Identification of a Potential Regulatory Variant for Colorectal Cancer Risk Mapping to 3p21.31 in Chinese Population

Juntao Ke, Jiao Lou, Rong Zhong, Xueqin Chen, Jiaoyuan Li, Cheng Liu, Yajie Gong, Yang Yang, Ying Zhu, Yi Zhang, Jiang Chang & Jing Gong

Genome-wide association studies (GWAS) have established chromosome 3p21.31 as a susceptibility locus for colorectal cancer (CRC) that lacks replication and exploration in the Chinese population. We searched potentially functional single nucleotide polymorphisms (SNPs) in the linkage disequilibrium (LD) block of 3p21.31 with chromatin immunoprecipitation-sequencing (ChIP-seq) data of histone modification, and tested their association with CRC via a case-control study involving 767 cases and 1397 controls in stage 1 and 528 cases and 678 controls in stage 2. In addition to the tag SNP rs8180040 (odds ratio (OR) = 0.875, 95% confidence interval (95% CI) = 0.793−0.966, \( P = 0.008, P\text{-FDR} = 0.040 \)), rs1076394 presented consistently significant associations with CRC risk at both stages with OR = 0.850 (95% CI = 0.771−0.938, \( P = 0.001, P\text{-FDR} = 0.005 \)) under the additive model in combined analyses. Supported by the analyses of data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO), it was suggested that rs1076394 served as an expression Quantitative Trait Loci (eQTL) for gene CCDC12 and NME6, while NME6’s expression was obviously higher in CRC tissues. Using biofeature information such as ChIP-seq and RNA sequencing (RNA-seq) data might help researchers to interpret GWAS results and locate functional variants for diseases in the post-GWAS era.

Colorectal cancer (CRC) was the third most commonly diagnosed cancer in males and the second in females worldwide in 2012, while it was the fifth in males and the third in females, with an estimated 310,244 new cases and 149,722 deaths in China in 2011. Several environmental factors, including diet, physical inactivity, obesity, cigarette smoking and alcohol consumption, were indicated as being involved in the occurrence and development of CRC. On the other hand, it has been well established that genetic factors play an important role in CRC etiology. Revolutionary genome-wide association studies (GWAS) and subsequent fine mapping researches have positioned over 30 susceptibility loci of CRC in Europeans and Asians, however most variants have been found to be only tag single nucleotide polymorphism (SNPs) residing in intergenic and intronic regions without a clear function.

A major challenge in the post-GWAS era is to identify the specific genetic variants that accounts for phenotype based on their functional biology. Recent reports showed that regulatory genome elements can greatly help to identify these causal SNPs, which could exert an effect on gene expression by modulating the activity of promoters, enhancers, insulators and silencers. Today, regulatory genomic regions are usually characterized by various histone modifications in the flanking nucleosomes. For example, enhancers are typically marked by H3K4me1 (histone H3 monomethylated at lysine 4) and promoters by H3K4me3 (histone H3 trimethylated at lysine 4), and both are regarded as active when additionally marked by H3K27ac (histone H3 acetylated at lysine 27). And chromatin immunoprecipitation-sequencing (ChIP-seq) of histone modifications has been widely used to map genome-wide enhancers and promoters.

Fernandez-Rozadilla et al. mapped 3p21.31 as a CRC-relevant genomic locus in a Spanish population with 2362 cases and 2517 controls, with a pooled \( P = 2.163E-06 \) (odds ratio (OR) = 0.784, 95% confidence interval...
Table 1. Candidate regulatory SNPs in chromosome locus 3p21.31. Abbreviations: CHB, Han Chinese in Beijing, China; MAF, minor allele frequency.

| SNP            | Position (hg19) | Major/Minor Allele | CHB MAF | Overlapping Peaks | Histone Modifications | Cell Line |
|----------------|-----------------|---------------------|---------|-------------------|-----------------------|-----------|
| rs2276854      | 47276968        | T/C                 | 0.69    | 47276401–47277264 | H3K4me1               | HCT116    |
| rs807936       | 47322496        | T/C                 | 0.48    | 47322237–47322873 | H3K4me1               | HCT116    |
| rs1076394      | 47322781        | G/A                 | 0.48    | 47322237–47322873 | H3K4me1               | HCT116    |
| rs807936       | 47320000        | T/G                 | 0.48    | 47322603–47325509 | H3K27ac               | HCT116    |
| rs8180040 (tag)| 47388947        | T/A                 | 0.48    | –                 | –                     | –         |

(95% CI) = 0.709–0.867 for the tag SNP rs8180040. Although it didn’t reach the common GWAS significance threshold of 10^-8 and wasn’t included in the larger-scale genetic study by Zhang et al. in East Asians, we considered this locus containing abundant genes as an attractive region to be researched in the Chinese population. Moreover, the strongest risk polymorphism rs8180040 was not in any known transcribed or regulatory sequences and not likely the causal SNP, which meant the real functional SNPs remained mined in 3p21.31.

Using epigenomic data obtained from relevant cell types represents a powerful approach to identifying functional SNPs in post-GWAS genetic researches. In this study, we analyzed ChIP-seq data of histone modifications from CRC cell lines, searched common variants within the regulatory elements of the risk-associated locus 3p21.31, investigated their associations with CRC risk via a two-stage case-control study in the Chinese population, and tried to explain the underlying function by analyzing the data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). To the best of our knowledge, this is the first replication and exploration study on 3p21.31 in East Asians.

Results

Selection of Candidate SNPs. The LD block of GWAS susceptibility loci 3p21.31 was chromosome 3: 47035735-47452118. After a bioinformatics analysis (details in Methods), four common polymorphisms, rs2276854, rs807936, rs1076394 and rs807937, situated within the peaks of histone modification ChIP-Seq data generated from HCT116 and Caco2, were found in the above loci. These four SNPs were in high LD with each other (r^2 > 0.9) according to the 1000 Genomes Project Phase 3 data of the CHB population. Among them, we saw rs1076394 as the most credibly functional variant due to its location in the overlapping region of H3k4me1 and H3k27ac ChIP-seq peaks (Table 1). The coexistence of these two histone modifications is broadly considered as a mark of active enhancer, while the appearance of either one of them is not. For the replication and exploration of 3p21.31, we genotyped the tag SNP rs8180040 and the potential regulatory SNP rs1076394 in Stage 1. We further validated the positive SNPs in another independent sample of Stage 2.

Population Characteristics. Descriptive characteristics of the subjects in this study are detailed in Table 2. In both stages, no significant differences were found between patients and controls in the distribution of sex and age. As expected, significantly more smokers were presented in the cases than in the controls, given that cigarette smoking was a well-established risk factor for CRC. We did not see the same distribution in drinking status.

Association Analysis. Both investigated polymorphisms, the presumably regulatory SNP rs1076394 and tag SNP rs8180040, were significantly associated with CRC risk in both stages and the combined analysis (Table 3).

In Stage 1, under a multivariable logistic regression model adjusted for gender, age, smoking and drinking status, individuals with AA genotype of rs1076394 had a significantly reduced risk of CRC (OR = 0.723, 95% CI = 0.559–0.936, P = 0.014, P-FDR = 0.035) compared to those with GG homozygotes. A dominant model was used to improve statistical power by combining the GA with AA into an A-carrier group (GA plus AA), and it showed that the allele A carriers had an obviously protective effect on CRC susceptibility (OR = 0.802, 95% CI = 0.669–0.963, P = 0.018, P-FDR = 0.03). Likewise, a positive outcome was found in the additive models, with a per-A-allele OR of 0.847 (95% CI = 0.748–0.960, P = 0.009, P-FDR = 0.045). As for rs8180040, we successfully replicated this GWAS tagSNP under the dominant model (OR = 0.811, 95% CI = 0.677–0.970, P = 0.022, P-FDR = 0.025) with nominal significance, but failed after FDR corrections.

Two promising variants were both further genotyped in the validation Stage 2. In agreement with the Stage 1, nominal significant associations were still exhibited between CRC risk and rs1076394 (dominant model: OR = 0.745, 95% CI = 0.584–0.951, P = 0.018, P-FDR = 0.09; additive model: OR = 0.833, 95% CI = 0.708–0.980, P = 0.027, P-FDR = 0.068), or rs8180040 (dominant model: OR = 0.747, 95% CI = 0.586–0.952, P = 0.018, P-FDR = 0.09; additive model: OR = 0.837, 95% CI = 0.713–0.983, P = 0.030, P-FDR = 0.075). When we combined two stages, positive results were still observed after FDR corrections (Table 3). And two polymorphisms were in high LD (r^2 = 0.895) with each other in our total samples, similar to the data of 1000 Genomes Phase 3 of CHB (r^2 = 1.000).

The results of interaction analysis between the promising SNP rs1076394 and smoking were detailed in Table S2, where no significant interactions were observed under either the multiplicative or the additive model in the two stages and the combined study.
TCGA and GEO Data Analyses. We downloaded the data of gene expression, germline genotypes, CpG methylation and somatic copy number for COAD (colorectal adenocarcinoma) and READ (rectal adenocarcinoma) from the TCGA portal (http://cancergenome.nih.gov/) up to October 2014. Then, we performed a modified eQTL (expression Quantitative Trait Loci) analysis of the correlation between rs8180040 (rs1076394 wasn’t included in the Affymetrix GenomeWide SNP 6.0 Array of genotype profiles) and expression of genes within 1 Mb flanking regions, with the effects of somatic copy number and methylation being adjusted. As shown in Table 4 and Fig. 1, rs8180040 was identified as an eQTL for genes NME6 (nucleoside diphosphate kinase 6, $R^2 = 0.019$, $P = 0.029$, $P$-FDR = 0.404) and CCDC12 (coiled-coil domain-containing protein 12, $R^2 = 0.031$,
As the presumed regulatory variant rs1076394 was in complete LD with rs8180040 in the 1000 Genomes database, rs1076394 was also indicated as being closely related to the expression of CCDC12 and NME6. In addition, we compared the two genes' expression between cancer and normal tissue, and found a significant difference for NME6 (P = 0.029; CRC tissues: 229.5 ± 3.2 RPKM, peritumoral tissue: 210.5 ± 5.4 RPKM), but not for CCDC12 (P = 0.258; CRC tissues: 707.7 ± 15.3 RPKM, peritumoral tissue: 661.6 ± 16.4 RPKM).

In GEO database, the datasets of expression profiles in Asian samples (Dataset Records: GDS2609 and GDS4718) were provided, while TCGA data was mostly from Caucasian samples. We compared the NME6 expression correlation between rs8180040 and flanking 1 Mb genes. Abbreviations: FDR, false discovery rate. The nominal significant results were in bold.

| Gene   | Correlation R2 | Correlation P | Correlation P-FDR |
|--------|----------------|---------------|-------------------|
| CCDC12 | 3.091E-02      | 4.952E-03     | 1.387E-01         |
| NME6   | 1.882E-02      | 2.884E-02     | 4.038E-01         |
| PTH1R  | 8.257E-03      | 1.487E-01     | 1.000             |
| CDC25A | 8.041E-03      | 1.542E-01     | 9.180E-01         |
| LTF    | 7.674E-03      | 1.639E-01     | 1.000             |
| MYL3   | 7.498E-03      | 1.689E-01     | 7.881E-01         |
| LRR2C2 | 6.062E-03      | 2.162E-01     | 8.649E-01         |
| DHX30  | 5.517E-03      | 2.382E-01     | 8.336E-01         |
| CAMP   | 5.229E-03      | 2.509E-01     | 7.805E-01         |
| CCR2   | 4.858E-03      | 2.684E-01     | 7.517E-01         |
| SMARCC1| 3.693E-03      | 3.348E-01     | 8.521E-01         |
| RTP3   | 2.954E-03      | 3.883E-01     | 9.061E-01         |
| SETD2  | 2.748E-03      | 4.055E-01     | 8.734E-01         |
| CCRL2  | 2.407E-03      | 4.362E-01     | 8.725E-01         |
| KIP9   | 2.382E-03      | 4.387E-01     | 8.189E-01         |
| NREAL2 | 1.133E-03      | 5.933E-01     | 1.000             |
| TDGF1  | 1.117E-03      | 5.959E-01     | 9.815E-01         |
| PRSS50 | 1.071E-03      | 6.037E-01     | 9.390E-01         |
| NRADD1| 6.032E-04      | 6.969E-01     | 1.000             |
| MAP4   | 5.666E-04      | 7.073E-01     | 9.902E-01         |
| PTP2N3 | 4.698E-04      | 7.310E-01     | 9.747E-01         |
| KLHL18 | 3.567E-04      | 7.645E-01     | 9.730E-01         |
| ALS2C2 | 3.426E-04      | 7.691E-01     | 9.363E-01         |
| ZNFS18 | 3.120E-04      | 7.794E-01     | 9.093E-01         |
| SCAP   | 3.056E-04      | 7.816E-01     | 8.754E-01         |
| TMIE   | 2.218E-04      | 8.133E-01     | 8.758E-01         |
| CSPG5  | 5.248E-05      | 9.085E-01     | 9.422E-01         |
| CCR5   | 4.253E-05      | 9.176E-01     | 9.176E-01         |

Table 4. Expression correlation between rs8180040 and flanking 1 Mb genes. Abbreviations: FDR, false discovery rate. The nominal significant results were in bold.

Figure 1. Association between rs8180040 and NME6/CCDC12 expression. (A) Association between rs8180040 and NME6 expression. (B) Association between rs8180040 and CCDC12 expression.

P = 0.005, P-FDR = 0.139, respectively. As the presumed regulatory variant rs1076394 was in complete LD with rs8180040 in the 1000 Genomes database, rs1076394 was also indicated as being closely related to the expression of CCDC12 and NME6. In addition, we compared the two genes' expression between cancer and normal tissue, and found a significant difference for NME6 (P = 0.029; CRC tissues: 229.5 ± 3.2 RPKM (reads per kilobases per million reads), peritumoral tissue: 210.5 ± 5.4 RPKM), but not for CCDC12 (P = 0.258; CRC tissues: 707.7 ± 15.3 RPKM, peritumoral tissue: 661.6 ± 16.4 RPKM).
expression between cancerous and normal colon tissues, and found the same higher expression in cancer tissues from the Chinese population ($P = 0.017$; CRC tissues: $585.8 \pm 29.3$ RPKM, normal colon tissue: $482.2 \pm 23.6$ RPKM), and the Japanese population ($P = 0.013$; CRC tissues: $973.6 \pm 79.9$ RPKM, normal colon tissue: $771.4 \pm 31.0$ RPKM). As for CCDC12, we could not observe significant differences between cancer and normal tissues in two datasets (Chinese samples: $P = 0.583$, CRC tissues: $2219.9 \pm 147.9$ RPKM, normal colon tissue: $2118.0 \pm 91.1$ RPKM; Japanese samples: $P = 0.062$, CRC tissues: $22137.9 \pm 971.7$ RPKM, normal colon tissue: $24648.9 \pm 870.9$ RPKM).

**Discussion**

The identification of tag SNPs through GWAS is the important first step in understanding the relationship between genomic variation and CRC risk. But now, the foremost goal in the post-GWAS era is to shed light on the causal SNPs and their functional consequences, progressing from indirectly statistical to directly biological associations between genetic variation and disease. Accumulating evidence showed that the most likely mechanistic basis that links those noncoding genetic variants to phenotype and disease is being regulatory.

In our study, by overlapping the LD boundaries of the locus 3p21.31 and regulatory regions predicted by CRC-specific histone modifications, we screened out the most promisingly functional rs1076394 among the four original polymorphisms in high LD. By conducting association studies in two independent Chinese populations containing 127 cases and 2075 controls in total, we replicated the significances of tag SNP rs8180040, and found a significant protective effect for the potentially regulatory variant rs1076394 that might serve as an eQTL for the genes CCDC12 and NME6, while NME6 presented significantly higher expression in cancer tissues.

The findings led us to assume that rs1076394 might influence CRC risk by altering the activity of an enhancer that controlled NME6 expression. The potentially functional variant rs1076394 lay within a region of the genome exhibiting chromatin modifications H3k4me1 and H3k27ac in a CRC cell line HCT116, consistent with characteristics of an active enhancer across diverse tissues. It was situated in the first intron of gene KIF9 (kinesin family member 9), which was involved in mitotic progression by maintaining correct spindle length, and the degradation of the matrix by regulating macrophage podosomes. However, according to our calculation of TCGA data, rs1076394 was related to the expression of CCDC12 and NME6, but not related to KIF9. We saw the SNP as a potential eQTL for CCDC12 and NME6, while we could not rule out the possibility that it was actually in LD with these two genes. The SNP rs1076394 was approximately 300kb upstream of CCDC12, which participated in promoting early erythroid differentiation, and over 1000 kb upstream of NME6, which might play a role in the regulation of cell growth and cell cycle progression. Either CCDC12 or NME6 was located in 3p21.3. But, NME6 was more suggested to be the actual contributing gene in this locus due to its significantly higher expression in cancer samples of both Caucasian and Asian populations. At the beginning, NME6 was discovered as a gene encoding a nucleoside diphosphate kinase that suppressed p53-induced apoptosis. In a shRNA functional screen, Nmeh was reported to be crucial for the renewal of embryonic stem cells (ESCs). And ESCs are characterized by immortalization ability, pluripotency, and oncogenicity. More recently, enriched somatic mutations in NME6 was suggested to be associated with the deregulation of pyrimidine metabolism and the promotion of malignant progression in human melanoma. Accordingly, NME6 may act as an oncogene that still needs more investigations.

The application of biofeature information such as ChIP-seq and RNA-seq data has represented an effective approach to identifying functionally regulatory SNPs, and different databases including UCSC, Encode, GEO and TCGA have provided easy access to massive amounts of relevant data. Incorporating epigenetic and expression analyses into traditional molecular epidemiology could assist in the interpretation of GWAS results and the discovery of functional variants for diseases in post-GWAS studies. Using a similar strategy to other CRC-associated loci should deepen our understanding of CRC risk.

Nevertheless, several limitations should be noted here. First of all, due to the lack of relevant functional experiments, biological reality beneath the statistically significant association is uncertain. In the analysis of TCGA data, we have not restricted it to Chinese samples, when its composition is mostly Caucasian samples. It may not reflect the exact outcome of the Chinese population that we researched on. Second, the strategy of retrieving candidate polymorphisms depended on the prediction from ChIP-seq data of relevant cell lines, which was not rigorous enough to define exact regulatory elements and all the functional variants inside. Focusing on common SNPs, we could not rule out the possibility that sets of rare variants or haplotypes in LD with the tag SNP are actually causal enough to define exact regulatory elements and all the functional variants inside. Focusing on common SNPs, we could not rule out the possibility that sets of rare variants or haplotypes in LD with the tag SNP are actually causal.

In summary, we discovered a probably regulatory SNP that was associated with CRC risk in the Chinese population. Researches on 3p21.31 and other CRC susceptibility loci with greater sample sizes and follow-up functional analyses are warranted to elaborate the biological mechanism of genetic etiology.

**Methods**

**Study Participants.** A two-stage case-control study was applied to evaluate the association between candidate variations and the risk of CRC. The discovery stage (Stage 1) consisted of 767 cases and 1397 controls, which were recruited from Tongji Hospital of Huazhong University of Science and Technology (HUST) between 2008 and 2012. The validation stage (Stage 2) involved 528 cases and 678 controls enrolled from 2013 to 2015 at the same hospital. All subjects were unrelated ethnic Han Chinese in both stages. Patients with histopathologically confirmed CRC and without previous chemotheraphy or radiotherapy, were included without restriction to gender and age. In the same time period, cancer-free controls were recruited form participants in physical examination programs of the same hospital, and were adequately matched to cases by gender and age ($\pm 5$ years). Definitions of smoking and drinking status were the same as in a previous study by our group. At recruitment of each subject, a written informed consent was obtained, and 2 millimeters peripheral venous blood was collected. This
study was approved by the ethics committee of Tongji Medical College of Huazhong University of Science and Technology, and the methods were carried out in accordance with the approved guidelines.

**SNP Selection and Genotyping.** Candidate SNPs were common genetic variants (minor allele frequency, MAF > 0.05) that located in the putative regulatory elements of the 3p21.31 locus. First, we applied the software HaploView to calculate the linkage disequilibrium (LD) block of 3p21.31 with the criterion of $r^2 > 0.8$, by inputting the Chinese Han Beijing (CHB) genotype information of 500 kb flanking the tagSNP rs8180040. This LD block was defined as the CRC susceptibility locus. Second, we downloaded ChIP-seq data regarding histone modification from two CRC cell lines, HCT116 and Caco2, form the UCSC database integrated with Encode data (S1 Table). And the extents of their signal peaks were considered as putative regulatory elements. Third, based on the CHB MAF data in dbSNP database, we only selected the common polymorphisms (MAF > 0.05) that situated within the overlapping parts between the LD block and peaks. Finally, four polymorphisms in high LD with each other ($r^2 > 0.9$) survived after this step-wise analysis. Among them, rs1076394 was chosen as the most potentially functional variant for the genotyping assays, because of its more indicative location in an active enhancer than others. At the same time, we also tried to replicate the tag SNP rs8180040 in our sample. In Stage 2, the nominal significant SNPs of Stage 1 were further validated. SNPs of both stages were genotyped by a TaqMan real-time polymerase chain reaction (PCR) assay (Applied Biosystems, Foster city, CA). Quality control was preformed by including 5% duplicate samples in blinded fashion, with a concordance rate of 100%.

**Statistical Analysis.** The differences in the distributions of gender, age, smoking, drinking status and genotypes between cases and controls were estimated by a χ² test or t-test, where appropriate. The Hardy-Weinberg equilibrium (HWE) in controls was evaluated with a goodness-of-fit χ² test. The odds ratio (ORs) and corresponding 95% confidence intervals (95% CIs) were used to measure the associations between SNPs and CRC susceptibility. And they were calculated after adjusting for gender, age, smoking and drinking status under a multivariate logistic regression model. For multiple comparison corrections, a simple procedure (Benjamini and Hochberg) was performed in two stages and combined study to control the false discovery rate. The LD of the candidate SNPs was analyzed using HaploView v4.2. With regard to TCGA and GEO data, the expression differences among three genotypes (TT, TA and AA) of rs8180040 were measured under a linear regression model adjusting the effects of somatic copy number and CpG methylation, and the differences between cancer and normal samples were measured by t-test. The gene-environment interactions were evaluated by a pair-wise analysis under multiplicative and additive interaction models. All the above statistical analyses were conducted using SPSS Software v20.0 (SPSS, Chicago, Illinois, USA), with the exception that the P values of additive interaction were assessed using Stata v11.0 (Stata Corporation, College Station, TX). P values in this study were two-sided with a significance criterion of $P < 0.05$.

**References**
1. Torre, L. A. et al. Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians 65, 87–108, doi:10.3322/caac.21262 (2015).
2. Chen, W., Zheng, R., Zhang, S. & He, J. Annual report on status of cancer in China, 2011. Chin J Cancer Res 27, 2–12, doi:10.3978/j.issn.1000-9604.2015.01.06 (2015).
3. Haggard, F. & Boushey, R. Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors. Clinics in Colon and Rectal Surgery 22, 191–197, doi:10.1055/s-0029-1242458 (2009).
4. Botteri, E. et al. Smoking and colorectal cancer: a meta-analysis. JAMA 300, 2765–2778, doi:10.1001/jama.2008.839 (2008).
5. Zhu, B., Zou, L., Qu, L., Zhong, R. & Xiao, X. Allium Vegetables and Garlic Supplements Do Not Reduce Risk of Colorectal Cancer, Based on Meta-analysis of Prospective Studies. Clinical Gastroenterology and Hepatology 12, 1991–2001.e1994, doi:10.1016/j.cgh.2014.03.019 (2014).
6. Lichtenstein, P. et al. 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, doi:10.1056/NEJM200007133434021 (2000).
7. de la Chapelle, A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 4, 769–780, doi:10.1038/nrc1453 (2004).
8. Gong, J. et al. A functional polymorphism in lnc-LAMC2-1:1 confers risk of colorectal cancer by affecting miRNA binding. Carcinogenesis, doi:10.1093/carcin/bgw024 (2016).
9. Zanke, B. W. et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 39, 989–994, doi:10.1038/ng2089 (2007).
10. Tomlinson, I. et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 39, 894–900, doi:10.1038/ng2085 (2007).
11. Broderick, P. et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 39, 1315–1317, doi:10.1038/ng.2007.18 (2007).
12. Jaeger, E. et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 40, 26–28, doi:10.1038/ng.2007.41 (2008).
13. Tenesa, A. et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 1p12 which replicates risk loci at 8q24 and 18q21. Nat Genet 40, 631–637, doi:10.1038/ng.133 (2008).
14. Tomlinson, I. P. et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 40, 623–630, doi:10.1038/ng.111 (2008).
15. Houlston, R. S. et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 40, 1426–1435, doi:10.1038/ng.262 (2008).
16. Houlston, R. S. et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 1q21.33 and 20q13.3. Nat Genet 42, 973–977, doi:10.1038/ng.670 (2010).
17. Tomlinson, I. P. et al. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. PLoS Genet 7, e1002105, doi:10.1371/journal.pgen.1002105 (2011).
18. Dunlop, M. G. et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet 44, 770–776, doi:10.1038/ng2293 (2012).
19. Peters, U. et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology 144, 799–807 (2013).
20. Jia, W. H. et al. Genome-wide association studies in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet 45, 191–196, doi:10.1038/ng2505 (2013).
Acknowledgements
The authors wish to thank all the study participants, research staff and students who participated in this work.

Author Contributions
Conception and design of the experiments: J.G. Conduct of the experiments: J.K. and J.L. DNA extraction: R.Z., X.C., J.L., C.L., Y.G., Y.Y., Ying, Z. and Yi, Z. Analysis and interpretation the data: J.K. Writing of the paper: J.K. Administrative, technical, and material support: J.C. and J.G. All authors reviewed the manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ke, J. et al. Identification of a Potential Regulatory Variant for Colorectal Cancer Risk Mapping to 3p21.31 in Chinese Population. Sci. Rep. 6, 25194; doi: 10.1038/srep25194 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/