NIH Public Access
Author Manuscript
Int J Cancer. Author manuscript; available in PMC 2015 August 15.

Published in final edited form as:
Int J Cancer. 2014 August 15; 135(4): 948–955. doi:10.1002/ijc.28733.

Genome-wide association study identifies a new *SMAD7* risk variant associated with colorectal cancer risk in East Asians

Ben Zhang[1], **Wei-Hua Jia**[2], **Keitaro Matsuo**[3], **Aesun Shin**[4], **Yong-Bing Xiang**[5], **Koichi Matsuda**[6], **Sun Ha Jee**[7], **Dong-Hyun Kim**[8], **Peh Yean Cheah**[9,10,11], **Zefang Ren**[2], **Qiuyin Cai**[1], **Jirong Long**[1], **Jiajun Shi**[1], **Wanqing Wen**[1], **Gong Yang**[1], **Bu-Tian Ji**[12], **Zhi-Zhong Pan**[2], **Fumihiko Matsuda**[13], **Yu-Tang Gao**[5], **Jae Hwan Oh**[14], **Yoon-Ok Ahn**[15], **Michiaki Kubo**[16], **Lai Fun Thean**[9], **Eun Jung Park**[7], **Hong-Lan Li**[5], **Ji Won Park**[14], **Jaeseong Jo**[7], **Jin-Young Jeong**[8], **Satoyo Hosono**[3], **Yusuke Nakamura**[6], **Xiao-Ou Shu**[1], **Yi-Xin Zeng**[2], and **Wei Zheng**[1]

[1] Division of Epidemiology, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA.

[2] State Key Laboratory of Oncology in South China, Cancer Center, Sun Yat-sen University, Guangzhou, China.

[3] Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan.

[4] Molecular Epidemiology Branch, National Cancer Center, Goyang-si, Korea.

[5] Department of Epidemiology, Shanghai Cancer Institute, Shanghai, P.R. China.

[6] Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.

[7] Institute for Health Promotion, Department of Epidemiology and Health Promotion, Graduate School of Public Health, Yonsei University, Seoul, Korea.

[8] Department of Social and Preventive Medicine, Hallym University College of Medicine, Okcheon-dong, Republic of Korea.

[9] Department of Colorectal Surgery, Singapore General Hospital, Singapore.

[10] Saw Swee Hock School of Public Health, National University of Singapore, Singapore.

[11] Duke-NUS Graduate Medical School, National University of Singapore.

[12] Division of Cancer Epidemiology & Genetics, National Cancer Institute, Bethesda, MD, USA.

[13] Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[14] Center for Colorectal Cancer, National Cancer Center, Goyang-si, Korea.

Corresponding author contact information: Wei Zheng, M.D., Ph.D. Vanderbilt Epidemiology Center and Vanderbilt-Ingram Cancer Center Vanderbilt University School of Medicine 2525 West End Avenue, 8th Floor, Nashville, TN 37203-1738 Phone: (615) 936-0682 Fax: (615) 936-8241 wei.zheng@vanderbilt.edu.

Conflict of interest
We declare that we have no conflicts of interest.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

[15] Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea.

[16] Center for Genomic Medicine, The Institute of Physical and Chemical Research (RIKEN), Kanagawa, Japan.

Abstract

Genome-wide association studies (GWAS) of colorectal cancer (CRC) have been conducted primarily in European descendants. In a GWAS conducted in East Asians, we first analyzed approximately 1.7 million single-nucleotide polymorphisms (SNPs) in four studies with 1,773 CRC cases and 2,642 controls. We then selected 66 promising SNPs for replication and genotyped them in three independent studies with 3,612 cases and 3,523 controls. Five SNPs were further evaluated using data from four additional studies including up to 3,290 cases and 4,339 controls. SNP rs7229639 in the *SMAD7* gene was found to be associated with CRC risk with an odds ratio (95% confidence interval) associated with the minor allele (A) of 1.22 (1.15-1.29) in the combined analysis of all 11 studies ($P = 2.93\times10^{-11}$). SNP rs7229639 is 2,487 bp upstream from rs4939827, a risk variant identified previously in a European-ancestry GWAS in relation to CRC risk. However, these two SNPs are not correlated in East Asians ($r^2 = 0.008$) nor in Europeans ($r^2 = 0.146$). The CRC association with rs7229639 remained statistically significant after adjusting for rs4939827 as well as three additional CRC risk variants (rs58920878, rs12953717, and rs4464148) reported previously in this region. SNPs rs7229639 and rs4939827 explained approximately 1% of the familial relative risk of CRC in East Asians. This study identifies a new CRC risk variant in the *SMAD7* gene, further highlighting the significant role of this gene in the etiology of CRC.

Keywords

Genome-wide association study; GWAS; colorectal cancer; *SMAD7*; genetic susceptibility; single-nucleotide polymorphisms; epidemiology

Introduction

Colorectal cancer (CRC) is the third-most common cancer and second-leading cause of cancer death worldwide (1). While environmental factors are believed to play an important role in the etiology of CRC, it is estimated that approximately 35% of CRC risk may be attributable to inherited factors (2). To date, approximately 19 common genetic susceptibility loci for CRC have been identified in genome-wide association studies (GWAS) (3-13). However, these genetic variants, along with high-penetrance germline mutations in known CRC susceptibility genes, including *APC*, the DNA mismatch repair genes, *SMAD4*, *AXIN2*, *BMPR1A*, *TGFBR2*, *POLD1*, *STK11*, and *MUTYH* (14;15), explain less than 15% of excess familial risk of CRC (3-15).

Most previous GWAS for CRC were conducted in European-ancestry populations. Given the potential difference in genetic architectures between East Asians and European ancestry populations, it is possible that some genetic risk variants for CRC identified in European descendants may not be generalizable to East Asians. Also, GWAS conducted in East Asians could possibly identify genetic risk variants unique to this population. In 2009, we

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

initiated a GWAS in East Asians, the Asia Colorectal Cancer Consortium, and identified three novel genetic susceptibility loci for CRC (16). In this paper, we reported additional findings from this consortium regarding the identification of a new risk variant for CRC in the *SMAD7* gene.

Materials and Methods

Study populations

This study, conducted as part of the Asia Colorectal Cancer Consortium, included 8,891 CRC cases and 10,547 cancer-free controls of East Asian ancestry recruited in eight centers located in China, Korea, Japan, and Singapore (**Table 1**). Specifically, Stage 1 consisted of four studies: Shanghai Study 1 (Shanghai-1, n = 982), Shanghai Study 2 (Shanghai-2, n = 553), Guangzhou Study 1 (Guangzhou-1, n = 1,666), Aichi Study 1 (Aichi-1, n = 1,439). Stage 2 consisted of seven studies: Guangzhou Study 2 (Guangzhou-2, n = 2,892), Korean-National Cancer Center Study (Korea-NCC, n = 2,721), Seoul Study (Korea-Seoul, n = 1,522), Korean Cancer Prevention Study-II (KCPS-II, n = 1,302), the Japan-BioBank Study (Japan-BioBank, n = 3,498), Singapore Chinese Study (SCH, n = 2,000), and Aichi Study 2 (Aichi-2, n = 863). Summary descriptions of these 11 participating studies from eight centers are provided in **Supplementary Methods**. Study protocols were approved by relevant institutional review boards for all study sites.

Laboratory procedures

Genomic DNA was extracted from either blood or saliva samples according to standard protocols. In Stage 1, genotyping was performed using Affymetrix Genome-Wide Human SNP Array 6.0 (Affy 6.0, 906,602 SNPs) for Shanghai-1 cases and controls; Illumina HumanOmniExpress BeadChip (Illumina OmniExpress, 729,462 SNPs) for Shanghai-2 cases and controls, Guangzhou-1 cases and Aichi-1 cases; Illumina Human610-Quad BeadChip (620,901 SNPs) for Guangzhou-1 controls; and Illumina Infinium HumanHap610 BeadChip (592,044 SNPs) for Aichi-1 controls. Genotype calling was performed using Birdseed algorithm for Affymetrix 6.0 or GenomeStudio software for Illumina GWAS platforms based on manufacturer's protocols.

Quality control (QC) protocols were applied to exclude samples and SNPs from all four studies in Stage 1 as described previously (17-19), including: 1) genotype call rate per sample < 95%, 2) genetically identical (PI_HAT > 0.9) or duplicated samples, 3) genetic sex inconsistent with survey/clinical data, 4) samples with close relative (PI_HAT > 0.25), 5) population structure inconsistent with HapMap Asians (see **Statistical analysis**), 6) genotype call rate per SNP < 95%, 7) minor allele frequency (MAF) < 0.05, 8) genotyping concordance < 95% in QC samples, 9) Hardy-Weinberg equilibrium (HWE) $P < 1\times10^{-5}$ in controls, or 10) SNPs not in autosomes. After these QC procedures, 580,086 SNPs for 971 individuals (474 cases and 497 controls) remained in the Shanghai-1 dataset; 515,701 SNPs for 485 individuals (254 cases and 231 controls) remained in the Shanghai-2 dataset; 522,096 SNPs for 641 cases and 435,925 SNPs for 972 controls remained in the Guangzhou-1 dataset; and 478,246 SNPs for cases and 443,065 SNPs for controls remained in the Aichi-1 dataset.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Stage 2 genotyping was performed using the Sequenom MassARRAY platform (Sequenom, San Diego, CA, USA) for the promising 66 SNPs selected from Stage 1. These SNPs were genotyped in Guangzhou-2, Korea-NCC, and Korea-Seoul studies. Again in Stage 2, standard QC protocols were applied to exclude SNPs, including: 1) genotype call rate per SNP < 95%, 2) unclear genotyping cluster, 3) genotyping concordance < 95% in QC samples, 4) HWE $P < 7.7\times10^{-4}$ (0.05/65) in controls. After these QC procedures, the number of eligible SNPs was 65 in Guangzhou-2, 64 in Korea-NCC, and 60 in Korea-Seoul studies. Five of these SNPs were taken forward for *in silico* replication in KCPS-II (n = 5), Japan-BioBank (n = 5), and SCH (n = 2; rs7229639 and rs2143619). Genotyping was conducted using either Affymetrix Genome-Wide Human SNP Array 5.0 in KCPS-II or Illumina HumanHap 610K and 550K in Japan-BioBank or Affymetrix Genome-Wide Human SNP Array 6.0 for SCH. Details of the QC procedures and data processing for samples included in these three studies have been previously reported in elsewhere (11;20;21). Finally, SNP rs7229639 was further genotyped in Aichi-2 using Sequenom along with SNPs for other projects.

Statistical analysis

Genome-wide imputation for samples in four Stage 1 studies was performed using program MACH 1.0 (22) based on data from the 90 CHB (Han Chinese in Beijing, China) and JPT (Japanese in Tokyo, Japan) samples included in the HapMap project (release 22). After exclusion of imputed SNPs with MAF < 0.05 and RSQ < 0.50, 1,695,815 genotyped or imputed SNPs remained for meta-analyses.

To evaluate the population structure and identify potential genetic outliers, we performed principal components analysis (PCA) using EIGENSTRAT, version 2.5 (23). We selected a set of ~6,000 uncorrelated SNPs (closest distance between two SNPs > 200kb, MAF > 0.2, $r^2 < 0.1$, and call rate > 99%) shared among all 4,415 samples included in Stage 1 and the HapMap Project using PLINK version 1.07 (24). Genotype data of these SNPs from the four Stage 1 studies were pooled together with HapMap data (release 23a) to generate the first ten principal components. Samples were removed from the final analysis if they were more than 6σ away from the means of PC1 and PC2.

Associations of SNPs with CRC risk in each of the four studies included in Stage 1 were assessed by assuming a log-additive effect of the allelic dosage of the SNPs. Odds ratios (ORs) and 95% confidence intervals (CIs) were generated from logistic regression models, adjusted for age, sex, and the first ten principal components. We coded 0, 1, or 2 copies of the effect alleles as dosage for genotyped SNPs and used the expected number of copies of the effect alleles as dosage score for imputed SNPs to account for imputation uncertainty (25). The meta-analysis was performed using an inverse-variance method assuming fixed-effects, with a Cochran's Q statistic to test for heterogeneity (26) and I^2 statistic to quantify heterogeneity (27) across studies. Summary statistics of genome-wide meta-analyses were generated using the METAL program (28). Similar to Stage 1, we evaluated associations of CRC risk with SNPs in each of the studies included in Stage 2 using logistic regression models with adjustment for age and sex. Summary estimates in Stage 2, all studies combined, and subgroups by populations (Chinese, Korean, and Japanese) and sex (male,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

female) were also obtained using a fixed-effects meta-analysis with METAL. SNPs showing an association at $P < 5\times10^{-8}$ in the combined analysis of all studies were considered genome-wide significant. We conducted haplotype association analysis for two SNPs in 18q21.1 using SAS Genetics v9.3 with logistic regression models. The familial relative risk (λ) to offspring of an affected individual due to a single locus is estimated using formula: $\lambda = (pr^2 + q)/(pr + q)^2$, where p is the frequency of the risk allele, $q = 1\text{-}p$ is the frequency of the reference allele, and r is the per-allele relative risk (29). The proportion of the familial relative risk explained by a locus, assuming a multiplicative interaction between markers in the locus and other loci, is calculated as $ln\left(\prod_i \lambda_i\right)/ln\left(\lambda_0\right)$, where λ_o is the overall familial relative risk which is assigned to be 2.2 for CRC estimated from a previous meta-analysis (30). Assuming that the risks associated with each locus combine multiplicatively, the combined contribution of the familial relative risks from multiple loci is equal to: $\left(\prod_i \lambda_i\right)/ln\left(\lambda_0\right)$.

To visualize population substructure, we drew a PCA plot using data from the 4,415 Stage 1 samples and 270 subjects from HapMap based on the first two principal components using R version 2.13.0 (http://www.r-project.org/). We also used R package to generate a forest plot to display the association of rs7229639 with CRC risk across studies. We generated regional association plots using the website-based software LocusZoom, version 1.1 (31). Haploview version 4.2 (32) was used to infer linkage disequilibrium (LD) structure.

Results

A total of 19,179 samples are included in the current analysis (**Table 1**). Cases and controls were reasonably well matched by age and sex in most of the participating studies. All samples in Stage 1 showed a clear East Asian origin and none of them were more than 6σ away from the means of PC1 and PC2 (**Supplementary Figure 1**). Cases and controls in each of the four studies were in the same cluster compared with East Asians in HapMap. After standard QC filter, a total of 1,695,815 SNPs were finally included in the association analyses. Using Stage 1 data, we evaluated association of CRC risk with the 26 previously reported SNPs. Of the 22 SNPs initially identified from GWAS conducted in European-ancestry populations, rs6691170 and rs16892766 are monomorphic in East Asians. One SNP, rs5934683 in Chromosome X, was excluded from analyses in this study. All other 19 SNPs showed an association with CRC risk in the same direction as reported initially (**Supplementary Table 1**). Nine of 19 SNPs showed a statistically significant association with CRC risk at $P < 0.05$. Of the four SNPs initially identified in East Asians, three (rs647161, rs10774214 and rs2423279) were also significantly associated with CRC in our Stage 1, and one (rs7758229) was not associated with CRC risk in these data. To identify new genetic risk variants for CRC, we selected the 66 most promising SNPs for replication in Stage 2 using the following criteria: 1) MAF > 0.05 in each of the four Stage 1 studies, 2) RSQ > 0.70 in all four studies, 3) $P_{meta} < 5.5 \times 10^{-5}$, 4) no heterogeneity ($P_{heterogeneity} > 0.05$ and $I^2 < 25\%$), 5) uncorrelated with SNPs in known CRC loci or with each other ($r^2 < 0.20$), and 6) data available in all four Stage 1 studies. These 66 SNPs were not evaluated in our previously published study (16).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Of the 66 SNPs selected for Stage 2 replication, 59 SNPs were successfully genotyped in 3,612 cases and 3,523 controls from three studies (Guangzhou-2, Korea-NCC, and Korea-Seoul) included in Stage 2 (**Supplementary Table 2**). Five SNPs (rs7923556, rs1539213, rs7229639, rs7247381, and rs2143619) from five different regions (10q21.2, 14q21.3, 18q21.1, 19q12, and 20p12.2) showed an association with CRC risk at *P*-value < 0.05 in the same direction as observed in Stage 1 (**Supplementary Table 2**). For these five SNPs, we conducted *in silico* replication using data from three additional studies (KCPS-II, Japan-BioBank, and SCH) with 2,899 cases and 3,867 controls. SNP rs7229639 was further genotyped in Aichi-2 including 391 cases and 472 controls. Of all the 59 SNPs evaluated in Stage 2, only rs7229639 was consistently associated with CRC risk across all seven studies, showing strong evidence of replication, with *P*-value 3.39×10^{-8}. Joint analysis of samples in Stages 1 and 2 yielded per-allele OR (95% CI) 1.22 (1.15-1.29) and *P*-value 2.93×10^{-11} (**Table 2**), which is substantially lower than the genome-wide significance level of 5×10^{-8}. This association was consistent across all eleven studies in Stages 1 and 2 (**Figure 1**), with little evidence of between-study heterogeneity (*P* for heterogeneity = 0.726, I^2 = 0%). Stratification analysis did not reveal any apparent heterogeneity across Chinese (OR = 1.20), Korean (OR = 1.21), or Japanese (OR = 1.23) subjects (*P* for heterogeneity = 0.958), or between men (OR = 1.28) and women (OR = 1.19) (*P* for heterogeneity = 0.268) for this SNP (**Table 2**). Finally, using genotype data from four studies included in Stage 2, we found that the risk of CRC was increased in a dose-response manner with the number of minor allele (A) of rs7229639 (*P* for trend = 4.78×10^{-7}), with ORs of 1.21 (95% CI: 1.10-1.34) and 1.62 (95% CI: 1.26-2.08) for heterozygotes and homozygotes, respectively. These data support an additive model rather than dominant (OR = 1.25, 95% CI: 1.14-1.38; $P = 4.39 \times 10^{-6}$) or recessive model (OR= 1.52, 95% CI: 1.19-1.96; $P = 9.95 \times 10^{-4}$).

SNP rs7229639 is located in intron 3 of the *SMAD7* gene at 18q21.1 (**Supplementary Figure 2**), where three other SNPs (rs4939827, rs12953717, and rs4464148) have been reported in previous GWAS conducted in European descendants to be associated with CRC risk (5). In a fine-mapping analysis, rs58920878 was identified as a potential causal variant in this region (33). SNP rs7229639 is not correlated with any of four previously reported risk variants in this region in East Asians with r^2 all under 0.05 (data from the 1000 Genomes Project). In European descendants, rs7229639 was weakly correlated with r4939827 (r^2 = 0.146) and not correlated with any of the four other SNPs (r^2 < 0.07). Data for these four previously reported SNPs were available in Stage 1, and three of the SNPs showed statistically significant (rs4939827) or marginally significant (rs58920878 and rs12953717) association with CRC risk in our Stage 1 (**Table 3**, **Supplementary Table 1**). Of these previously reported risk variants, rs4939827 showed the strongest association with CRC risk (OR =0.89, 95% CI, 0.80-0.98, *P* = 0.022) in Stage 1 of our study. Because rs4939827 is in LD with both rs58920878 and rs12953717 in Asians (r^2 > 0.8), we selected rs4939827 for additional genotyping in Stage 2 (*P* = 0.008) (data not shown). A combined analysis of samples from both Stages 1 and 2 yielded a per-allele OR (95% CI) of 0.90 (0.85-0.96) (*P* = 0.001) (data not shown).

Conditional analyses were performed to determine whether the observed association with rs7229639 was independent of the other GWAS-identified SNPs in this region (**Table 3**).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The association with rs7229639 remained statistically significant after adjusting for any of these previously GWAS-identified SNPs individually or in combination. Interestingly, adjusting for rs7229639 strengthened the associations of CRC risk with three (rs4939827, rs58920878, and rs12953717) of the four SNPs reported from previous GWAS. Because these three previously reported SNPs were strongly correlated, we selected rs4939827 for further evaluation with rs7229639 in haplotype analysis. Haplotype analysis of rs7229639 and rs4939827 revealed two common haplotypes, A-C and G-T, to be statistically significantly associated with CRC risk (**Table 4**). For haplotype A-C, the strength of association with CRC (OR = 1.68, 95% CI, 1.37-2.07; $P = 9.92 \times 10^{-7}$) was greater than that of either rs7229639 (OR = 1.22, 95% CI, 1.11-1.33; $P = 3.43 \times 10^{-5}$) or rs4939827 (OR = 0.90, 95% CI, 0.83-0.97; $P = 0.007$) alone. When dominant model was applied, three minor haplotypes were associated with a 1.49-fold (95% CI, 1.29-1.72) increased risk of CRC ($P = 5.01 \times 10^{-8}$).

Discussion

In this two-stage GWAS of CRC including 8,675 cases and 10,504 controls from China, Korea, Japan, and Singapore, we identified a new genetic variant (rs7229639) in the *SMAD7* gene to be associated with CRC risk at genome-wide significance level. The association of this SNP with CRC risk remained highly statistically significant after adjusting for all four other risk variants reported previously in this region (5;33), providing evidence for the presence of multiple genetic risk variants in the *SMAD7* gene for CRC. SNP rs7229639 explained approximately 0.75% of the familial relative risk of CRC in East Asians, while rs4939827, the risk variant identified originally in GWAS, explained about 0.3% of familial relative risk. When haplotypes of these two variants are considered, these variants would explain 1.9% of familial relative risk, approximately six-fold the 0.3% explained by the original risk variant (rs4939827).

SNP rs7229639 (46,450,976 bp, NCBI Human Genome Build 37.3) is located in intron 3 of the *SMAD7* gene at chromosome 18q21.1. In this region, three other variants including rs4939827 (46,453,463 bp), rs12953717 (46,453,929 bp), and rs4464148 (46,459,032 bp) have been found to be associated with CRC risk through GWAS conducted in European-ancestry populations (5). A subsequent fine-mapping through resequencing 17 kb region of 18q21.1 in a study conducted in European descendants identified a new variant, rs58920878 (46,449,565 bp) as a potential causal variant at this locus (33). SNP rs58920878 showed the strongest association in that study and is correlated with rs4939827 (r^2=0.533), rs12953717 (r^2=0.927), and rs4464148 (r^2=0.327) in European descendants. Although rs58920878 is also in high LD with both rs4939827 (r^2=0.830) and rs12953717 (r^2=0.785) in East Asians, it was only weakly associated with CRC risk in our study. The SNP (rs7229639) we identified in this study showed the strongest association with CRC risk and is not correlated with any of the four previously reported SNPs in both East Asian and European populations. Therefore, as expected, adjustment for any or all of these four SNPs did not attenuate the association of rs7229639 with CRC risk in the current study. Haplotype analyses further confirmed the independent association of rs7229639 and rs4939827 with CRC risk in this locus. It is possible that multiple causal variants may be present in this locus among East Asians.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The *SMAD7* gene is a key member in TGF-β family signaling pathway which has been shown to play a dual role in carcinogenesis, including tumor suppressor in early stage and oncogene in advanced stage of cancers (34). Multiple genes in this pathway are known to be involved in colorectal pathogenesis (35). *SMAD7* gene encodes Smad family member 7 (Smad7), an inhibitory protein that functions as an antagonist of TGF-β signaling by blocking phosphorylation of receptor-activated Smads or by competitive inhibition of complex formation of receptor-activated Smads with the common-mediator Smad4 in the cytoplasm and nucleus (36). Smad7 also serves as an important cross-talk mediator of the TGF-β signaling pathway with other signaling pathways including Wnt signaling (36). Expression of Smad7 was found in both normal colon mucosa and tumor cells (37), and aberrant Smad7 expression may influence CRC progression (38). Recently, Smad7 was shown to induce colorectal tumorigenicity through blocking TGF-beta-induced growth inhibition and inhibiting apoptosis, and a certain proportion of human colorectal tumors may become refractory to tumor suppressive actions of TGF-β that might lead to increased tumorigenicity (39).

A recent study has showed that the previously-identified CRC risk variant rs4939827 may be associated with CRC survival (40). This SNP was associated with certain characteristics of CRC, including invasiveness of the cancer and *RUNX3* methylation status (41). No survival data, however, were collected in our study, and thus we could not evaluate these *SMAD7* variants, including the newly-identified risk variant rs7229639, with CRC survival in our study.

In summary, this study identified a new CRC risk variant in the *SMAD7* gene among East Asians, which further highlights the significant role of this gene in the etiology of CRC. To date, multiple CRC susceptibility loci have been identified by GWAS in TGFβ pathway genes, including *SMAD7*, supporting an important role of this pathway in the pathogenesis of CRC. Future studies are warranted to investigate these CRC susceptibility loci to identify causal variants underlying the associations discovered in GWAS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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, Regina Courtney for DNA preparation, Jing He for data processing and analyses, and Mary Jo Daly for clerical support in manuscript preparation. This research was supported in part by U.S. National Institutes of Health grants R37CA070867, R01CA082729, R01CA124558, R01CA148667, and R01CA122364, as well as 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 (R37CA070867), Shanghai Men's Health Study (R01CA082729), Shanghai Breast and Endometrial Cancer Studies (R01CA064277 and R01CA092585, contributing only controls), Guangzhou Colorectal Cancer Study (National Key Scientific and Technological Project – 2011ZX09307-001-04; 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, Nos. 17015018 and 221S0001), Korea-NCC Colorectal Cancer Study (Basic Science Research Program through the National Research Foundation of Korea,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2010-0010276 and National Cancer Center Korea, 0910220), Japan-BioBank Colorectal Cancer Study (Grant from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government), KCPS-II colorectal cancer study (National R&D Program for cancer control, 0920330; Seoul R&D Program, 10526), Korea-Seoul Colorectal Cancer Study (None reported), and Singapore Chinese Study (NMRC/1193/2008).

Abbreviations

CHB	Han Chinese in Beijing, China
CI	confidence intervals
CRC	colorectal cancer
EAF	effect allele frequency
GWAS	genome-wide association study
HERPACC-II	Hospital-based Epidemiologic Research Program at Aichi Cancer Center
HWE	Hardy-Weinberg equilibrium
JPT	Japanese in Tokyo, Japan
KCPS-II	Korean Cancer Prevention Study-II
LD	linkage disequilibrium
MAF	minor allele frequency
NCC	National Cancer Center
OR	odds ratio
PCA	principal components analysis
QC	quality control
SCH	Singapore Chinese Study
SMHS	Shanghai Men's Health Study
SNP	single-nucleotide polymorphism
SWHS	Shanghai Women's Health Study

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011; 61:69–90. [PubMed: 21296855]
2. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000; 343:78–85. 13. [PubMed: 10891514]
3. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet. 2007; 39:989–94. [PubMed: 17618283]
4. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet. 2007; 39:984–8. [PubMed: 17618284]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

5. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet. 2007; 39:1315–7. [PubMed: 17934461]
6. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, Walther A, Spain S, Pittman A, Kemp Z, Sullivan K, Heinimann K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet. 2008; 40:26–8. [PubMed: 18084292]
7. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet. 2008; 40:631–7. [PubMed: 18372901]
8. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K, Jaeger E, Fielding S, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet. 2008; 40:623–30. [PubMed: 18372905]
9. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, Carvajal-Carmona L, Howarth K, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 2008; 40:1426–35. [PubMed: 19011631]
10. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, Spain SL, Broderick P, Domingo E, Farrington S, Prendergast JG, Pittman AM, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet. 2010; 42:973–7. [PubMed: 20972440]
11. Cui R, Okada Y, Jang SG, Ku JL, Park JG, Kamatani Y, Hosono N, Tsunoda T, Kumar V, Tanikawa C, Kamatani N, Yamada R, et al. Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. Gut. 2011; 60:799–805. [PubMed: 21242260]
12. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, Tenesa A, Spain S, Broderick P, Ooi LY, Domingo E, Smillie C, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet. 2012; 44:770–6. [PubMed: 22634755]
13. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, Berndt SI, Bezieau S, Brenner H, Butterbach K, Caan BJ, Campbell PT, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013; 144:799–807. [PubMed: 23266556]
14. de la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer. 2004; 4:769–80. [PubMed: 15510158]
15. Ma X, Zhang B, Zheng W. Genetic variants associated with colorectal cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Gut. 2013
16. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, Kim DH, Ren Z, Cai Q, Long J, Shi J, Wen W, et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet. 2013; 45:191–6. [PubMed: 23263487]
17. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009; 41:324–8. [PubMed: 19219042]
18. Bei JX, Li Y, Jia WH, Feng BJ, Zhou G, Chen LZ, Feng QS, Low HQ, Zhang H, He F, Tai ES, Kang T, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. Nat Genet. 2010; 42:599–603. [PubMed: 20512145]
19. Nakata I, Yamashiro K, Yamada R, Gotoh N, Nakanishi H, Hayashi H, Tsujikawa A, Otani A, Saito M, Iida T, Oishi A, Matsuo K, et al. Association between the SERPING1 gene and age-related macular degeneration and polypoidal choroidal vasculopathy in Japanese. PLoS One. 2011; 6:e19108. [PubMed: 21526158]
20. Jee SH, Sull JW, Lee JE, Shin C, Park J, Kimm H, Cho EY, Shin ES, Yun JE, Park JW, Kim SY, Lee SJ, et al. Adiponectin concentrations: a genome-wide association study. Am J Hum Genet. 2010; 87:545–52. [PubMed: 20887962]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

21. Thean LF, Li HH, Teo YY, Koh WP, Yuan JM, Teoh ML, Koh PK, Tang CL, Cheah PY. Association of caucasian-identified variants with colorectal cancer risk in singapore chinese. PLoS One. 2012; 7:e42407. [PubMed: 22879968]
22. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–34. [PubMed: 21058334]
23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]
25. Zheng J, Li Y, Abecasis GR, Scheet P. A comparison of approaches to account for uncertainty in analysis of imputed genotypes. Genet Epidemiol. 2011; 35:102–10. [PubMed: 21254217]
26. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med. 1997; 127:820–6. [PubMed: 9382404]
27. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002; 21:1539–58. [PubMed: 12111919]
28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]
29. Zheng W, Zhang B, Cai Q, Sung H, Michailidou K, Shi J, Choi JY, Long J, Dennis J, Humphreys MK, Wang Q, Lu W, et al. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. Hum Mol Genet. 2013; 22:2539–50. [PubMed: 23535825]
30. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol. 200196:2992–3003. [PubMed: 11693338]
31. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–7. [PubMed: 20634204]
32. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005; 21:263–5. [PubMed: 15297300]
33. Pittman AM, Naranjo S, Webb E, Broderick P, Lips EH, van WT, Morreau H, Sullivan K, Fielding S, Twiss P, Vijayakrishnan J, Casares F, et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. Genome Res. 2009; 19:987–93. [PubMed: 19395656]
34. Blobe GC, Schiemann WP, Lodish HF. Role of transforming growth factor beta in human disease. N Engl J Med. 2000; 342:1350–8. [PubMed: 10793168]
35. Markowitz SD, Bertagnolli MM. Molecular origins of cancer: Molecular basis of colorectal cancer. N Engl J Med. 2009; 361:2449–60. [PubMed: 20018966]
36. Yan X, Chen YG. Smad7: not only a regulator, but also a cross-talk mediator of TGF-beta signalling. Biochem J. 2011; 434:1–10. [PubMed: 21269274]
37. Korchynskyi O, Landstrom M, Stoika R, Funa K, Heldin CH, ten DP, Souchelnytskyi S. Expression of Smad proteins in human colorectal cancer. Int J Cancer. 1999; 82:197–202. [PubMed: 10389752]
38. Levy L, Hill CS. Alterations in components of the TGF-beta superfamily signaling pathways in human cancer. Cytokine Growth Factor Rev. 2006; 17:41–58. [PubMed: 16310402]
39. Halder SK, Beauchamp RD, Datta PK. Smad7 induces tumorigenicity by blocking TGF-beta-induced growth inhibition and apoptosis. Exp Cell Res. 2005; 307:231–46. [PubMed: 15922743]
40. Phipps AI, Newcomb PA, Garcia-Albeniz X, Hutter CM, White E, Fuchs CS, Hazra A, Ogino S, Nan H, Ma J, Campbell PT, Figueiredo JC, et al. Association between colorectal cancer susceptibility Loci and survival time after diagnosis with colorectal cancer. Gastroenterology. 2012; 143:51–4. [PubMed: 22580541]
41. Garcia-Albeniz X, Nan H, Valeri L, Morikawa T, Kuchiba A, Phipps AI, Hutter CM, Peters U, Newcomb PA, Fuchs CS, Giovannucci EL, Ogino S, et al. Phenotypic and tumor molecular

characterization of colorectal cancer in relation to a susceptibility SMAD7 variant associated with survival. Carcinogenesis. 2013; 34:292–8. [PubMed: 23104301]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

What's new?

GWAS, conducted primarily in European ancestry populations, have identified common genetic variants in approximately 20 loci associated with CRC risk. These variants account for approximately 10% of familial risk of this common cancer. In a large GWAS of CRC including approximately 19,000 individuals of East Asian ancestry, SNP rs7229639 in the *SMAD7* gene was found to be associated with CRC risk at genome-wide significance level ($P<5\times10^{-8}$), independent of rs4939827 reported previously in this region by an European GWAS. Findings from this study highlight the significant role of this gene in the etiology of CRC and suggest that genetic risk variants may differ by study populations.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

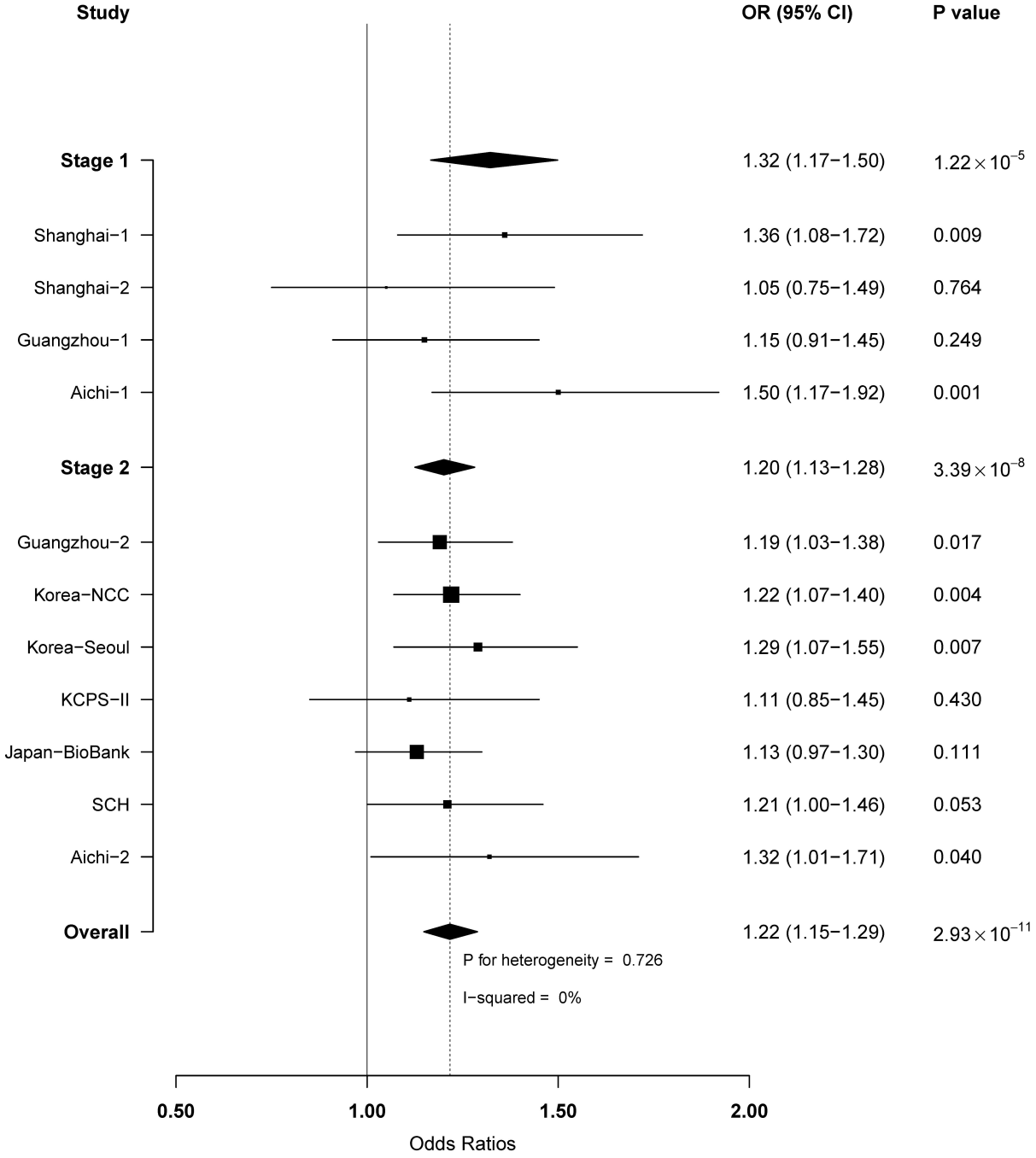

Figure 1. Association of rs7229639 with colorectal cancer risk by studies

Per-allele ORs are presented with the area of the box proportion to the inverse variance weight of the estimate. Horizontal lines show 95% CIs.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Descriptions of participating studies and subjects included in this analysis

		Sample size[1]		Mean age (years)[2]		Female (%)[2]		
Study	**Population**	**Genotyped**	**After QC**	**Cases**	**Controls**	**Cases**	**Controls**	**Genotyping platform**
Stage 1								
Shanghai-1	Chinese	481/501	474/497	60.02	60.20	73.84	72.64	Affymetrix 6.0
Shanghai-2	Chinese	296/257	254/231	61.16	60.75	54.72	56.71	Illumina OmniExpress
Guangzhou-1	Chinese	694/972	641/972	54.86	47.40	36.51	27.06	Illumina OmniExpress/Human610-Quad
Aichi-1	Japanese	497/942	404/942	59.43	47.88	37.38	47.77	Illumina OmniExpress/HumanHap610
Subtotal		1,968/2,672	1,773/2,642					
Stage 2								
Guangzhou-2[3]	Chinese	1,371/1,521	1,371/1,521	58.22	54.64	37.96	39.25	Sequenom
Korea-NCC[3]	Korean	1,392/1,329	1,392/1,329	58.19	55.59	37.64	38.60	Sequenom
Korea-Seoul[3]	Korean	849/673	849/673	59.05	57.19	40.99	47.85	Sequenom
KCPS-II[4]	Korean	325/977	325/976	51.38	41.27	27.08	43.34	Affymetrix 5.0
Japan-BioBank[4]	Japanese	1,595/1,903	1,583/1,898	58.37	52.49	39.17	36.22	Illumina HumanHap 610K/550K
SCH[4]	Chinese	1,000/1,000	991/993	69.99	69.40	42.99	43.00	Affymetrix 6.0
Aichi-2[3]	Japanese	391/472	391/472	59.87	60.08	36.06	35.17	Sequenom
Subtotal		6,923/7,875	6,902/7,862					
Total		8,891/10,547	8,675/10,504					

[1] Number of cases/controls. QC, quality control.

[2] Among samples remained after QC exclusion.

[3] Studies for direct genotyping.

[4] Studies for in silico replication.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Association of rs7229639 in the *SMAD7* gene with colorectal cancer risk in East Asians

	Sample size[1]		EAF[2]		Per-allele association		Heterogeneity	
Study	**Cases**	**Controls**	**Cases**	**Controls**	**OR (95% CI)[3]**	***P* value**	$P_{heterogeneity}$[4]	I^2 (%)
Stage 1	1,773	2,642	0.186	0.152	1.32 (1.17-1.50)	1.22×10^{-5}	0.269	24
Stage 2	6,800	7,761	0.176	0.153	1.20 (1.13-1.28)	3.39×10^{-8}	0.897	0
Overall	8,573	10,403	0.178	0.152	1.22 (1.15-1.29)	2.93×10^{-11}	0.726	0
Chinese	3,670	4,204	0.170	0.145	1.20 (1.10-1.32)	6.65×10^{-5}	0.958	0
Korean	2,525	2,889	0.215	0.185	1.21 (1.09-1.33)	1.85×10^{-4}		
Japanese	2,378	3,310	0.151	0.133	1.23 (1.10-1.38)	3.21×10^{-4}		
Male	3,542	4,338	0.190	0.156	1.28 (1.17-1.40)	2.88×10^{-8}	0.268	19
Female	2,456	3,175	0.200	0.174	1.19 (1.07-1.32)	8.13×10^{-4}		

[1] Data from Japan-BioBank and SCH studies were not included in the stratified analysis by sex.

[2] Effect allele (A) frequency of cases and controls.

[3] OR was estimated based on the effect allele (A).

[4] *P* for heterogeneity across studies was calculated using a Cochran's *Q* test

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Evaluation of independent association of rs7229639 and other four risk variants in the *SMAD7* gene in relation to colorectal cancer risk

Test SNP (allele)	Adjusted SNP(s)	Cases/controls	EAF[1]	OR (95% CI)	*P* value
rs58920878 (C)	None	1,773/2,642	0.734	0.92 (0.82-1.02)	0.097
rs7229639 (A)	None	4,840/5,925	0.156	1.25 (1.16-1.34)	5.25×10^{-9}
rs4939827 (C)	None	4,840/5,925	0.727	0.90 (0.85-0.96)	8.87×10^{-4}
rs12953717 (C)	None	1,773/2,642	0.733	0.92 (0.83-1.01)	0.085
rs4464148 (C)	None	1,773/2,642	0.059	0.97 (0.80-1.18)	0.770
rs7229639 (A)	rs58920878	1,773/2,642	0.152	1.40 (1.23-1.60)	4.19×10^{-7}
rs7229639 (A)	rs4939827	4,840/5,925	0.156	1.30 (1.20-1.40)	1.87×10^{-11}
rs7229639 (A)	rs12953717	1,773/2,642	0.152	1.40 (1.23-1.60)	3.97×10^{-7}
rs7229639 (A)	rs4464148	1,773/2,642	0.152	1.32 (1.17-1.50)	1.25×10^{-5}
rs7229639 (A)	All four SNPs[2]	1,773/2,642	0.152	1.41 (1.23-1.61)	8.96×10^{-7}
rs58920878 (C)	rs7229639	1,773/2,642	0.734	0.84 (0.76-0.94)	0.002
rs4939827 (C)	rs7229639	4,840/5,925	0.727	0.86 (0.81-0.92)	2.57×10^{-6}
rs12953717 (C)	rs7229639	1,773/2,642	0.733	0.85 (0.76-0.94)	0.002
rs4464148 (C)	rs7229639	1,773/2,642	0.059	1.02 (0.84-1.24)	0.854

Abbreviations: EAF, effect allele frequency; OR, odds ratio; and CI, confidence interval.

[1] Effect allele frequency of the tested SNP in controls.

[2] rs58920878, rs4939827, rs12953717 and rs4464148.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

Association of colorectal cancer risk with the haplotypes comprising rs7229639 and rs4939827

Haplotype[1]	Frequency[2]		OR (95% CI)[3]	*P* value
	Cases	Controls		
G-C	0.537	0.582	1.00 (Reference)	
G-T	0.273	0.258	1.41 (1.18-1.67)	1.01×10^{-4}
A-C	0.176	0.149	1.68 (1.37-2.07)	9.92×10^{-7}
A-T	0.013	0.010	1.33 (0.54-3.26)	0.532
G-T/A-C/A-T	0.463	0.418	1.49 (1.29-1.72)	5.01×10^{-8}

[1] Haplotypes of rs7229639 and rs4939827.

[2] Analyses were based on 3,067 cases and 3,283 controls included in Stage 2.

[3] Adjusted for age, sex and study site.