A Common SMAD7 Variant Is Associated with Risk of Colorectal Cancer: Evidence from a Case-Control Study and a Meta-Analysis

Qibin Song 1, Beibei Zhu 1,2,*, Weiguo Hu 1, Liming Cheng 3, Hongyun Gong 1, Bin Xu 1, Xiawen Zheng 2, Li Zou 2, Rong Zhong 2, Shengyu Duan 2, Wei Chen 2, Rui Rui 2, Jing Wu 2, Xiaoping Miao 2*

1 Department of Oncology, Renmin Hospital of Wuhan University, Wuhan, China, 2 Department of Epidemiology and Biostatistics and MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, 3 Department of Laboratory Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Abstract

Background: A common genetic variant, rs4939827, located in SMAD7, was identified by two recent genome-wide association (GWA) studies to be strongly associated with the risk of colorectal cancer (CRC). However, the following replication studies yielded conflicting results.

Method and Findings: We conducted a case-control study of 641 cases and 1037 controls in a Chinese population and then performed a meta-analysis, integrating our and published data of 34313 cases and 33251 controls, to clarify the relationship between rs4939827 and CRC risk. In our case-control study, the dominant model was significant associated with increased CRC risk (Odds Ratio (OR) = 1.46; 95% confidence interval (95% CI), 1.19–1.80). The following meta-analysis further confirmed this significant association for all genetic models but with significant between-study heterogeneity (all \( P \) for heterogeneity <0.1). By stratified analysis, we revealed that ethnicity, sample size, and tumor sites might constitute the source of heterogeneity. The cumulative analysis suggested that evident tendency to significant association was seen with adding study samples over time; whilst, sensitive analysis showed results before and after removal of each study were similar, indicating the highly stability of the current results.

Conclusion: Results from our case-control study and the meta-analysis collectively confirmed the significant association of the variant rs4939827 with increased risk of colorectal cancer. Nevertheless, fine-mapping of the susceptibility loci defined by rs4939287 should be imposed to reveal causal variant.

Introduction

Colorectal cancer (CRC) is the third most common cancer and the fourth leading cause of cancer mortality worldwide [1]. Among the risk factors and causes for CRC, genetic component has strongly contributed to CRC development, which accounts for approximately 35% of total cases as reflected by twin- and family-based studies [2]. However, so far genetic factors have incompletely been characterized.

Genome-wide association (GWA) study has greatly contributed to identification of common genetic variants associated to common disease without prior knowledge of gene function. Several recent GWA studies have reported multiple novel susceptibility loci to colorectal cancer [3–11]. Among these loci, the single nucleotide polymorphism (SNP, rs4939827), located in 16q21, has been strongly associated with risk of CRC by multiple GWA studies [3,10]. Broderick et al. firstly identified rs4939827 in a GWA set of 620 cases and 960 controls and 3 replication sets of 7377 cases and 5867 controls [10], and then Tenesa et al. further refined this finding in another comprehensive, phased-based GWA study comprising 16739 cases and 15545 controls [3]. Interestingly, rs4939827 maps to Mothers against decapentaplegic homolog 7 (SMAD7), a strong candidate gene in the famous transformation growth factor-β (TGF-β) pathway. SMAD7 acts as an intracellular antagonist of TGF-β signaling by recruiting SMURF to receptors for inactivation. Perturbation of SMAD7 and suppression of TGF-β signaling has been documented to be involved in CRC [12]. Much attention has been drawn to this SNP, however, several follow-up studies cannot replicate the association [13–16], which may be due to the sample size. For instance, in Chinese population, Xiong et al. reported a significant association of this SNP with CRC risk [17], whereas Li et al. failed to replicate this association [13]. Similar
obtained. This study was approved by the ethics committee of Tongji Hospital of Huazhong University of Science and Technology, Wuhan, China. Cases had been histopathologically confirmed with primary colorectal cancer and had not received any treatment prior to blood samples collection. Controls were randomly selected from a subject pool of individuals who participate in health check-up programs at the same hospital in the same time period as the patients were enrolled. Controls were frequency matched to patients by age (±5 years) and gender. All subjects were unrelated ethnic Han Chinese living in Wuhan region. At recruitment, a 5-ml peripheral venous blood sample was collected from each subject after written informed consent was obtained. This study was approved by the ethics committee of Tongji Hospital of Huazhong University of Science and Technology.

Materials and Methods

Study populations

In this study, a total of 641 new CRC cases and 1037 cancer-free controls were enrolled from between 2009 and 2011 from Tongji Hospital of Huazhong University of Science and Technology, Wuhan, China. Cases had been histopathologically confirmed with primary colorectal cancer and had not received any treatment prior to blood samples collection. Controls were randomly selected from a subject pool of individuals who participate in health check-up programs at the same hospital in the same time period as the patients were enrolled. Controls were frequency matched to patients by age (±5 years) and gender. All subjects were unrelated ethnic Han Chinese living in Wuhan region. At recruitment, a 5-ml peripheral venous blood sample was collected from each subject after written informed consent was obtained. This study was approved by the ethics committee of Tongji Hospital of Huazhong University of Science and Technology.

DNA isolation and genotyping

Genomic DNA was extracted from 5-ml of peripheral blood sample using the RelaxGene Blood System DP319-02 (Tiangen, Beijing, China) according to the manufacturer’s instructions. The genotypes of rs4939827 were determined by using the TaqMan SNP Genotyping Assay (Applied Biosystems, Foster city, CA) on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster city, CA). For quality control, 5% duplicated samples were randomly selected for to assess the reproducibility, with a concordance rate of 100%

Statistical analysis

Pearson χ² test, fisher exact test, and t test were employed to evaluate the differences in distribution of demographic characteristics and genotypes between case and control groups, where appropriate. Goodness-of-fit χ² test was adopted to assess Hardy-Weinberg Equilibrium (HWE) in the controls. Unconditional multivariate logistic regression analysis was used to estimate odds ratios (ORs) and their 95% confidence intervals (CIs) for the effect of rs4939827 genotypes on CRC risk, after adjusting for age and sex. To avoid the assumptions of genetic models, additive and dominant models for rs4939827 were also assessed. All statistical analyses were performed with the SPSS 12.0 software. A value of P<0.05 was considered representative of statistical significance.

Meta-analysis of rs4939827 in association with CRC risk

To further investigate the association between rs4939827 and CRC risk, a meta-analysis based on the published studies was carried out according to the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [19]. Systematic literature search updated to September, 2011 were performed in the PubMed and EMBASE databases (Figure S1), using the search strategy based on combinations of the keywords “rs4939827” and “colorectal cancer, colorectal neoplasia or colorectal adenoma” without language restriction. References listed in the retrieved articles were also scanned. Reviews, comments, and letters were also checked for additional studies. Studies were included if they met all of the following criteria: (a) assessment of the association between rs4939827 and CRC risk; (b) use of a case-control study or nested case-control study design; (c) information provided on genotype or allele frequency for risk estimates; (d) the genotype of controls is in Hardy-Weinberg equilibrium; (e) studies of humans. If the studies had overlapping subjects, only the study containing the largest population was finally included. Three reports were excluded due to lack of sufficient data for calculation of ORs after contacting with individual authors by E-mail [20–22].

The following data were extracted by two independent authors (B. Zhu & Q. Song): first author’s last name, country of origin, publication year, predominant ethnicity of participants, sample size, study method and design, source of control groups (population- or hospital-based controls), genotyping method. Counts of alleles and genotypes in cases and controls were extracted or calculated from published data. Pooled frequency of the T allele in various ethnic populations was estimated using the inverse variance method previously described by Thakkinstian et al [23]. ORs and their 95% CIs as the metrics of effect size were re-calculated for the genotypes TT versus CC and CT versus CC. A dominant model was assumed for rs4939827, and an additive “per-allele” model and a recessive model were also considered. In this study, we used the Cochran’s Q statistic to assess heterogeneity (heterogeneity was considered significant at P<0.01) [24]. The I² metric was applied to quantify heterogeneity irrespective of numbers of studies (I²=0–25%, no heterogeneity; I²=25–50%, moderate heterogeneity; I²=50–75%, large heterogeneity; I²=75–100%, extreme heterogeneity) [25]. A fixed-effects model, using Mantel-Haenszel method [26], was applied to pool data from studies when heterogeneity was negligible; otherwise, a random-effects model, using DerSimonian and Laird method, was applied [27]. Stratified analyses were performed, if feasible, according to ethnicity (European, Asian and mixed population), sample size (≤1000 and >1000 subjects), study design (GWA and replication study) and tumor site (colon, rectum and colorectal cancers). Sensitivity analysis was also performed to assess the influence of each individual study on overall estimates by

| Variables | Cases (N=641) | Control (N=1037) | P |
|-----------|--------------|-----------------|---|
| Sex       |              |                 |   |
| Male      | 384 (59.9)   | 613 (59.1)      |   |
| Female    | 257 (40.1)   | 424 (40.9)      |   |
| Age (years) | 56.31±12.59  | 57.24±10.86     |   |
| Tumor site|              |                 |   |
| Colon     | 250 (39%)    |                 |   |
| Rectum    | 391 (61%)    |                 |   |
| Duke’s stage | 83 (12.9%) | 228 (35.6%)   |   |
| A         | 227 (35.4%)  |                 |   |
| B         | 310 (47.9%)  |                 |   |
| C         | 103 (16.1%)  |                 |   |

*P value was calculated by the χ² test; *P value was calculated by the t test. doi:10.1371/journal.pone.0033818.001
The sequential removal of individual studies [28]. Cumulative analysis was performed to investigate the dynamic trend of the association between the SNP and CRC with accumulation of studies by published year [29]. Publication bias was estimated by funnel plot and Egger’s test [30,31]. All statistical analyses were carried out by Stata version 10.0.

### Results

#### Population characteristics.

A total of 641 incident cases of colorectal cancer and 1037 frequency-matched controls were enrolled in this study. As shown in Table 1, males were 59.9%.

### Table 2. The association between rs4939827 and colorectal cancer risk in a Chinese population.

|                  | Controls  | Cases  | OR (95% CI)<sup>b</sup> |
|------------------|-----------|--------|--------------------------|
|                  | CC/CT/TT  | CC/CT/TT | CT vs. CC | TT vs. CC | Dominant model | Additive model |
| Total            | 732/272/33| 399/232/10| 1.57 (1.27–1.94) | – | 1.46 (1.19–1.80) | 1.27 (1.05–1.52) |
| Tumor site       | –         | –       | –           | – | – | – |
| Colon            | 732/305<sup>a</sup> | 149/101<sup>a</sup> | – | – | 1.65 (1.24–2.20) | – |
| Rectum           | 732/305<sup>a</sup> | 250/141<sup>a</sup> | – | – | 1.35 (1.06–1.73) | – |
| Duke’s stage     | –         | –       | –           | – | – | – |
| A+B              | 732/305<sup>a</sup> | 184/127<sup>a</sup> | – | – | 1.67 (1.28–2.17) | – |
| C+D              | 732/305<sup>a</sup> | 215/115<sup>a</sup> | – | – | 1.28 (0.98–1.67) | – |

<sup>a</sup>CC/(CT+TT);
<sup>b</sup>ORs and their corresponding 95% CIs were calculated by multivariate logistic regression model after adjusting for age and sex.

doi:10.1371/journal.pone.0033318.t002

### Table 3. Characteristics of studies on rs4939827 polymorphisms and risk of colorectal cancer included in the meta-analysis.

| First author | Published year | Country | Ethnicity      | Study method | Study design | Genotyping method | Tumor site | Case/control |
|--------------|----------------|---------|----------------|--------------|--------------|-------------------|------------|--------------|
| Broderick P  | 2007           | UK      | European       | Nested CC    | GWAS         | Illumina         | CRC        | 620/960      |
|              |                |         |                | Replication  | Allele-PCR   | CRC               | 4422/3844  |
|              |                |         |                |               |              | CRC               | 1992/1680  |
|              |                |         |                |               |              | CRC               | 963/343    |
| Tenesa A     | 2008           | Scotland| European       | Nested CC    | GWAS         | Illumina         | CRC        | 2895/3059    |
|              |                |         |                | Replication  | TaqMan-PCR   | CRC               | 830/923    |
|              |                |         |                |               |              | CRC               | 1318/1140  |
|              |                |         |                |               |              | CRC               | 1173/1182  |
|              |                |         |                |               |              | CRC               | 2232/2250  |
|              |                |         |                |               |              | CRC               | 2150/2182  |
|              |                |         |                |               |              | CRC               | 1352/1336  |
|              |                |         |                |               |              | CRC               | 418/295    |
|              |                |         |                |               |              | CRC               | 4391/3178  |
| Curtin K     | 2009           | Leeds, UK | European   | CC            | Replication  | SNPlex           | CRC        | 245/216      |
|              |                |         |                |               |              | CRC               | 398/400    |
|              |                |         |                |               |              | CRC               | 422/425    |
| Thompson CL  | 2009           | USA     | Mixed          | CC            | Replication  | TaqMan-PCR       | Colon      | 554/709      |
|              |                |         |                |               |              | CRC               | 257/271    |
| Mates IN     | 2010           | Romania | European       | CC            | Replication  | Centaurus        | Colon, rectum | 92/95       |
|              |                |         |                |               |              | CRC               | 970/969    |
| Niittymaki I | 2010           | Finland | European       | CC            | Replication  | Sequencing       | CRC        | 1475/2287    |
| Slattery ML  | 2010           | USA     | Americans      | CC            | Replication  | TaqMan-PCR       | Colon      | 1475/2287    |
|              |                |         |                |               |              | CRC               | 1475/2287  |
| Xiong F      | 2010           | China   | Asians         | CC            | Replication  | T-ARMS-PCR       | Colon, rectum | 2124/2180   |
| von Holst     | 2010           | Sweden  | European       | Nested CC    | Replication  | deCode test      | CRC        | 1782/1679    |
| Li X         | 2011           | China   | Asians         | CC            | Replication  | Sequenom         | CRC        | 138/168      |
| Ho JW        | 2011           | HK, China | Asians        | CC            | Replicatoin  | Sequenom         | CRC        | 716/714      |
| Zhu B        | 2011           | China   | Asians         | CC            | Replication  | TaqMan-PCR       | Colon, rectum | 641/1037    |

Abbreviation: CC, case-control study; CRC, colorectal cancer.
doi:10.1371/journal.pone.0033318.t003
among cases compared with 59.1% among controls. Mean age was 56.31 years (±12.59) for cases and 57.24 years (±10.86) for controls. There was no significant difference in distribution of sex (P = 0.748) and age (P = 0.119) between case and control group. Of Cases, 39% had colon cancer and 61% had rectum cancer. Regarding tumor state, 12.9%, 35.6%, 35.4% and 16.1% of cases were classified as Duke’s A, B, C and D stage at the time of diagnosis, respectively.

Genotypes in the controls conformed to Hardy-Weinberg equilibrium (P = 0.214). Significant difference in genotype distribution was observed between cases and controls (χ² = 21.25, P < 0.001). In the multivariate logistic regression model, individuals with the CT genotype had a significant, 57% increased risk of CRC (OR = 1.57; 95% CI, 1.27–1.94, P < 0.001) compared to those with the CC homozygote. Due to the low frequency of the TT genotype (3.2% in controls and 1.6% in cases) in this study population, a dominant model was perform, by combining the TT with the CT into an T carrier (TT plus CT) group, and result showed that the T carrier also present significantly increased risk, compared with those carrying the CC genotype (OR = 1.46; 95%CI, 1.19–1.80, P < 0.001). In the allelic model, T allele carriers also showed significantly increased risk compared to those with the C allele (OR = 1.26; 95% CI, 1.05–1.51, P = 0.01). In the additive model, per-T allele similarly conferred an OR of 1.27 (95% CI, 1.05–1.52, P = 0.01) (Table 2).

We then stratified data according to the pathological factors under the dominant model. The CT plus TT genotypes were both associated with increased risk of colon and rectal cancers. Interestingly, the effect of the CT plus TT genotypes was larger in

| Genetic model | OR (95%CI)  | P     | P for Heterogeneity | I² (%) |
|---------------|------------|-------|---------------------|-------|
| Overall (n = 25) | T vs. C | 1.18 (1.14–1.22) | <0.001 | 0.001 | 55.0 |
| (n = 21) | TT vs. CC | 1.33 (1.21–1.47) | <0.001 | <0.001 | 60.6 |
| | CT vs. CC | 1.17 (1.09–1.26) | <0.001 | <0.001 | 59.1 |
| | Dominant | 1.20 (1.14–1.26) | <0.001 | <0.001 | 61.6 |
| | Recessive | 1.24 (1.17–1.32) | <0.001 | 0.039 | 38.4 |
| | Additive | 1.17 (1.14–1.20) | <0.001 | <0.001 | 54.8 |

doi:10.1371/journal.pone.0033318.t004

Figure 1. Forest plot of association of rs4939827 with CRC under allelic model.
doi:10.1371/journal.pone.0033318.g001
colon cancer (OR = 1.65; 95% CI = 1.24–2.20) than that in rectum cancer (OR = 1.35; 95% CI = 1.06–1.73). Regarding the Duke’s stage, the CT plus TT genotypes were associated with increased risk in early stage (A+B: OR = 1.67; 95% CI, 1.28–2.17) but not in advanced cancer (C+D: OR = 1.28; 95% CI, 0.98–1.67; Table 2).

Results of meta-analysis

Study characteristics. A total of 11 publications plus the current study, comprising 25 case-control studies of 34313 cases and 33251 controls, were finally included in this meta-analysis[3,10,13–17,32–35], of which, 19 studies were conducted in European [3,10,16–17,32–34], 5 in Asian [3,13,14,35], and 1 in the mixed population [15] (Table 3). The report by Broderick et al. only provided data on allele frequency and thus was only included in the pooled analysis of allelic OR [10].

Frequency of risk allele in control population. Both significant between-study heterogeneity were observed in European and Asian groups (P for heterogeneity <0.001). Under random-effects
model, the pooled frequency of the T allele was 51.2% (95% CI, 50.1%–52.2%) in European controls, which was markedly higher than that of 23.4% in Asian controls (95% CI, 10.4%–28.3%) (Figure S2). These pooled frequencies were similar to those reported in HapMap database of 0.508 and 0.256 for European and Asian, respectively.

**Overall meta-analysis of rs4939827 in associated with CRC.** As shown in Table 4, significant evidence of heterogeneity was seen in all genetic models (all P for heterogeneity < 0.05), and ORs for all genetic models were pooled under random-effects model. In allelic model, the T allele conferring a pooled OR of 1.18 compared to the C allele (95% CI, 1.14–1.22; Figure 1). Genotypic ORs of the TT versus CC and CT versus CC were 1.35 (95% CI, 1.21–1.47) and 1.17 (95% CI, 1.09–1.26), respectively. Similarly, the dominant, recessive and additive models were all associated with significantly increased risk of CRC.

**Stratified analysis.** To investigate the potential source of between-study heterogeneity, stratified analysis was performed (Table 5). After stratifying by ethnicity, significant heterogeneity still existed in European, whereas in Asian heterogeneity was effectively reduced. In European population, the variant in all genetic models presented significantly increased risk of CRC. In Asian population, all the genetic models except for the TT genotypic and recessive models were associated with increased risk of CRC, potentially suggesting that the T variant act in various manners between different ethnic populations. When stratified by sample size, we define the large group when the sample size was more than 1000, otherwise was small group, heterogeneity was almost removed in small sample subgroup but not large sample subgroup. Significant association of CRC risk with the variant remained in large sample studies for all genetic models, whereas only the recessive model showed significant result in small sample subgroup. According to tumor site, only the data on dominant model was available. For colon cancer, heterogeneity was still observed and no significant association was found, whereas the dominant model was significantly associated with increased risk of rectum cancer without evidence of heterogeneity. Regarding study design, GWA studies were merely pooled in allelic model due to limited studies for assessing genotypic model. Heterogeneity did not change after stratifying by GWAS and replication and significant association still existed.

**Sensitivity analyses and cumulative meta-analysis.** Due to the significant between-study heterogeneity for all genetic models, sensitivity analysis was performed, by removing the individual studies sequentially under random-effects model, to assess the effect of each study on the pooled estimate. As shown in Table 6, the pooled OR for the allelic model was similar before and after elimination of each study. Similar results were seen for other genetic models that no single study dramatically change the pooled ORs, indicating the robust stability of the current results.

Accumulative meta-analysis was carried out via the assortment of studies by publication time. As shown by Figure 2, in the allelic model, the 95% CIs for the pooled OR became increasingly narrower with each accumulation of more studies, indicating the progressively boosted precision of the estimation by continual adding more samples. Simultaneously, inclinations toward significant association were evident over time. Similar results were seen in other genetic models.

**Publication Bias.** As reflected by the funnel plot (Figure S3.) and Egger’s test, no evidence of publication bias was observed in all genetic models (all P for egger’s test > 0.05).

**Discussion**

rs4939827 located at 18q21 was revealed to be associated with CRC risk by two GWA studies, but inconsistent results have been reported by multiple following replication studies. In this study, we initially found a significant association between the variant 4939827 and increased risk of CRC in a case-control set of a Chinese population. Then the following meta-analysis, first to integrate GWA and replication data from 25 case-control studies of 34313 cases and 33251 controls, consistently indicated the significant association of rs4939827 with the risk of CRC. This significant association was further confirmed by cumulative meta-analysis, presenting the effect of the variant got increasingly significant with each accumulating of more data over time. rs4939827 is located in intron 3 of SMAD7, which encodes an inhibitory SMAD protein that function as a negatively feedback regulator of TGF-β signals [36]. There was evidence that the over-expression of SMAD7 could promote tumorogenesis via disturbing TGF-β-induced growth inhibition and apoptosis. Although we herein confirmed the association between the rs4939827 and CRC risk, whether this SNP is causative was still uncertain. Intriguingly, Houlston et al. have identified a novel C to G SNP (rs4939827) in dbSNP (MAF = 0.47), through re-sequencing the linkage disequilibrium (LD) region tagged by rs4939827 in 2532 CRC cases and 2607 controls, was maximally associated with CRC risk [37]. The following functional models further provided evidence for the role of this SNP in transcription factor binding, proposing that this functional SNP was likely to be one of the causal variants in susceptibility loci tagged by rs4939827.

---

**Table 6. Sensitivity analysis of allelic model.**

| Study omitted                  | OR (95%CI)       | P for heterogeneity | I²   |
|-------------------------------|------------------|---------------------|------|
| Broderick 2007 (GWAS) [10]    | 1.17 (1.13–1.22) | 0.001               | 53.5%|
| Broderick 2007 (B)            | 1.18 (1.14–1.23) | 0.001               | 54.6%|
| Broderick 2007 (C)            | 1.17 (1.13–1.22) | <0.001              | 56.1%|
| Broderick 2007 (D)            | 1.18 (1.14–1.23) | <0.001              | 56.2%|
| Tenesa 2008 (Scotland GWAS) [3] | 1.17 (1.13–1.22) | <0.001              | 56.9%|
| Tenesa 2008 (Scotland Replication) | 1.19 (1.14–1.23) | 0.001               | 53.7%|
| Tenesa 2008 (Australia)       | 1.18 (1.13–1.22) | <0.001              | 56.3%|
| Tenesa 2008 (Cananda)         | 1.18 (1.13–1.22) | <0.001              | 56.9%|
| Tenesa 2008 (England)         | 1.17 (1.13–1.22) | 0.001               | 55.6%|
| Tenesa 2008 (German)          | 1.17 (1.13–1.22) | 0.001               | 54.9%|
| Tenesa 2008 (Israel)          | 1.17 (1.13–1.20) | 0.046               | 35.3%|
| Tenesa 2008 (Spain)           | 1.18 (1.14–1.22) | <0.001              | 56.9%|
| Tenesa 2008 (Japan)           | 1.18 (1.13–1.22) | <0.001              | 56.9%|
| Curtin 2009 (UK-Leeds) [17]   | 1.18 (1.13–1.22) | <0.001              | 56.7%|
| Curtin 2009 (UK-Sheffield)    | 1.18 (1.14–1.23) | 0.001               | 53.9%|
| Curtin 2009 (USA-Utah)        | 1.18 (1.13–1.22) | <0.001              | 56.9%|
| Thompson 2009 [15]            | 1.19 (1.14–1.23) | 0.001               | 52.6%|
| Xiong 2010 [13]               | 1.18 (1.13–1.22) | <0.001              | 56.9%|
| Mates 2010 [16]               | 1.18 (1.14–1.22) | 0.001               | 54.3%|
| Nittymaki 2010 [32]           | 1.18 (1.13–1.22) | <0.001              | 56.9%|
| Slattery 2010 [33]            | 1.18 (1.14–1.22) | 0.001               | 55.6%|
| von Holst 2010 [34]           | 1.18 (1.14–1.23) | 0.001               | 54.7%|
| Li 2011 [14]                  | 1.18 (1.14–1.23) | 0.001               | 54.9%|
| Ho 2011 [35]                  | 1.18 (1.13–1.22) | <0.001              | 56.9%|
| Zhu 2011                      | 1.18 (1.13–1.22) | <0.001              | 56.6%|
| Combined                     | 1.18 (1.14–1.22) | 0.001               | 55.0%|

DOI: 10.1371/journal.pone.0033318.t006
Nevertheless, the obvious evidence of between-study heterogeneity in this meta-analysis should be issued. We have applied a comprehensive stratified analysis to interrogate the potential source of heterogeneity. After stratifying by ethnicity, heterogeneity was largely reduced in Asian, reflecting ethnicity could partly explain the heterogeneity. The further supports of that came from the evidence that various manners the T variant likely act in and different allele frequencies between European and Asian populations. When stratified by tumor sites, rectum cancer subgroup did not show heterogeneity anymore, suggesting tumor sites might also be a potential source of heterogeneity. Additionally, significant association only presented for rectum cancer, which was inconsistent with our case-control study, possibly due to our small sample size for stratification by tumor sites. Regarding sample size, heterogeneity was almost removed in small sample studies but not large sample subgroup, possibly due to more complex confounding factors introduced into large sample. Significant association remained in large sample studies for all genetic models, whereas only the recessive model showed significant result in small sample size to detect the modest effect of the variant. Taken together, we revealed that the ethnicity, tumor sites, and sample size might constitute source of heterogeneity in this meta-analysis. Whilst, the significant association of rs4939827 presented in the subgroup of replication studies, consistent with the result from GWA study subgroup, suggested this meta-analysis succeed in amplifying power to detect the modest effect of this variant by pooling data across studies.

Furthermore, the sensitivity analysis and publication bias assessment indicated the current results from this meta-analysis were stable. Despite the clear strengthen of this study that applied a comprehensive analysis strategy, several limitations should be Figured out. First, the sample size of our case-control study was relatively small. Nevertheless, the following meta-analysis with enough power has drawn the consistent result with our case-control study. Second, the analysis of separate effect for colon or rectum cancer conferred limited power, and more studies are needed. Additionally, CRC is a complex trait corporately influenced by genetic and environmental factors; however, lacking the environment data limited us to further assess gene-environment interaction.

In conclusion, our study in a Chinese population and this meta-analysis collectively confirm the significant association between SNP rs4939827 and the colorectal cancer risk in European and Asian populations. However, although a novel SNP in highly LD with rs4939827 has been proposed to be causal, further fine-mapping of the CRC susceptibility loci tagged by rs4939827 is warranted to uncover more causal variants, especially for the low-frequency or rare functional variants.

**Supporting Information**

*Figure S1 Flow chart of study selection.* (TIF)
Figure S2  Pooled frequency of T allele in European and Asian population.

(TIF)

Figure S3  Funnel plot of publication bias under allelic model.

(TIF)

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, et al. (2011) Global cancer statistics. CA Cancer J Clin 61: 69–90.
2. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, et al. (2000) Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343: 78–85.
3. Tenesa A, Farrington SM, Pendergast JGD, Porteous ME, Walker M, et al. (2008) Genome-wide association study identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 10q21. Nature Genetics 40: 631–637.
4. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, et al. (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nature Genetics 39: 989–994.
5. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, et al. (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 40: 625–630.
6. Zanke BW, Greenwood CM, Rangej J, Kustra R, Tenesa A, et al. (2007) Common genetic variants at the CRAG1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 40: 26–28.
7. Mates IN, Csiki I, Mates D, Constantinescu V, Badea P, et al. (2010) Association of common genetic variants with colorectal cancer risk in a Romanian sample. Molecular Biology Reports.
8. Halder SK, Rachakonda G, Deane NG, Datta PK (2008) Smad7 induces expression. Genome Research 19: 987–993.
9. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, et al. (2008) Common genetic variants at the CRAG1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 40: 937–977.
10. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, et al. (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q11, 8q26, 12q13.13 and 20q13.33. Nat Genet 42: 973–977.
11. Cui R, Okada Y, Jang SG, Ku JL, Park JG, et al. (2011) Common variant in 6q26-q27 is associated with distal colon cancer in an Asian population. Gut.
12. He J, Wilkins LR, Stram DO, Kolonel LN, Henderson BE, et al. (2010) Generalizability and Epidemiologic Characterization of Eleven Colorectal Cancer GWAS Hits in Multiple Populations. Cancer Epidemiology Biomarkers & Prevention 19: 78–85.
13. Li X, Yang X-X, Hu N-Y, Sun J-Z, Li F-X, et al. (2011) A risk-associated single nucleotide polymorphism of SMAD7 is common to colorectal, gastric, and lung cancers in a Han Chinese population. Molecular Biology Reports.
14. Thompson CL, Pummer SJ, Acheson LS, Tucker TG, Casey G, et al. (2009) Association of common genetic variants in SMAD7 and risk of colon cancer. Carcinogenesis 30: 902–906.
15. Li X, Yang X-X, Hu N-Y, Sun J-Z, Li F-X, et al. (2011) A risk-associated single nucleotide polymorphism of SMAD7 is common to colorectal, gastric, and lung cancers in a Han Chinese population. Molecular Biology Reports.
16. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, et al. (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q21.2. Nature Genetics 39: 984–988.
17. Curtin K, Lin WY, George R, Katory M, Shorto J, et al. (2009) Meta Association of Colorectal Cancer Confirms Risk Alleles at 8q24 and 10q21. Cancer Epidemiology Biomarkers & Prevention 18: 616–621.
18. Pittman AM, Naranjo S, Webb E, Broderick P, Lips EH, et al. (2009) The colorectal cancer risk at 10p14 is caused by a novel variant altering SMAD7 expression. Genome Research 19: 987–993.

Author Contributions

Conceived and designed the experiments: XM. Performed the experiments: BZ, LZ, RZ, SD, WC, RR. Analyzed the data: QS, BZ, JW. Contributed reagents/materials/analysis tools: WH, LC, HG, BX. Wrote the paper: QS, BZ, XM.

PLoS ONE | www.plosone.org 8 March 2012 | Volume 7 | Issue 3 | e33318

SMAD7 rs4939827 and Colorectal Cancer