix Genome-Wide Human SNP Array 5.0 (ref. 16). Controls for the Guangzhou-1 and Aichi-1 studies were genotyped previously using the Illumina Humana160-Quad BeadChip 19 and Illumina Infinium HumanHap610 BeadChip 19 platforms, respectively. Details of quality control procedures for these samples have been described previously 15–17. We excluded from the analysis samples that were genetically identical or duplicated, had a genotype-determined sex that was inconsistent with self-reported data, had unclear population structure, had close relatives with a PI-HAT estimate greater than 0.25 or had a call rate of <98%. Within each study, SNPs were excluded if (i) MAF was <5%, (ii) the call rate was <95%; (iii) the genotyping concordance percentage was <95% in quality control samples; (iv) the P value for Hardy-Weinberg equilibrium was <1.0 × 10−5 in controls; or (v) there was no LD (r2 < 0.2) with the other SNPs identified in this study; (vi) there was high imputation quality in each of the five studies (RSQ > 0.5); and (vii) P < 0.01 in combined analysis of all stage 1 studies.

Evaluation of population structure. We evaluated population structure in each of the five participating studies included in stage 1 by using principal-components analysis (PCA). Genotyping data for uncorrelated genome-wide SNPs were pooled with data from HapMap to generate the first ten principal components using EIGENSTRAT software 40 (see URLs). The first two principal components for each sample were plotted using R (see URLs). We identified and excluded one participant of KCPS-II who was more than 6 s.d. away from the means of principal components 1 and 2 (Supplementary Fig. 1). The remaining 7,847 samples showed clear east Asian origin, and these samples were included in the final genome-wide association analysis. Cases and controls in each of the five studies were in the same cluster as HapMap Asian samples. The estimated inflation factor λ ranged from 1.02 to 1.04 in these studies after adjusting for age, sex and the first ten principal components, with a λ of 1.01 for combined stage 1 data (Supplementary Fig. 2 and Supplementary Table 1).

Imputation. We used the MacH 1.0 program 46 (see URLs) to impute genotypes for autosomal SNPs that were present in HapMap Phase 2 release 22 separately for each of the five stages included in stage 1. Genotype data from the 90 Asian subjects from HapMap were used as the reference. For Guangzhou-1 and Aichi-1, cases and controls were genotyped using different platforms. To improve imputation quality 46, we identified SNPs for which data were available in both cases and controls (250,612 SNPs in Guangzhou-1 and 232,426 SNPs in Aichi-1) and used them to impute genotyping data. A total of 1,636,380 genotyped SNPs or imputed SNPs with high imputation quality (RSQ > 0.50) in all five studies were tested for association with CRC. To directly evaluate the imputation quality for the top four SNPs identified in our study, we genotyped them in approximately 2,500 samples included in stage 1. The agreement of genotype calls derived from direct genotyping and imputation was very high, with mean concordance rates of 98.05%, 95.61%, 99.84% and 97.90% for rs647161, rs10774214, rs2423279 and rs1665650, respectively (Supplementary Table 7).

Statistical analyses. Dosage data for genotyped and imputed SNPs for participants in each stage 1 study were analyzed using the program mach2dat 18 (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 43 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 package 39 (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 heterogeneity 44 and the I2 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 × 10−8 in the combined analysis was considered statistically significant.

We used Haploview version 4.2 (see URLs; ref. 47) to generate a genome-wide Manhattan plot for results from the stage 1 meta-analysis. Forest plots.

Supplementary Information

doi:10.1038/ng.2505

NATURE GENETICS
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 Haplovlew47 and the UCSC Genome Browser (see URLs).

37. Long, J. et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. PLoS Genet. 8, e1002532 (2012).
38. Shu, X.O. et al. Identification of new genetic risk variants for type 2 diabetes. PLoS Genet. 6, pii: e1001127 (2010).
39. Zheng, W. et al. Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. J. Natl. Cancer Inst. 102, 972–981 (2010).
40. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
41. Sinnott, J.A. & Kraft, P. Artifact due to differential error when cases and controls are imputed from different platforms. Hum. Genet. 131, 111–119 (2012).
42. Jiao, S., Hsu, L., Hutter, C.M. & Peters, U. The use of imputed values in the meta-analysis of genome-wide association studies. Genet. Epidemiol. 35, 597–605 (2011).
43. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
44. Lau, J., Ioannidis, J.P. & Schmid, C.H. Quantitative synthesis in systematic reviews. Ann. Intern. Med. 127, 820–826 (1997).
45. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558 (2002).
46. Zhang, B., Beeghly-Fadiel, A., Long, J. & Zheng, W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Lancet Oncol. 12, 477–488 (2011).
47. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haplovlew: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265 (2005).
48. Pruim, R.J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).