A Rare Missense Variant in TCF7L2 Associates with Colorectal Cancer Risk by Interacting with a GWAS-Identified Regulatory Variant in the MYC Enhancer

Jiang Chang1, Jianbo Tian1, Yang Yang1, Rong Zhong1, Jiaoyuan Li1, Kan Zhai2, Juntao Ke1, Jiao Lou1, Wei Chen1, Beibei Zhu1, Na Shen1, Yi Zhang1, Yajie Gong1, Ying Zhu1, Danyi Zou1, Xiating Peng1, Kun Huang3, and Xiaoping Miao1

Abstract

Genome-wide association studies (GWAS) of colorectal cancer have identified several common susceptible variants in gene regulatory regions. However, low-frequency or rare coding risk variants have not been systematically investigated in patients with colorectal cancer from Chinese populations. In this study, we performed an exome-wide association analysis with 1,062 patients with colorectal cancer and 2,184 controls from a Chinese population. Promising associations were further replicated in two replication sets: replication stage I with 2,478 cases and 3,880 controls, and replication stage II with 3,761 cases and 4,058 controls. We identified two variants significantly associated with colorectal cancer risk: a novel rare missense variant in TCF7L2 [rs138649767, OR = 2.08, 95% confidence interval (CI): 1.69–2.57, \( P = 5.66 \times 10^{-12} \)] and a previous European GWAS-identified 3'-UTR variant in ATF1 (rs11169571, OR = 1.18, 95% CI: 1.13–1.24, \( P = 1.65 \times 10^{-12} \)). We found a significant interaction between the TCF7L2 missense variant rs138649767 and a previous GWAS-identified regulatory variant rs6983267 in the MYC enhancer (\( P_{\text{interaction}} = 0.0002 \)). Functional analysis of this variant revealed that TCF7L2 with rs138649767-A allele harbored the ability to activate the MYC enhancer with rs6983267-G allele and enhance colorectal cancer cell proliferation. In addition, the ATF1 rs11169571 variant significantly correlated with ATF1 expression by affecting hsa-miR-1283 and hsa-miR-520d-5p binding. Further ChIP-seq and gene coexpression analyses showed that oncogenes NRAS and BRAF were activated by ATF1 in colorectal cancer. These results widen our understanding of the molecular basis of colorectal cancer risk and provide insight into pathways that might be targeted to prevent colorectal cancer.

Significance: Exome-wide association analysis identifies a rare missense variant in TCF7L2 and a common regulatory variant in ATF1 as susceptibility factors of colorectal cancer.

Graphical Abstract: http://cancerres.aacrjournals.org/content/canres/78/17/5164/F1.large.jpg. Cancer Res; 78(17): 5164–72. ©2018 AAGR.

Introduction

Colorectal cancer is the third most common cancer in the world and the fifth leading cause of cancer-related death in China (1, 2).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

An increase in age-standardized incidence and mortality rate has been observed for colorectal cancer in Chinese populations over the past decades (2). Several environmental factors such as...
overnutrition, obesity, and smoking have been associated with the development of sporadic colorectal cancer (3, 4). However, only a portion of exposed individuals develop colorectal cancer, indicating that an individual’s genetic makeup may also play an important role in the etiology of colorectal cancer (5, 6).

Genome-wide association study (GWAS) has proved to be an efficient approach for identifying genetic variants correlated with risk of colorectal cancer and more than 20 loci have been found in East Asian populations (7–11). However, most of these risk variants are common regulatory variants with minor allele frequency (MAF) >0.05 and are located near genes or in gene introns, explaining only a small fraction of heritability of colorectal cancer (12). Therefore, novel approaches are needed to find more susceptible variants in colorectal cancer, especially functional ones. The Illumina HumanExome Beadchip (also known as the “exome chip”) has been developed to mainly focus on variants in coding regions with low or rare frequency in populations, which may help us to identify more functional variants associated with colorectal cancer. It has served as a good complement to GWAS, and several groups already have remarkable findings using this approach (13, 14).

In this study, we genotyped 247,870 variants in 1,062 patients with colorectal cancer and 2,184 controls from a Chinese population using the exome chip. Promising associations were replicated in two replication sets consisting of 6,239 colorectal cancer cases and 7,938 controls in total. We identified a rare missense variant rs138649767 in transcription factor 7 like 2 (TCF7L2) and a common 3′-UTR variant rs11169571 in activating transcription factor 1 (ATF1) that were significantly associated with colorectal cancer risk in a total of 7,301 colorectal cancer cases and 10,122 controls. Through further analysis of rs138649767, we found a statistical and functional interaction between this variant and a previous GWAS-identified regulatory variant rs6983267 in MYC proto-oncogene (MYC) enhancer. In addition, functional analyses of rs11169571 showed that it was correlated with ATF1 expression and NRAS proto-oncogene (NRAS) and B-Raf proto-oncogene (BRAF) activation.

Patients and Methods

Study subjects

The discovery and the replication stage I contained 3,540 individuals with colorectal cancer who were collected from local hospitals in Wuhan, China. The replication stage II consisted of 3,761 individuals with colorectal cancer recruited from the cancer hospital of Chinese Academy of Medical Sciences in Beijing, China. All cases were histopathologically or cytologically confirmed by at least two local pathologists according to the World Health Organization classification. Demographic characteristics including age and sex (Supplementary Table S1) were obtained from the medical records of these individuals. All the controls were cancer-free individuals who lived in the same residential areas and were seeking routine physical examination in the outpatient departments at the hospital where the cases were collected (15–17). They were volunteered to participate in the epidemiology survey during the same period. The cancer-free information was obtained through routine physical examination and self-report. Written informed consent was obtained from each subject at recruitment, and this study was approved by the Institutional Review Board of Tongji Medical College, Huazhong University of Science and Technology. This study was conducted in accordance with the Declaration of Helsinki principles.

Genotyping and quality control

Genomic DNA was extracted from blood samples. Genotyping was performed using the Illumina HumanExome BeadChip in the discovery stage. All procedures were based on the standard protocol (18) and regular screening methods (Fig. 1). A total of 168,930 variants were excluded from subsequent analysis because they (i) were monomorphic variants (166,568), (ii) were not in autosomal chromosomes (1,538), (iii) had a call rate of <95% (292), or (iv) presented a P value < 0.0001 in a Hardy–Weinberg equilibrium test (532). A total of 23 subjects were excluded because they had an overall genotyping rate of <95%. In the replication stage I, the Sequenom MassARRAY system was used for genotyping. In the replication stage II, the TaqMan genotyping system was used. The previous GWAS-identified rs6983267 was genotyped using TaqMan genotyping system. The genotyping consistency for the discovery stage was assessed on the basis of 300 replicate samples genotyped using both exome chip and Sequenom MassARRAY system for all 18 promising SNPs, and the concordance rate of each variant was between 99.7% and 100%. We also employed the direct sequencing of PCR products to replicate 50 randomly selected GG carriers and 50 randomly selected GA carriers for rs138649967, and the concordance rate was 100%.

Association analysis

The principal component analysis (PCA) for population stratification was run by EIGENSTRAT (19) based on 4,604 autosomal ancestry information markers included on the exome chip (18). Logistic regression models, applying an additive genetic model for each allele, were used to test the association between genetic variants and genetic association adjusting for sex, age, and the top three PCs in the discovery stage. Adjustment for sex and age were considered in the replication stages or all samples combined, respectively. Nineteen variants with P < 0.0005 were considered as significant in the discovery stage (Supplementary Table S2) and 18 variants (except rs1129406, which is in perfect LD with rs11169571) were selected for further replication (Supplementary Table S3). A P value below 0.0028 (0.05/18) was considered as significant in the replication stage I. A Manhattan plot was created by Haploview (20) and QQ plots were drawn with the R software (version 3.1.1).

Genotype imputation

We phased the haplotypes with SHAPEIT (21) and performed imputations with IMPUTE2 (22, 23) software to impute ungenotyped SNPs in a 2-Mb region centered on rs11169571. This analysis was based on the linkage disequilibrium (LD) and haplotypes information from the 1000 Genomes Project Phase 3 ASN samples as references. Poorly imputed variants with an information measure (Is) < 0.40 (output from IMPUTE2 info file) were excluded from subsequent analyses. After imputation, 76 genotyped variants and 4,379 well-imputed ones were used for the conditional analysis. Regional plots were created using LocusZoom (24) with hg19/1000Genomes Nov 2014 ASN for the LD analysis.

Cell lines

LoVo and SW480 cells were obtained from the China Center for Type Culture Collection and were cultured in DMEM (Gibco)
supplemented with 10% FBS (Gibco) and 1% antibiotics (100 U/mL penicillin and 0.1 mg/mL streptomycin) at 37°C in a humidified atmosphere of 5% CO2. All cell lines were tested routinely by DNA sequencing using the AmpF/STR Identifier Kit (Applied Biosystems) and the latest date the cells were tested was June 1st, 2017. Cells were tested for the absence of Mycoplasma contamination (MycoAlert, Lonza) and the latest date the cells were tested was June 1st, 2017. Cells were seeded in 24-well plates (1 x 10^5 cells per well) 24 hours prior to transfection experiments. The cells used in experiments were within 10 passages from thawing.

Construction of reporter plasmids and transient transfections

For transfection assays of rs138649767, LoVo and SW480 cells were transiently cotransfected with pcDNA3.1(+) vector containing full-length TCF7L2 cDNA with A or G allele of rs138649767 and the pGL3-basic empty vector, or pGL3-basic vector containing the MYC promoter, or pGL3-basic vector containing the MYC promoter and the 8q24 MYC enhancer with G or T allele of rs6983267. For transfection assays of rs11169571, LoVo and SW480 cells were transiently cotransfected with psiCHECK-2 empty vector or psiCHECK-2 vector containing either the T or C allele of rs1169571 and simultaneously were treated with miRNA mimics or miRNA inhibitor. A luciferase reporter assay was performed using the Dual Luciferase Reporter Kit (Luciferase Assay System, Promega) according to the manufacturer’s recommendations. Three independent experiments were performed, and triplicate wells were transfected in each experiment.

Analysis of cell proliferation

Cells were seeded in 96-well flat-bottomed plates, and each well contained 2,000 cells in 100 μL of cell suspension. After a certain time in culture, cell viability was measured using CCK-8 assays (Dojindo Laboratories). Each experiment included six replicates and was repeated three times.

Analysis of ATF1 ChIP-seq and coexpression data

Chromatin immunoprecipitation sequencing (ChIP-seq) data for ATF1 in the LoVo cell line was obtained from the Cistrome database. ATF1 binding peaks were intersected with promoter regions of genes annotated by Gencode v24. The promoter is defined as the region between 2 kb upstream and 250 bp downstream from any transcription start site of a coding transcript gene. Coexpression of selected genes was obtained from the eBiportal database. Reads per kilobase per million mapped reads (RPKM) data for 276 colorectal cancer tissue samples from The Cancer Genome Atlas were used. Gene coexpression was tested using Spearman correlation, and genes with \( P < 0.05 \) and \( r > 0.3 \) were considered significant.

Data availability statement

The data that support the findings of this study have been deposited in the Genome Sequence Submission (Gsub) of Beijing Institute of Genomics, Chinese Academy of Sciences (http://bigd.big.ac.cn/gsub/) with accession number PRJCA000872.

Results

Characteristics of study subjects

Basic clinical statistics of all study subjects are presented in Supplementary Table S1. In total, 7,301 colorectal cancer cases and 10,122 controls were included in this study with average ages of 58.33 and 61.37 years for cases and controls, respectively. There were 4,517 (61.9%) men and 2,784 (38.1%) women who were diagnosed with colorectal cancer and 6,409 (63.3%) men and

![Figure 1. The flowchart of this study.](image-url)
Figure 2. Manhattan plot for associations between genetic variants and colorectal cancer risk. We assessed association using a logistic regression model, with adjustment for age, sex and the top three PCs of population stratification. The $-\log_{10} P$ values ($y$-axis) of 78,940 SNPs in 1,062 colorectal cancer cases and 2,184 controls are presented on the basis of the chromosomal positions of the SNPs ($x$-axis). The red horizontal line represents $P = 0.005$.

3,713 (36.7%) women were recruited as controls. No significant population stratifications were observed in discovery stage through PCA (Supplementary Fig. S1A–S1D). The genetic inflation factor ($\lambda$ = 1.041) indicated a good match between the expected association test statistics $P$ values distributions and the observed values; founders were not at work (Supplementary Fig. S2).

Two variants were associated with risk of colorectal cancer

The Manhattan plot of genetic variations associated with colorectal cancer gives an overview of $P$ values in the discovery stage (Fig. 2). Of the 78,940 variants passing quality control, 19 promising variants with MAF >0.1% and $P<0.0005$ were selected for further replication (Supplementary Table S2). Two variants were successfully verified in the both two replication stages: rs138649767 located in the exon of TCF7L2 [OR = 2.08, 95% confidence interval (CI): 1.69–1.57, $P = 5.66 \times 10^{-12}$] and rs11169571 located in the 3’-UTR of ATF1 [OR = 1.18, 95% CI: 1.13–1.24, $P = 1.65 \times 10^{-12}$] were significantly correlated with risk of colorectal cancer (Table 1). The rare missense rs138649767 variant was a novel signal and the rs11169571 variant were only identified in a meta-analysis of four European GWAS.

rs138649767 activated the MYC enhancer and enhanced colorectal cancer cells proliferation

Previous studies have found that the TCF7L2 bind an enhancer region in 8q24 and activate MYC. The binding affinity was affected by a GWAS-identified rs983267 variant. We then have the hypothesis that the identified missense rs138649767 variant in TCF7L2 may also affect the binding affinity for TCF7L2 and its target region and have functional interaction with rs6983267. Therefore, we performed a series of analyses to test this hypothesis. We transfected the pGL3-basic vector containing the MYC promoter and enhancer with different rs6983267 alleles into LoVo and SW480 cells, and overexpressed different TCF7L2 rs138649767 variants (Fig. 3A). The reporter assay result showed that transfection of TCF7L2[A] significantly activates the MYC enhancer, especially for enhancer with rs6983267[G], compared with transfection of TCF7L2[G] (Fig. 3B). This functional interaction result was consistent with our findings through population analysis. To further characterize the function of TCF7L2 variant in colorectal cancer cells, we overexpressed different TCF7L2 variants in LoVo and SW480 cells and tested the rate of cell proliferation. The TCF7L2[A] overexpression significantly enhanced LoVo and SW480 cells’ proliferation compared with overexpression of

Table 1. Result of the identified variants that were associated with colorectal cancer risk in the discovery, replication, and combined samples

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Gene</th>
<th>Stage</th>
<th>Genotype</th>
<th>Cases no. (%)</th>
<th>Controls no. (%)</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs138649767</td>
<td>10q25</td>
<td>TCF7L2</td>
<td>Discovery</td>
<td>GG</td>
<td>1,015 (95.6)</td>
<td>2,143 (98.1)</td>
<td>1.00 (Reference)</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GA</td>
<td>47 (4.4)</td>
<td>41 (1.9)</td>
<td>2.35 (1.53–3.60)</td>
<td>9.95 x 10^{-5}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication I</td>
<td>GG</td>
<td>2,402 (96.9)</td>
<td>5,821 (98.5)</td>
<td>1.00 (Reference)</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GA</td>
<td>76 (3.1)</td>
<td>59 (1.5)</td>
<td>2.12 (1.49–3.02)</td>
<td>3.16 x 10^{-5}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Replication II</td>
<td>GG</td>
<td>3,658 (97.3)</td>
<td>4,004 (98.7)</td>
<td>1.00 (Reference)</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GA</td>
<td>103 (2.7)</td>
<td>54 (1.3)</td>
<td>2.09 (1.50–2.92)</td>
<td>1.37 x 10^{-5}</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Combined</td>
<td>GG</td>
<td>7,075 (96.9)</td>
<td>9,968 (98.5)</td>
<td>1.00 (Reference)</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GA</td>
<td>226 (3.1)</td>
<td>154 (1.5)</td>
<td>2.08 (1.69–2.57)</td>
<td>5.66 x 10^{-12}</td>
</tr>
</tbody>
</table>

| rs11169571 | 12q13 | ATF1      | Discovery  | TT       | 1,073 (49.3) | 1,073 (49.3)   | 1.00 (Reference) | 0.001  |
|           |       |           |           | TC       | 495 (46.6)   | 926 (42.4)     | 1.30 (1.11–1.52) | 9.58 x 10^{-4} |
|           |       |           |           | CC       | 123 (11.6)   | 185 (8.5)      | 1.55 (1.20–2.00) | 8.45 x 10^{-4} |
|           |       |           |           | Additive |            |               | 1.27 (1.13–1.42) | 3.91 x 10^{-5} |
|           |       |           | Replication I | TT   | 1,077 (45.5) | 1,868 (48.1)  | 1.00 (Reference) | 0.001  |
|           |       |           |           | TC       | 1,054 (46.6) | 1,669 (43.0)  | 1.21 (1.08–1.34) | 7.93 x 10^{-4} |
|           |       |           |           | CC       | 247 (10.0)   | 343 (8.8)     | 1.22 (1.01–1.47) | 0.0359 |
|           |       |           |           | Additive |            |               | 1.14 (1.05–1.24) | 0.0011 |
|           |       |           | Replication II | TT  | 1,603 (42.6) | 1,947 (48.0)  | 1.00 (Reference) | 0.001  |
|           |       |           |           | TC       | 1,784 (47.4) | 1,771 (43.6)  | 1.22 (1.12–1.34) | 2.07 x 10^{-5} |
|           |       |           |           | CC       | 574 (9.9)    | 340 (8.4)     | 1.34 (1.14–1.57) | 4.20 x 10^{-4} |
|           |       |           |           | Additive |            |               | 1.18 (1.10–1.27) | 2.07 x 10^{-6} |
|           |       |           | Combined  | TT       | 3,124 (42.8) | 4,888 (48.3)  | 1.00 (Reference) | 0.001  |
|           |       |           |           | TC       | 3,453 (47.0) | 4,366 (45.1)  | 1.23 (1.16–1.31) | 1.32 x 10^{-10} |
|           |       |           |           | CC       | 744 (10.2)   | 868 (8.8)     | 1.35 (1.19–1.48) | 2.32 x 10^{-7} |
|           |       |           |           | Additive |            |               | 1.18 (1.13–1.24) | 1.65 x 10^{-12} |

NOTE: The novel identified variant in this study is in bold. $P$ values are two sided and were calculated in logistic regression analysis adjusted for sex and age. Abbreviation: Chr, Chromosome region.
TCF7L2[G], or the control vector (Fig. 3C). All the above results suggested that the rs138649767 variant was associated with increased risk of colorectal cancer through activating the MYC enhancer (Fig. 3D).

Statistical interactions were identified between rs138649767 and rs6983267

To further characterize the interaction between rs138649767 and rs6983267, we performed a statistical interaction between these two variants. Interestingly, we found a significant interaction between rs138649767 in TCF7L2 and rs6983267 in MYC enhancer ($P = 0.0002$, Table 2). Compared with individuals with both protective alleles of these two variants, individuals with either one risk allele have 1.19- and 1.84-fold risk of colorectal cancer, and individuals with both two risk alleles have 5.11-fold risk of colorectal cancer (Table 2).

The rs11169571 was the only signal in 12q13.12 region in Chinese populations

The identified rs11169571 variant located in the 12q13.12 region. Four SNPs (rs7136702 and rs11169552, rs4768903, and rs34245511) in this region were identified in previous colorectal cancer GWAS in European populations (25–27), but not in colorectal cancer GWAS in Chinese populations (7, 8, 10). To further characterize colorectal cancer associations at the 12q13.12 region in the Chinese population, we performed imputation analysis for a 2-Mb region centered on the rs11169571 (Supplementary Fig. S3A). We found that only the rs11169571 signal was significant in this region. When conditioning on rs11169571, no other independent significant SNPs in this region were observed (Supplementary Fig. S3B), while conditioning on either of the four variants identified in European GWAS (rs4768903, rs7136702, rs11169552, and

<table>
<thead>
<tr>
<th>rs138649767 genotype</th>
<th>rs6983267 genotype</th>
<th>Cases no. (%)</th>
<th>Controls no. (%)</th>
<th>OR (95% CI)</th>
<th>$P$</th>
<th>$P_{interaction}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>TT + TG</td>
<td>5,645 (77.3)</td>
<td>8,231 (81.3)</td>
<td>1.00 (Reference)</td>
<td>1.00</td>
<td>0.0002</td>
</tr>
<tr>
<td>GG</td>
<td>GG</td>
<td>1,430 (19.6)</td>
<td>1,737 (17.2)</td>
<td>1.19 (1.10–1.29)</td>
<td>9.50 x 10^{-5}</td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>TT + TG</td>
<td>176 (2.4)</td>
<td>159 (1.4)</td>
<td>1.84 (1.47–2.31)</td>
<td>1.39 x 10^{-2}</td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>GG</td>
<td>50 (0.7)</td>
<td>51 (0.5)</td>
<td>5.15 (2.86–9.35)</td>
<td>3.76 x 10^{-3}</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: $P$ values for different haplotypes were two sided and were calculated by an additive model in logistic regression analysis adjusted for sex and age. $P$ values for G x G interaction were calculated by conducting a one degree-of-freedom Wald test of a single interaction parameter (SNP x SNP) as implemented in an unconditional logistic regression, with age and sex as covariates.
Figure 4.
Functional analysis of the ATP1 rs11169571 variant. A, Predicted effect of allelic variation at rs11169571 on recognition of miRNAs. SNP rs11169571 occurs in the 8-bp seed sequence of complementarity at the 5' end of hsa-miR-520d-5p and hsa-miR-1283. Base pairing is indicated by a solid (Watson–Crick) vertical line. B, Significant correlation between rs11169571 genotypes and ATP1 expression in colon tissues. The eQTL results were obtained from 124 colon tissues in the GTEx database, with the P value being $6.7 \times 10^{-8}$ and effect size being 0.5. C, Luciferase reporter assays to measure T versus C allele difference at rs11169571 with the presence of hsa-miR-1283 in LoVo and SW480 cells. Cells were transiently cotransfected with constructs and 20, 40, 80 nmol/L hsa-miR-1283 mimic, hsa-miR-1283 inhibitor, or negative control. The results are shown as relative luciferase activity versus negative control. Data are from three independent transfection experiments with assays conducted in three replications. *P < 0.01; **P < 0.001; ***P < 0.0001 compared with control by unpaired Wilcoxon rank-sum test. D, Luciferase reporter assays to measure T versus C allele difference at rs11169571 with the presence of hsa-miR-520d-5p in LoVo and SW480 cells. Cells were transiently cotransfected with constructs and 20, 40, 80 nmol/L hsa-miR-520d-5p mimic, hsa-miR-1283 inhibitor, or negative control. The results are shown as relative luciferase activity versus negative control. Data are from three independent transfection experiments with assays conducted in three replications. *P < 0.01; **P < 0.001; ***P < 0.0001 compared with control by unpaired Wilcoxon rank-sum test. E, ChIP-seq results for H3K4me3 (top) and ATP1 (bottom) at gene promoter regions of NRAS and BRAF. Data was obtained from the Cistrome database. F, Gene coexpression results between ATP1 and oncogenes NRAS or BRAF. Reads per kilobase per million mapped reads (RPKM) data for 276 colorectal cancer tissue samples from The Cancer Genome Atlas were used. Gene coexpression was tested using Spearman correlation.
rs34245551), the rs11169571 signal was still significant (Supplementary Fig. S3C–S3F).

The rs11169571 variant influences ATF1 expression by affecting mRNA binding

As rs11169571 is located in the 3′-UTR of ATF1, we then examined whether rs11169571 could affect mRNA binding. Through an in silico analysis using MirSNP (28) and miRNASNP (29), we found that rs11169571 SNPs lies within a binding site for the seed region of hsa-miR-524-5p, hsa-miR-520d-5p and hsa-miR-1283. The T allele matches the predicted three miRNAs seed–binding domain, whereas the C allele represents a mismatch base pairing (Fig. 4A; Supplementary Fig. S4A). A previous study also found that rs11169571 could affect hsa-miR-1283 binding in HA-VSMCs cells (30). By using data from the Genotype-Tissue Expression (GTEx) database, we found that different rs11169571 genotypes were significantly associated with ATF1 expression in 124 sigmoid colon tissues ($P = 6.7 \times 10^{-8}$, effect size $= 0.5$; Fig. 4B). We then cloned ATF1 3′-UTR fragments containing the rs11169571 T or C allele into the luciferase reporter vector psi-check2 to compare the luciferase activity between the two alleles. The results showed that hsa-miR-1283 suppressed luciferase expression more efficiently for the rs11169571T-containing vector than the rs11169571C-containing counterpart in both cell lines (Fig. 4C). More specifically, a declining trend occurred in a concentration-dependent manner for the constructs containing the rs11169571 T allele in both LoVo and SW480 cells (Fig. 4C). A similar result was also observed for hsa-miR-520d-5p (Fig. 4D), but not for hsa-miR-524-5p (Supplementary Fig. S4B).

The ATF1 activates NRAS and BRAF in colorectal cancer cells

As ATF1 is an important transcription factor, we obtained its ChIP-seq data from the Cistrome database and coexpression data from the cbioportal database to find downstream genes activated by ATF1. We found that ATF1-binding peaks were enriched in the promoter region of 905 genes in LoVo cell, with 357 of them showing coexpression with ATF1 in 276 colorectal cancer tissues (Spearman correlation $P < 0.05$ and $r > 0.3$). When intersecting these genes with oncogenes annotated by the Catalogue Of Somatic Mutations In Cancer (COSMIC) database, we found that ATF1 binds in the promoter region of oncogenes NRAS and BRAF (Fig. 4E) and correlates with their RNA expression with correlation values being 0.676 and 0.302, respectively (Fig. 4F). These results indicated that ATF1 may increase the susceptibility to colorectal cancer by activating NRAS and BRAF.

Discussion

In this study, through an exome-wide association analysis, we identified two variants associated with increased risk of colorectal cancer that reached genome-wide significance and one of them (rs138649767) was a rare missense variant with higher effect size (OR = 2.08) than most GWAS-identified common susceptible variants (OR = 1.2–1.5).

This rs138649767 is located in the exon of TCF7L2 (previously known as TCF4), which is a key transcription factor in the Wnt pathway and plays an important role in the development of colorectal cancer (31). Deficiency or mutations of TCF7L2 has been proved to cause colorectal cancer formation and progression (32, 33). The well-known mechanism of TCF7L2 is that it binds to a MYC enhancer containing rs6983267, which is located in a famous GWAS hotpot 8q24 (34, 35). However, no study report whether germline mutations of TCF7L2 could affect its ability to activate MYC. Interestingly, in this study, we found a statistical and functional interaction between the missense TCF7L2 variant and the MYC enhancer variant. In addition, overexpression of the risk allele of rs138649767 significantly enhances colorectal cancer cells’ proliferation. These results highlight an important role of a missense variant of TCF7L2 in the development of colorectal cancer.

The other significant SNP identified in this study is rs11169571 in the 3′-UTR of ATF1. In a recently joint study based on imputation data from six GWAS in European populations, significant association was observed for rs1129406, which is in complete LD with rs11169571 (36). In this study, we found that the rs11169571 signal was the only one in the Chinese population. Each of the four previous European GWAS identified SNPs is in low linkage equilibrium with rs11169571 in Asian populations. Previous Asian colorectal cancer GWAS did not find the rs11169571 variant may due to the different GWAS chip used.

ATF1 encodes a sequence-specific activating transcription factor containing a bZIP DNA binding domain. It influences cellular physiologic processes by regulating the expression of downstream target genes that are related to growth, survival, and other cellular activities. Previous studies have identified that ATF1 is often fused with Ewing sarcoma gene (EWS) and associated with development of clear cell sarcoma (37, 38). In this study, we found that rs11169571 was correlated with ATF1 expression by affecting hsa-miR-1283 and hsa-miR-520d-5p binding. However, there are 124 variants that are in high LD ($r^2 > 0.8$) with rs11169571 and also correlated with ATF1 expression. These variants may also be causal ones that need to be further explored in the future. This study elucidated that the rs11169571 is the only significant signal in the 12q13.12 region in Chinese populations and the ATF1 is an important susceptibility in this region.

One limitation in this study is that only sex and age were adjusted in the regression model and other clinical or demographic factors may affect the assessment of variation risk. The second limitation is that the statistical analyses that were performed for individual polymorphisms were only under additive model (per-allele effect). This model detects both additive and dominant genetic loci effectively, but performs poorly for recessive alleles (39, 40).

In summary, through a systematic analysis of coding variants, we identified two variants significantly associated with the risk of colorectal cancer. Especially, we identified a significant interaction between a rare missense variant in TCF7L2 and a previous GWAS-identified common regulatory variant in MYC enhancer. These results highlight an important role of rare missense variant in the development of colorectal cancer and provide more insights for the prevention of this cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: J. Chang, X. Miao
Development of methodology: J. Tian, B. Zhu, N. Shen
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Tian, Y. Yang, R. Zhong, J. Li, K. Zhai, J. Lou, Y. Zhang, Y. Gong

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Chang, J. Tian, Y. Yang, J. Ke, W. Chen, Y. Zhu

Writting, review, and/or revision of the manuscript: J. Chang, Y. Yang, K. Huang, X. Miao

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Tian, K. Huang

Study supervision: X. Miao

Acknowledgments

This work was supported by the National Key Research and Development Plan Program (2016YFC1302702 to X. Miao); National Natural Science Foundation of China (81525004 to J. Lou); National Key Research and Development Program of China (2016YFC1302702 to X. Miao); and the Fok Ying Tong Foundation for Young Teachers in the Higher Education Institutions of China (131038 to X. Miao).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 26, 2018; revised May 16, 2018; accepted July 10, 2018; published first July 19, 2018.

References


35. Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, et al. The common colorectal cancer predisposition SNP rs6983267 at
Chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet 2009;41:885–90.


A Rare Missense Variant in TCF7L2 Associates with Colorectal Cancer Risk by Interacting with a GWAS-Identified Regulatory Variant in the MYC Enhancer

Jiang Chang, Jianbo Tian, Yang Yang, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-18-0910

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2018/07/19/0008-5472.CAN-18-0910.DC1

Visual Overview
A diagrammatic summary of the major findings and biological implications:
http://cancerres.aacrjournals.org/content/78/17/5164/F1.large.jpg

Cited articles
This article cites 40 articles, 2 of which you can access for free at:
http://cancerres.aacrjournals.org/content/78/17/5164.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/78/17/5164.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/78/17/5164.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.