The Associations between RNA Splicing Complex Gene *SF3A1* Polymorphisms and Colorectal Cancer Risk in a Chinese Population

Xiaohua Chen[1,2☯], Hua Du[2☯], Binjian Liu[2], Li Zou[3], Wei Chen[3], Yang Yang[3], Ying Zhu[3], Yajie Gong[3], Jianbo Tian[3], Feng Li[1], Shan Zhong[1,4]*

1 Department of Medical Genetics, School of Basic Medical Science, Wuhan University, Wuhan, 430071, China, 2 Department of Laboratory Medicine, No 161 Hospital of PLA, Wuhan, 430010, China, 3 Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, 4 Hubei Provincial Key Laboratory of Developmentally Originated Disease, Wuhan, 430071, China

☯ These authors contributed equally to this work.
* zhongshan@whu.edu.cn

Abstract

Background

Aberrant alternative splicing included alterations in components of the mRNA splicing machinery often occurred in colon cancer. However, the role of *SF3A1*, one key component of the mRNA splicing machinery, on colorectal cancer (CRC) risk was still not elucidated.

Method and Findings

We performed a hospital-based case-control study containing 801 CRC patients and 817 cancer-free controls to examine the association between *SF3A1* polymorphisms and CRC risk in a Chinese population. Four candidate SNPs (rs10376, rs5753073, rs2839998 and rs2074733) were selected based on bioinformatics analysis and previous findings. The results showed no significant associations between these SNPs and CRC risk ($P > 0.05$). Besides, the stratified analysis based on the smoking and alcohol use status obtained no statistically significant results.

Conclusion

Our study was the first one to investigate the association between *SF3A1* polymorphisms and CRC risk. The results suggested these four SNPs in *SF3A1* were not associated with CRC risk in a Chinese population, however, further more studies are needed to confirm our findings.

Introduction

Colorectal cancer (CRC) remains a major health problem and is a leading cause of morbidity and mortality worldwide, representing the third most common cause of cancer-related death [1]. It was estimated to cause 142,820 new cases and 50,830 deaths of the colon and rectum cancer in the United States for both men and women in 2013[2]. Although several environmental risk factors [3,4] have been detected to be associated with risk of CRC, genetic susceptibility was found to be involved in the development of this disease. Genome-wide association studies (GWAS) have been successful applied in identifying susceptibility loci for cancer and other diseases [5,6]. In colorectal cancer, recent GWAS studies have revealed more than 20 susceptibility single nucleotide polymorphisms (SNPs) in multiple different loci in European and Asian populations [7–20]. However, most of them are located in non-coding regions and can explain less than 10% of the familial relative risk of CRC in European populations [13,14]. These indicated that there may be a substantial fraction of genetic components undiscovered and the biological mechanisms are required to be explored.

RNA splicing can remove introns from pre-messenger RNAs and is essential to all eukaryotic organisms to generate considerable numbers of alternative isoforms with altered coding potential or regulatory regions in order to guarantee the functional diversity of their protein in the face of a limited number of genes[21,22]. However, aberrant alternative splicing resulted from mutations within splicing elements in cancer genes or transcripts from non-mutated genes occurred in many cancers [23]. For example, a few studies have investigated the aberrant alternative splicing in colon cancer and detected many colon cancer specific alternative splicing events affecting several proteins or pathways[24–28]. These colon cancer-related splicing events often involved alterations in components of the mRNA splicing machinery, which was exemplified by the recent finding that amplification or overexpression of *PRPF6* could be a driver of colon tumorigenesis[29].

RNA splicing is a well-ordered process that recruits, rearranges and disengages of a set of small nuclear ribonucleoprotein (snRNP) complexes, as well as many other protein components onto the pre-mRNAs. During splicing, SF3A1, together with the U2 snRNP and other proteins, are recruited to the 3' splicing site to generate the splicing complex A after the recognition of the 3' splicing site[30].Therefore, *SF3A1* is critical for spliceosome assembly and normal splicing events. *SF3A1* is located in 22q12.2, where has been reported to be associated with susceptibility of lung cancer [31], breast cancer [32] and inflammatory bowel disease [33] by genome-wide association studies. Several studies have reported the associations between mutations of *SF3A1* and other diseases. Yoshida et al[30] have recently discovered lower mutational rates for *SF3A1* in the majority of the patients with myelodysplastic syndromes (MDS). Additionally, curated information from the Catalogue of Somatic Mutations in Cancer (COSMIC) database revealed that mutations in the coding-region of *SF3A1* were associated with several cancers, including esophageal adenocarcinoma, myxoid liposarcomas, synovial sarcomas, osteosarcomas, endometrial tumors, lung cancer, breast cancer, ovarian carcinoma, gastric cancer and glioblastoma. Collectively, these findings suggested the link between *SF3A1* and cancer risk, and emphasized a need of additional researches for the association of *SF3A1* polymorphisms and CRC risk.

Considering the common occurrence of aberrant alternative splicing in CRC and the role of *SF3A1* in alternative splicing, we hypothesized the polymorphisms of *SF3A1* might also contribute to the susceptibility of CRC. In the present study, we carried out a hospital-based case-control study in a Chinese population to investigate the association between polymorphisms of *SF3A1* and CRC risk.

Materials and Methods

Ethics Statement

At the recruitment, written informed consent was obtained from each subject. The personal information about sex, birth year, smoking and drinking habits of all participants were also collected by interviews. Meanwhile, 5ml peripheral blood sample from each subject was collected and stored in the -80°C refrigerator before DNA extraction. This study was approved by ethics committee of Tongji Hospital of Huazhong University of Science and Technology.

Study participants

A total of 801 CRC cases and 817 cancer-free controls were investigated in this study, all of whom were unrelated ethnic Han Chinese living in Wuhan region. Patients who had been histopathologically confirmed with primary colorectal cancer were enrolled from Tongji Hospital between 2009 and 2013, and had not received radiotherapy or chemotherapy before blood samples collection, part of which were described in our previous studies[34–36]. Cancer-free controls were selected from physical examination participants in the same hospital and during the same period without any history of cancer or seriously chronic disease. The control subjects were frequency matched to CRC patients by age (±5 years) and gender.

Identification of candidate SNPs

The candidate SNPs were identified based on bioinformatics analysis and related findings. The screening procedure was described as follows. First, we input the gene "*SF3A1*" into a web-based bioinformatics tool "SNPinfo—SNP Function Prediction" (http://snpinfo.niehs.nih.gov/snpinfo/snpfunc.htm) with allele frequency restricted to "CHB" and "Asian" populations. The tool integrates GWAS and candidate gene information to predict functional characteristics of both non-coding and coding SNPs, such as transcription-factor-binding site (TFBS), micro-RNA-binding site, splice site, regulatory potential score, Polyphen and so on. As a result, 32 SNPs with MAF of CHB or Asian > 5% were retrieved, among which only rs10376 and rs5753073 were predicted to locate in the binding sites of microRNA with the miRanda scores for two alleles differed by ≥ 16. Then, another bioinformatics tool "miRNA SNP v2.0" (http://bioinfo.life.hust.edu.cn/miRNASNP2/) was adopted to predict the function of rs2839998 because of its ambiguous result by "SNPinfo—SNP Function Prediction". As expected, the SNP was also predicted to be a microRNA-binding site. Besides, the SNP of rs2074733 was revealed to be associated with increased risk of pancreatic cancer in our previous study. Therefore, four SNPs (rs10376, rs5753073, rs2839998 and rs2074733) were finally selected in our study.

DNA isolation and genotyping

5ml peripheral blood sample from each case and control subject was used to isolate genomic DNA by using RelaxGene Blood System DP319-02 (Tiangen, Beijing, China) according to the manufacturer's instructions. The 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster city, CA) was applied to determine the genotypes of rs2074733, rs10376, rs2839998 and rs5753073 by using the TaqMan SNP Genotyping Assay. (Applied Biosystems, Foster City, CA). Amplification was done under the following conditions: 95°Cfor 10 min followed by 45 cycle of 94°C for 30 s and 62°C for 1 min. Data were analyzed using Allelic Discrimination Program (Applied Biosystems). In addition, we randomly selected 5% samples and genotyped twice as quality control with a concordance rate of 100%.

Statistical analysis

To evaluate the differences in distribution of sex, age, smoking, drinking status and genotypes between case and control groups, Pearson χ^2 test and t test were employed where appropriate. Hardy–Weinberg equilibrium for genotypes was tested in controls by a goodness-of-fit χ^2-test. The associations between rs2074733, rs10376, rs2839998 and rs5753073 and CRC risk were evaluated by the OR and its 95% confidence interval (95% CI) using unconditional logistic regression analysis adjusted by age, sex, smoking habit and alcohol use. SPSS 12.0 software was used to perform the two sided statistical analyses and $P < 0.05$ was considered statistically significant.

Results

Characteristics of the study population

In the present study, we have recruited 801 CRC patients and 817 cancer-free healthy individuals, whose characteristics are shown in **Table 1**. The mean age of patients and controls were 58.18 years (±11.68) and 57.51 years (±11.44), respectively. Among cases 58.4% were males compared with 58.0% among controls. Additionally, 35.3% cases were smokers and 31.6% cases with drinking habits. There were no significant differences in the distribution of age ($P = 0.244$), sex ($P = 0.867$), smoking habit ($P = 0.122$) and alcohol use ($P = 0.453$) between case and control groups.

Association analysis

Genotype distributions of the *SF3A1* polymorphisms in the CRC patients and control individuals are shown in **Table 2**. The genotypes of the rs5753073, rs2839998, rs10376 and rs2074733 in controls conformed to Hardy-Weinberg equilibrium (HWE) with P values of 0.802, 0.734, 0.668 and 0.208, respectively. The logistic regression analysis showed no significant associations between the heterozygote and homozygote variations of all these 4 SNPs and CRC risk after adjusted by age, sex, smoking habit and alcohol use. For example, individuals with rs5753073 GG or GA genotype showed similar CRC risk compared to those with the AA genotype (OR = 0.73, 95% CI: 0.37–1.44 and OR = 0.86, 95% CI: 0.68–1.08). Similarly, there were

Table 1. Distribution of characteristics among cases and controls.

Variables	Case (N = 801) No. (%)	Control (N = 817) No. (%)	*P*
Age (years)	58.18±11.68	57.51±11.44	0.244 [a]
Sex			0.867 [b]
Male	468(58.4)	474(58.0)	
Female	333(41.6)	343(42.0)	
Smoking			
Yes	283 (35.3)	259 (31.7)	0.122 [b]
No	518 (64.7)	558 (68.3)	
alcohol use			
Yes	253 (31.6)	244 (29.9)	0.453 [b]
No	548 (68.4)	573 (70.1)	

a: *P* value was calculated by *t* test

b: *P* value was calculated by χ^2 test

doi:10.1371/journal.pone.0130377.t001

no differences in the genotype distributions of rs2839998, rs10376 and rs2074733 between cases and controls.

We next stratified our study subjects to investigate the relationship of the *SF3A1* polymorphisms with smoking and alcohol use status. However, we still observed no significant associations between these SNPs and CRC risk in smoking and alcohol use subgroups (*P*>0.05) (**Tables 3 and 4**).

Table 2. Association between selected polymorphisms and CRC risk.

SF3A1-genotype	Case (N = 801) No. (%)	Control (N = 817) No. (%)	*P*-HWE	MAF in 1000 genomes project	MAF in Case	MAF in Control	*P* [a]	*P* [b]	OR [b](95% CI)
rs5753073			0.802	0.120	0.134	0.153			
AA	594 (75.0)	584 (71.8)					0.325		1.00
AG	183 (23.1)	209 (25.7)						0.183	0.86 (0.68–1.08)
GG	15 (1.9)	20 (2.5)						0.361	0.73 (0.37–1.44)
A/G							0.132	0.119	0.85 (0.70–1.04)
Additive[c]								0.120	0.86 (0.70–1.04)
rs2839998			0.734	0.350	0.337	0.315			
GG	340 (44.6)	368 (46.7)					0.288		1.00
GA	330 (43.3)	344 (43.7)						0.715	1.04 (0.84–1.29)
AA	92 (12.1)	76 (9.6)						0.105	1.32 (0.94–1.86)
G/A							0.180	0.166	1.11 (0.96–1.29)
Additive								0.168	1.11 (0.96–1.29)
rs10376			0.668	0.070	0.093	0.105			
CC	658 (83.0)	640 (80.2)					0.256		1.00
AC	123 (15.5)	148 (18.5)						0.124	0.81 (0.63–1.06)
AA	12 (1.5)	10 (1.3)						0.760	1.14 (0.49–2.66)
C/A							0.235	0.241	0.87 (0.69–1.10)
Additive								0.252	0.88 (0.70–1.10)
rs2074733			0.208	0.470	0.467	0.467			
TT	221 (27.9)	221 (27.4)					0.884		1.00
TC	402 (50.8)	420 (52.0)						0.669	0.95 (0.75–1.20)
CC	169 (21.3)	167 (20.7)						0.917	1.02 (0.76–1.35)
T/C							0.973	0.962	1.00 (0.87–1.15)
Additive								0.962	1.00 (0.87–1.16)

a: *P* values were calculated by χ^2 test

b: Adjusted by age, sex, smoking and alcohol use

c: Additive model

doi:10.1371/journal.pone.0130377.t002

Table 3. Subgroup analysis according to the smoking status.

SF3A1-genotype	No Smokers				Smokers			
	Case (N = 518) No. (%)	Control (N = 558) No. (%)	P [a]	OR (95% CI)	Case (N = 283) No. (%)	Control (N = 259) No. (%)	P [a]	OR (95% CI)
rs5753073								
AA	387 (75.3)	406 (73.0)		1.00	207 (74.5)	178 (69.3)		1.00
AG	116 (22.6)	138 (24.8)	0.486	0.90 (0.68–1.20)	67 (24.1)	71 (27.6)	0.277	0.80 (0.54–1.19)
GG	11 (2.1)	12 (2.2)	0.883	0.94 (0.41–2.16)	4 (1.4)	8 (3.1)	0.141	0.40 (0.12–1.36)
rs2839998								
GG	219 (44.4)	254 (47.1)		1.00	121 (45.0)	114 (45.8)		1.00
GA	215 (43.6)	229 (42.5)	0.514	1.09 (0.84–1.42)	115 (42.8)	115 (46.2)	0.744	0.94 (0.65–1.36)
AA	59 (12.0)	56 (10.4)	0.339	1.22 (0.81–1.84)	33 (12.3)	20 (8.0)	0.169	1.55 (0.83–2.87)
rs10376								
CC	424 (82.8)	436 (79.4)		1.00	234 (83.3)	204 (81.9)		1.00
AC	82 (16.0)	106 (19.3)	0.149	0.79 (0.58–1.09)	41 (14.6)	42 (16.9)	0.560	0.87 (0.54–1.40)
AA	6 (1.2)	7 (1.3)	0.776	0.85 (0.28–2.57)	6 (2.1)	3 (1.2)	0.489	1.65 (0.40–6.74)
rs2074733								
TT	148 (28.9)	156 (28.4)		1.00	73 (26.1)	65 (25.2)		1.00
TC	250 (48.8)	281 (51.1)	0.714	0.95 (0.71–1.26)	152 (54.3)	139 (53.9)	0.779	0.94 (0.63–1.42)
CC	114 (22.3)	113 (20.5)	0.683	1.08 (0.76–1.52)	55 (19.6)	54 (20.9)	0.625	0.88 (0.53–1.47)

a: Adjusted by age, sex and alcohol use

doi:10.1371/journal.pone.0130377.t003

Discussion

In the present study, we hypothesized that polymorphisms of *SF3A1* might contribute to the genetic susceptibility of CRC. Since the RNA splicing system is indispensable for functional diversity of protein and gene control in eukaryotic organisms, dysregulation of such machinery, including mutations of the core genes, can cause various diseases and cancers [37,38]. As a member of splicing complex, *SF3A1* has been implicated in various cancers. We thus proposed that polymorphisms might affect the *SF3A1* expression, then resulting in aberrant alternative splicing events in CRC.

The four candidate polymorphisms selected based on bioinformatics analysis and previous findings are located within non-coding regions of *SF3A1*. More than 1,200 GWAS studies have detected nearly 6,500 susceptibility loci[39] and it is noteworthy that 93% of which are located within non-coding regions[5]. Some of SNPs located within regulatory non-coding regions can affect gene expression and are major components in complex disease predisposition [40]. Among the four polymorphisms, rs10736, rs5753073 and rs2839998 are located in the 3'untranslated region (UTR), where microRNA binds and regulates the mRNA expression. These bindings can be affected by SNPs that reside in the microRNA target site, which can either abolish existing binding sites or create illegitimate binding sites, having a different effect on gene expression and representing another type of genetic variability that can influence the risk of certain human diseases[41]. Accumulative evidences have revealed that polymorphisms

Table 4. Subgroup analysis according to the drinking status.

SF3A1-genotype	No Drinkers				Drinkers			
	Case (N = 548) No. (%)	Control (N = 573) No. (%)	P [a]	OR (95% CI)	Case (N = 253) No. (%)	Control (N = 244) No. (%)	P [a]	OR (95% CI)
rs5753073								
AA	414 (76.2)	416 (73.0)		1.00	180 (72.3)	168 (69.1)		1.00
AG	118 (21.7)	141 (24.7)	0.464	0.86 (0.57–1.30)	65 (26.1)	68 (28.0)	0.235	0.84 (0.64–1.12)
GG	11 (2.0)	13 (2.3)	0.171	0.41 (0.12–1.46)	4 (1.6)	7 (2.9)	0.663	0.83 (0.37–1.89)
rs2839998								
GG	241 (46.1)	255 (46.2)		1.00	99 (41.4)	113 (47.9)		1.00
GA	222 (42.4)	243 (44.0)	0.255	1.26 (0.85–1.86)	108 (51.7)	101 (42.8)	0.786	0.97 (0.75–1.24)
AA	60 (11.5)	54 (9.8)	0.060	1.82 (0.98–3.41)	32 (13.4)	22 (9.3)	0.469	1.16 (0.77–1.75)
rs10376								
CC	448 (82.7)	447 (79.5)		1.00	210 (83.7)	193 (81.8)		1.00
AC	87 (16.1)	109 (19.4)	0.558	0.86 (0.52–1.43)	36 (14.3)	39 (16.5)	0.142	0.79 (0.58–1.08)
AA	7 (1.3)	6 (1.1)	0.942	0.95 (0.25–3.62)	5 (2.0)	4 (1.7)	0.812	1.14 (0.38–3.43)
rs2074733								
TT	166 (30.6)	156 (27.6)		1.00	55 (22.1)	65 (26.9)		1.00
TC	265 (48.8)	294 (51.9)	0.366	1.23 (0.79–1.92)	137 (55.0)	126 (52.1)	0.240	0.85 (0.64–1.12)
CC	112 (20.6)	116 (20.5)	0.284	1.34 (0.79–2.29)	57 (22.9)	51 (21.1)	0.550	0.90 (0.64–1.27)

a: Adjusted by age, sex and smoking

doi:10.1371/journal.pone.0130377.t004

within micro-RNA-binding sites are associated with breast cancer[42], bladder cancer[43] and colorectal cancer[44]. In addition to microRNA target sites, the vast majority of CRC SNPs revealed by GWAS overlapped with at least one enhancer in colon crypt, some of which were significantly associated with low-frequency lost variant enhancer loci (VELs)[45] and linked to altered expression level of their target genes. Hence, the above researches suggested the putative roles of our non-coding SNPs in CRC carcinogenesis.

To investigate the roles of *SF3A1* polymorphisms in contributing to the susceptibility of CRC, we performed an association analysis of rs5753073, rs2839998, rs10376 and rs2074733 in 801 cases and 817 cancer-free controls in a Chinese population. However, all of these SNPs have failed to be associated with CRC risk. When stratified analyses were performed by the smoking and alcohol use status, we still obtained no statistically significant results. We proposed that the limited sample size might be insufficient for this association study, therefore more cases and controls are required in the future.

Some limitations also existed in our current study. First, as a hospital-based case-control study, selection bias might not avoid. Therefore, larger prospective studies are participated to confirm our results. Second, CRC is a heterogeneous disease in which environmental factors play an important role. Therefore, more other risk factors should be considered to further elucidate the etiology of CRC.

In summary, we firstly investigated the associations between polymorphisms of *SF3A1* and CRC risk. Our study demonstrated that rs5753073, rs2839998, rs10376 and rs2074733 did not confer to the risk of CRC based on the current hospital-based study. Certainly, larger population-based studies with comprehensive design are needed to further clarify the role of polymorphisms of *SF3A1* in the etiology of CRC.

Acknowledgments

Author Contributions

Conceived and designed the experiments: XC HD SZ. Performed the experiments: XC HD. Analyzed the data: LZ WC. Contributed reagents/materials/analysis tools: BL YY YZ YG JT FL. Wrote the paper: XC HD SZ.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. CA Cancer J Clin 61: 69–90. doi: 10.3322/caac.20107 PMID: 21296855

2. Yao L, Tak YG, Berman BP, Farnham PJ (2014) Functional annotation of colon cancer risk SNPs. Nat Commun 5: 5114. doi: 10.1038/ncomms6114 PMID: 25268989

3. Brenner H, Kloor M, Pox CP (2014) Colorectal cancer. Lancet 383: 1490–1502. doi: 10.1016/S0140-6736(13)61649-9 PMID: 24225001

4. Zhu B, Zou L, Qi L, Zhong R, Miao X (2014) Allium vegetables and garlic supplements do not reduce risk of colorectal cancer, based on meta-analysis of prospective studies. Clin Gastroenterol Hepatol 12: 1991–2001 e1991-1994; quiz e1121. doi: 10.1016/j.cgh.2014.03.019 PMID: 24681077

5. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106: 9362–9367. doi: 10.1073/pnas.0903103106 PMID: 19474294

6. Wei Z, Wang K, Qu HQ, Zhang H, Bradfield J, Kim C, et al. (2009) From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. PLoS Genet 5: e1000678. doi: 10.1371/journal.pgen.1000678 PMID: 19816555

7. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 39: 989–994. PMID: 17618283

8. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 39: 984–988. PMID: 17618284

9. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, et al. (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 39: 1315–1317. PMID: 17934461

10. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, et al. (2008) Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 40: 26–28. PMID: 18084292

11. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, et al. (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 40: 631–637. doi: 10.1038/ng.133 PMID: 18372901

12. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, et al. (2008) 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 PMID: 18372905

13. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. (2008) 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 PMID: 19011631

14. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet 42: 973–977. doi: 10.1038/ng.670 PMID: 20972440

15. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. (2011) 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 PMID: 21655089

16. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, et al. (2012) Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet 44: 770–776. doi: 10.1038/ng.2293 PMID: 22634755

17. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. (2013) Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology 144: 799–807 e724. doi: 10.1053/j.gastro.2012.12.020 PMID: 23266556

18. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, et al. (2013) Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet 45: 191–196. doi: 10.1038/ng.2505 PMID: 23263487

19. Zhang B, Jia WH, Matsuo K, Shin A, Xiang YB, Matsuda K, et al. (2014) Genome-wide association study identifies a new SMAD7 risk variant associated with colorectal cancer risk in East Asians. Int J Cancer 135: 948–955. doi: 10.1002/ijc.28733 PMID: 24448986

20. Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, et al. (2014) Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat Genet 46: 533–542. doi: 10.1038/ng.2985 PMID: 24836286

21. Chen M, Manley JL (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 10: 741–754. doi: 10.1038/nrm2777 PMID: 19773805

22. Padgett RA (2012) New connections between splicing and human disease. Trends Genet 28: 147–154. doi: 10.1016/j.tig.2012.01.001 PMID: 22397991

23. Pal S, Gupta R, Davuluri RV (2012) Alternative transcription and alternative splicing in cancer. Pharmacol Ther 136: 283–294. doi: 10.1016/j.pharmthera.2012.08.005 PMID: 22909788

24. Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, Veitch J, et al. (2006) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics 7: 325. PMID: 17192196

25. Thorsen K, Sorensen KD, Brems-Eskildsen AS, Modin C, Gaustadnes M, Hein AM, et al. (2008) Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. Mol Cell Proteomics 7: 1214–1224. doi: 10.1074/mcp.M700590-MCP200 PMID: 18353764

26. Mojica W, Hawthorn L (2010) Normal colon epithelium: a dataset for the analysis of gene expression and alternative splicing events in colon disease. BMC Genomics 11: 5. doi: 10.1186/1471-2164-11-5 PMID: 20047688

27. Thorsen K, Mansilla F, Schepeler T, Oster B, Rasmussen MH, Dyrskjot L, et al. (2011) Alternative splicing of SLC39A14 in colorectal cancer is regulated by the Wnt pathway. Mol Cell Proteomics 10: M110 002998.

28. Bisognin A, Pizzini S, Perilli L, Esposito G, Mocellin S, Nitti D, et al. (2014) An integrative framework identifies alternative splicing events in colorectal cancer development. Mol Oncol 8: 129–141. doi: 10.1016/j.molonc.2013.10.004 PMID: 24189147

29. Adler AS, McCleland ML, Yee S, Yaylaoglu M, Hussain S, Cosino E, et al. (2014) An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. Genes Dev 28: 1068–1084. doi: 10.1101/gad.237206.113 PMID: 24788092

30. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478: 64–69. doi: 10.1038/nature10496 PMID: 21909114

31. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, et al. (2011) A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat Genet 43: 792–796. doi: 10.1038/ng.875 PMID: 21725308

32. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 45: 353–361, 361e351-352. doi: 10.1038/ng.2563 PMID: 23535729

33. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491: 119–124. doi: 10.1038/nature11582 PMID: 23128233

34. Zhu B, Tian J, Zhong R, Tian Y, Chen W, Qian J, et al. (2014) Genetic variants in the SWI/SNF complex and smoking collaborate to modify the risk of pancreatic cancer in a Chinese population. Mol Carcinog.

35. Zhong R, Liu L, Zou L, Zhu Y, Chen W, Zhu B, et al. (2013) Genetic variations in TERT-CLPTM1L locus are associated with risk of lung cancer in Chinese population. Mol Carcinog 52 Suppl 1: E118–126. doi: 10.1002/mc.22043 PMID: 23908149

36. Zhong R, Liu L, Zou L, Sheng W, Zhu B, Xiang H, et al. (2013) Genetic variations in the TGFbeta signaling pathway, smoking and risk of colorectal cancer in a Chinese population. Carcinogenesis 34: 936–942. doi: 10.1093/carcin/bgs395 PMID: 23275154

37. Pajares MJ, Ezponda T, Catena R, Calvo A, Pio R, Montuenga LM (2007) Alternative splicing: an emerging topic in molecular and clinical oncology. Lancet Oncol 8: 349–357. PMID: 17395108

38. David CJ, Manley JL (2010) Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev 24: 2343–2364. doi: 10.1101/gad.1973010 PMID: 21041405

39. Pennisi E (2011) The Biology of Genomes. Disease risk links to gene regulation. Science 332: 1031. doi: 10.1126/science.332.6033.1031 PMID: 21617055

40. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, et al. (2012) Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet 44: 1084–1089. doi: 10.1038/ng.2394 PMID: 22941192

41. Chen K, Song F, Calin GA, Wei Q, Hao X, Zhang W (2008) Polymorphisms in microRNA targets: a gold mine for molecular epidemiology. Carcinogenesis 29: 1306–1311. doi: 10.1093/carcin/bgn116 PMID: 18477647

42. Nicoloso MS, Sun H, Spizzo R, Kim H, Wickramasinghe P, Shimizu M, et al. (2010) Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. Cancer Res 70: 2789–2798. doi: 10.1158/0008-5472.CAN-09-3541 PMID: 20332227

43. Yang H, Dinney CP, Ye Y, Zhu Y, Grossman HB, Wu X (2008) Evaluation of genetic variants in micro-RNA-related genes and risk of bladder cancer. Cancer Res 68: 2530–2537. doi: 10.1158/0008-5472.CAN-07-5991 PMID: 18381463

44. Landi D, Gemignani F, Naccarati A, Pardini B, Vodicka P, Vodickova L, et al. (2008) Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. Carcinogenesis 29: 579–584. doi: 10.1093/carcin/bgm304 PMID: 18192692

45. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, et al. (2012) Epigenomic enhancer profiling defines a signature of colon cancer. Science 336: 736–739. doi: 10.1126/science.1217277 PMID: 22499810