RESEARCH ARTICLE

Identification of a Low-Frequency Missense Variant in E2F Transcription Factor 7 Associated with Colorectal Cancer Risk In A Chinese Population

Ai Ye Guo[1], Kan Zhai[2], Jin Ling Xu[1], Jin Long Hu[3]*, Lan Gao[1]*

Abstract

Background: Transcription factors regulate gene expression and play important role in tumor genesis. Especially, the E2F transcription factor family controls the cell cycle and regulate many tumor suppressors. Missense variants in E2F family genes, which change the amino acid sequence, may alter the capacity for DNA binding or the protein structure, leading to a functional alteration. **Material and Methods:** We here searched for missense variants in E2F transcription family genes (E2F1~E2F8) and identified two (rs2075995 for E2F2 and rs3829295 for E2F7) with minor allele frequencies >0.01 in Chinese Han Beijing population from the 1000 genome project. We genotyped these two variants in 1,055 colorectal cancer (CRC) patients and 1,936 healthy controls using Taqman genotyping assays and assessed associations between SNPs and risk of CRC using logistic regression adjusted for gender and age. **Results:** We found rs3829295 at E2F7 to be significantly associated with risk of CRC. Compared with TT genotype carriers, CT and CT+CC genotype carriers had lower risks of CRC with ORs of 0.61 (95% CI: 0.44-0.85, P=0.003) and 0.61 (95% CI: 0.44-0.84, P=0.003), respectively. When stratified by gender and age, significant associations were observed in males (OR= 0.56, 95% CI: 0.38-0.83, P=0.004) for rs3829295, but not females (OR= 0.73, 95% CI: 0.43-1.22, P=0.232). **Conclusion:** Through a systematic assessment of variants in the E2F transcription factor family, we identified a low-frequent missense variant in E2F7 significantly associated with CRC risk, indicating that E2F7 may play an important role in development of this tumor type.

Keywords: Colorectal cancer- E2F transcription factor 7- missense variant- susceptibility- case-control

Asian Pac J Cancer Prev, **18 (1)**, 271-275

Introduction

Colorectal cancer (CRC) ranks the third most commonly cancer in males and the second in females, with an estimation of 71,830 men and 65,000 women diagnosed with colorectal cancer in 2014 (Siegel et al., 2014; Torre et al., 2015). Although the incidence rates of CRC decreased by approximately 3% per year during the past decades in western countries (Siegel et al., 2014), with the progressive 'westernization' of lifestyles, the incidence rates of CRC in China is sharply increasing (Chen et al., 2016). Although some environmental factors were established to play an important role in the etiology of CRC, it is estimated that approximately 35% of CRC risk may be attributable to inherited factors (Lichtenstein et al., 2000). Previous genome-wide association study (GWAS) have identified nearly 20 genetic loci in Asian associated with CRC susceptibility (Jia et al., 2013; Zhang et al., 2014; Wang et al., 2016; Zeng et al., 2016), however, could only explain a small fraction of the heritability. Therefore, more variants, especially functional variants with low-frequency still need to be explored.

Transcription factors are proteins that controls the rate of transcription of genetic information form DNA to mRNA by binding to a specific DNA sequence. In turn, they regulate the expression of genes and paly a very important role in both cell activity and tumor genesis. A better understanding of the regulatory factors that contribute in the development of CRC could provide new insights into precision medicine of this disease. The E2F transcription factor family is a crucial group of transcription factors that involved in the cell cycle regulation and carcinogenesis (Chen et al., 2009; Polager and Ginsberg, 2009). There are eight genes in this family (E2F1~E2F8) and three of them are activators (E2F1, E2F2 and E2F3a). Missense variants, that change the amino acid of proteins, may alter the function of transcription factors and lead to a dysregulation of downstream genes.Thus, we conjecture

*[1]Department of Clinical Laboratory , [3]Department of Oncology, People's Hospital of Zhengzhou University and Henan Provincial People's Hospital, Zhengzhou, [2]Medical Research Center, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China. *For Correspondence: gaolan8@126.com, hujinlong0305@sina.com. Ai Ye Guo and Kan Zhai have equal contribution in this study*

that missense variants in E2F family genes may associate with CRC susceptibility.

In the present study, we searched for missense variants in the E2F transcription factors family (E2F1~E2F8) and performed a case-control study to test whether these variants is associated with CRC risk.

Materials and Methods

Study subjects

This study consists of 1,055 CRC patients and 1,936 healthy controls. CRC patients were enrolled from People's Hospital of Zhengzhou University and Henan Provincial People's Hospital, Zhengzhou, China between January 1st, 2012 to November 30th, 2015. Health controls were selected from a community cancer screening program for early detection conducted in the same region during the same period as cases were collected. All participants were unrelated Han Chinese descent, and the inclusion criteria included pathologically confirmed primary CRC, without any radiotherapy or chemotherapy treatment prior to blood samples collected. The informed consent was obtained from every participant at recruitment and peripheral blood samples and demographic characteristics such as gender, age and ethnicity were collected by interviewers. This study was conducted under the approval of the Institutional Review Board of Zhengzhou University.

SNP selection and genotyping

We searched for missense variant in E2F transcription factor family genes (E2F1~E2F8) using Ensembl (http://asia.ensembl.org/). rs2075995 at E2F2 and rs3829295 at E2F7 with MAF in Chinese Han Beijing (CHB) > 0.01 were selected for genotyping. Genotyping were performed using genomic DNA extracted from 2 ml peripheral blood sample collected from each participant at recruitment. SNPs were genotyped by Taqman SNP Genotyping Assay (Applied Biosystems). The PCR primers and probes for genotyping rs2075995 and rs3829295 were 5′-GAG GAT ATC TCT TGT TGG CCT TGT-3′/5′-GAC CTG GGA AGC AGC AAC AG-3′, 5′-FAM-CTG GAT GAG CTG GTC-MGBNFQ-3′/5′-HEX-CTG GAT GAG ATG GTC-MGBNFQ-3′ and 5′-AGC TCA GGG CTA ACA GGA TGT C-3′/5′-AAA GGC AAA GTA GGG AAT ATG AAG AC-3′, 5′-FAM-CCT TAC CTT GGG CAC GA-MGBNFQ-3′/5′-HEX-TTA CCT TGG GCA TGA C-MGBNFQ-3′. Several genotyping quality controls were implemented, including (i) the case and control samples were mixed in the plates, and persons who performed the genotyping assay were not aware of case or control status, (ii) positive and negative (no DNA) samples were included on every 384-well assay plate, and (iii) we further employed the direct sequencing of PCR products to replicate sets of 50 randomly selected, TaqMan-genotyped samples for the two SNPs and the accordance between the two methods was 100%.

Statistical analysis

Odds ratios (ORs) and 95% confidence intervals (CIs) adjusted for age and sex gender were used to assess the strength of associations between selected polymorphisms and CRC risk by using unconditional multivariate logistic regression analysis. All statistical analyses were performed using SAS software (version 9.1; SAS Institute, Cary, NC), with a significance level of 0.05. All tests were two-sided.

Results

Characteristic of Study Subjects

The distributions of selected demographic characteristics including sex gender and age of the CRC patients and healthy controls were summarized in Table 1. The present study included 1,055 CRC patients and 1,936 gender matched healthy controls. There were 66.9% and 62.4% males in cases and controls, respectively. The average age of cases and controls were 60.5 and 64.8, respectively.

Association between Missense SNPs in E2F family Genes and CRC Risk

Missense variant in E2F transcription factor family genes (E2F1~E2F8) were searched. SNPs with global minor allele frequency (MAF)>0.001 were listed (Table 2). Two SNPs (rs2075995 at E2F2 and rs3829295 at E2F7) with MAF in Chinese Han Beijing (CHB) > 0.01 were selected for genotyping.

The genotype frequencies of the two selected SNPs and their associations with CRC risk were showed in Table 3. We did not observe any significant associations between rs2075995 at E2F2 and CRC susceptibility (TG vs. TT: OR=0.88, 95% CI: 0.75-1.04, P=0.135; GG vs. TT: OR=0.85, 95% CI: 0.67-1.08, P=0.189). However, we identified a missense SNP (rs382929) in E2F7 significantly associated with CRC susceptibility. Compared with TT genotype carriers, CT genotype and CT+CC genotype carriers were associated with risk of CRC with OR being 0.61 (95% CI: 0.44-0.85, P=0.003) and 0.61 (95% CI: 0.44-0.84, P=0.003), respectively.

Stratified Analysis of Missense SNPs in E2F Family Genes and CRC Risk

We performed stratified analyses by age and gender to evaluate the effects of variant genotypes on the risk of CRC (Table 4). Among the males, rs3829295 significantly associated with CRC susceptibility with an odds ratio of 0.56 (95% CI: 0.38-0.83, P= 0.004). However, the association were not observed for females (OR= 0.73, 95% CI: 0.43-1.22, P=0.232). When stratified by age, rs3829295 were significantly associated with CRC risk in both younger (< 65 years) and older (≥ 65 years) participants with the OR being 0.59 (95% CI: 0.36-0.97, P=0.037) and 0.64 (95% CI: 0.43-0.96, P=0.031), respectively. For rs2075995, no significant associations

Table 1. Summary of Characteristics of Study Subjects

	Cases (n=1,055)	Controls (n=1,936)
Age (years), mean±S.D.	60.5 ± 12.8	64.8 ± 8.9
Gender, n (%)		
Male	706 (66.9)	1208 (62.4)
Female	349 (33.1)	728 (37.6)

Table 2. Genetic Variants in Exon Region of E2F Family Genes with Global Minor Allele Frequency >0.001

Chr	SNP	Position	Gene	Minor allele	Global MAF	CHB MAF	SIFT	PolyPhen
20	rs3213176	33676869	E2F1	T	0.012	0.000	1.0	0.01
20	rs149272498	33677168	E2F1	C	0.001	0.000	0.2	0.045
20	rs144481343	33677203	E2F1	T	0.001	0.000	0.9	0.003
20	rs3213173	33677440	E2F1	T	0.006	0.000	0.1	0.013
20	rs35385772	33678328	E2F1	T	0.013	0.000	1.0	0.002
20	rs145741678	33680374	E2F1	T	0.002	0.000	0.1	0.564
20	rs560305222	33680399	E2F1	T	0.001	0.000	0.2	0.21
20	rs574523664	33686231	E2F1	T	0.001	0.000	1.0	0.000
1	rs114788023	23516368	E2F2	T	0.004	0.000	0.4	0.002
1	rs116576730	23516451	E2F2	A	0.003	0.000	0.3	0.124
1	rs139052092	23519074	E2F2	A	0.001	0.000	0.0	0.838
1	**rs2075995**	**23520972**	**E2F2**	**A**	**0.355**	**0.413**	**0.2**	**0.62**
1	rs2229297	23521037	E2F2	T	0.006	0.000	0.1	0.008
1	rs41306580	23521042	E2F2	A	0.001	0.000	0.2	0.025
1	rs116694174	23521045	E2F2	A	0.001	0.000	0.0	0.097
1	rs3218125	23530550	E2F2	T	0.001	0.000	0.0	0.959
6	rs577488642	20402458	E2F3	G	0.002	0.000	0.0	0.153
6	rs189058327	20402579	E2F3	T	0.008	0.000	0.3	0.001
6	rs4134982	20490197	E2F3	A	0.003	0.000	0.4	0.013
6	rs115470365	20490212	E2F3	T	0.002	0.000	0.4	0.001
16	rs377030974	67195911	E2F4	A	0.001	0.005	0.4	0.001
16	rs201338753	67195926	E2F4	A	0.001	0.005	0.5	0.002
8	rs4150841	85177473	E2F5	C	0.029	0.000	0.9	0.000
8	rs187526876	85209401	E2F5	C	0.001	0.000	0.6	0.001
2	rs574653171	11447654	E2F6	G	0.001	0.000	0.0	0.000
2	rs142815155	11447668	E2F6	A	0.001	0.000	0.0	0.000
2	rs141189123	11450078	E2F6	T	0.005	0.000	1.0	0.002
12	rs146918336	77024094	E2F7	C	0.001	0.000	0.8	0.001
12	rs310831	77025561	E2F7	T	0.042	0.000	1.0	0.000
12	rs566695974	77025683	E2F7	A	0.001	0.000	0.0	0.424
12	rs560349395	77025706	E2F7	A	0.002	0.000	0.1	0.007
12	rs532446600	77025757	E2F7	T	0.002	0.000	0.0	0.221
12	rs34429867	77027946	E2F7	C	0.004	0.000	0.1	0.004
12	rs36110778	77028089	E2F7	G	0.006	0.000	0.8	0.000
12	**rs3829295**	**77029839**	**E2F7**	**C**	**0.008**	**0.0485**	**1.0**	**0.001**
12	rs73133236	77029847	E2F7	A	0.001	0.000	0.0	0.219
12	rs114349810	77030151	E2F7	T	0.003	0.000	0.1	0.019
12	rs201813841	77033069	E2F7	G	0.001	0.000	0.7	0.062
12	rs139349075	77033913	E2F7	C	0.002	0.000	0.0	0.013
12	rs148596563	77033971	E2F7	T	0.001	0.000	0.7	0.002
12	rs147478050	77044696	E2F7	C	0.001	0.000	0.0	0.661
12	rs61754233	77046043	E2F7	C	0.007	0.000	0.7	0.235
12	rs200583330	77055922	E2F7	T	0.001	0.005	0.0	0.925
12	rs310791	77056010	E2F7	A	0.144	0.005	1.0	0.000
12	rs149646617	77056111	E2F7	G	0.003	0.000	0.1	0.081
12	rs61744271	77064629	E2F7	A	0.002	0.000	0.0	0.17

Table 2. Continued

Chr	SNP	Position	Gene	Minor allele	Global MAF	CHB MAF	SIFT	PolyPhen
11	rs137938238	19224728	E2F8	C	0.008	0.000	0.3	0.003
11	rs145907915	19225288	E2F8	G	0.001	0.000	0.4	0.003
11	rs793274	19225738	E2F8	C	0.02	0.000	0.4	0.041
11	rs148337173	19225740	E2F8	C	0.001	0.000	0.0	1.000
11	rs80272893	19229879	E2F8	A	0.001	0.000	0.1	0.98
11	rs150203629	19230787	E2F8	A	0.001	0.000	0.0	0.462
11	rs77599073	19230828	E2F8	G	0.015	0.000	0.3	0.05
11	rs541607823	19234783	E2F8	C	0.001	0.000	0.2	0.005
11	rs201855427	19234818	E2F8	G	0.001	0.005	0.7	0.002
11	rs141999878	19234829	E2F8	A	0.001	0.000	0.0	0.99
11	rs144450449	19235058	E2F8	C	0.002	0.000	0.1	0.734
11	rs562521751	19237988	E2F8	T	0.001	0.000	0.1	0.463
11	rs200312339	19238051	E2F8	T	0.001	0.000	0.6	0.005

Note, SNP, single nucleotide polymorphism; MAF, minor allele frequency; CHB, Chinese Han Beijing; SIFT is an algorithm which predicts whether an amino acid substitution affects protein function.SNPs with CHB MAF >0.01 are in bold.

Table 3. Association between Two Exonic SNPs in E2F Family Genes and Risk of CRC in a Chinese Population

SNP	Chr	Position	Gene	Genotype	CasesNo. (%)	ControlsNo. (%)	OR (95% CI)[a]	P[a]
rs3829295	12	77423619	E2F7	TT	1,001 (94.9)	1,779 (91.9)	1.0 (Reference)	
				TC	53 (5.0)	153 (7.9)	0.6 (0.4-0.8)	0.003
				CC	1 (0.1)	4 (0.2)	0.4 (0.1-3.8)	0.445
				TC+CC	54 (5.1)	157 (8.1)	0.6 (0.4-0.8)	0.003
rs2075995	1	23847464	E2F2	TT	427 (40.5)	716 (37.0)	1.0 (Reference)	
				TG	496 (47.0)	952 (49.2)	0.9 (0.7-1.0)	0.135
				GG	132 (12.5)	268 (13.8)	0.8 (0.7-1.1)	0.189
				TG+GG	628 (59.5)	1,220 (63.0)	0.9 (0.7-1.0)	0.093

[a], Calculated by logistic regressionmodel adjusted for gender and age

Table 4. Association of rs3829295 and CRC Risk Stratified by Gender and Age

SNP	Male		Female	
	OR (95% CI)	*P*	OR (95% CI)	*P*
rs3829295	0.56 (0.38-0.83)	0.0038	0.73 (0.43-1.22)	0.2315
rs2075995	0.95 (0.83-1.10)	0.5099	0.82 (0.68-1.01)	0.0569
	> 65 years		≤ 65 years	
	OR (95% CI)	*P*	OR (95% CI)	*P*
rs3829295	0.59 (0.36-0.97)	0.0368	0.64 (0.43-0.96)	0.0311
rs2075995	0.92 (0.77-1.09)	0.3375	0.89 (0.77-1.03)	0.1217

were observed in stratified analyses.

Discussion

In this hospital-based case-control study, we explored the association between missense variants in E2F transcription factors family and CRC susceptibility in 1,055 CRC patients and 1,936 controls.We found the frequency of the TC/CC genotypes of the rs3829295 were significantly lower than that of TT genotypes, especially in males.The results from the present study suggest thatrs3829295 T>C polymorphism in E2F7 was significant associated with CRC risk, indicating an important role of E2F7 in CRC carcinogenesis.

The E2F7 gene is located at chromosome 12q21.2, containing 14 exons. As a member of E2F transcription factor family, E2F7 play an essential role in the regulation of cell cycle progression (Di Stefano et al., 2003). However, in contrast to the E2F activators (eg. E2F1 and E2F2), E2F7 can block the E2F-dependent activation of a subset of E2F target genes as well as mitigate cellular proliferation of mouse embryo fibroblasts (de Bruin et al., 2003). It was also reported that E2F7 and E2F8 works as a unique repressive arm of the E2F transcriptional network that controls the E2F1-p53 apoptotic axis (Li et al., 2008; Liu et al., 2013). E2F7 and E2F8 expression is induced in

response to DNA damage, which is partially dependent on P53 (Zalmas et al., 2008; Aksoy et al., 2012; Carvajal et al., 2012). Accumulating evidence showed that E2F7 may act as a tumor suppressor in multiple types of cancer by blocking cell proliferation (de Bruin et al., 2003; Endo-Munoz et al., 2009; Mitxelena et al., 2016; Thurlings et al., 2016).

E2F7 has two DNA binding domains and binds to the E2F DNA binding consensus site independently of DP co-factors (Logan et al., 2004). A mutational analysis indicates that the integrity of both DNA-binding domains is required for cell cycle delay and transcriptional modulation (Logan et al., 2004). However, the rs3829295 T>C polymorphism cause a 626th amino acid change (Met > Val), not an amino acid change in the DNA binding domain of E2F7. Although not lead to a direct alteration to the DNA binding capacity of E2F7, the variant may function through other ways, such as influencing on the formation of E2F7 homodimer (Logan et al., 2004). Significant association between this variant and CRC risk with a relatively high effect size (OR=0.61 in dominant model) was observed in this study. The high effect size and not been identified by previous GWAS explain by its relatively low frequency in Chinese population (CHB MAF = 0.0485) and potential important function.The association was significant especially in males but not females, indicating a gender disparities for CRC. This results all indicated that further functional analysis is worthy to be performed to elucidate the relationship between rs3829295 and CRC in the future.

In summary, through a case-control study in a Chinese Han population, we find a significant association between E2F7 missense variant rs3829295 and CRC susceptibility, especially in males. This results expand our insights of CRC carcinogenesis and provide more evidence for the precision medicine of this disease.

Author Contributions

J.L.H. and L.G. conceived and designed the experiments; A.Y.G. and K.Z. performed the experiments. J.L.X. analyzed the data. All authors wrote the paper and approved the final version.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This work was supported by grants from Foundation of Henan Provincial Science and Technology Program (142102310392) and Beijing Municipal Administration of Hospitals' Youth Programme (QML20150307).

References

Aksoy O, Chicas A, Zeng T, et al (2012). The atypical E2F family member E2F7 couples the p53 and RB pathways during cellular senescence. *Genes Dev*, **26**, 1546-57.

Carvajal LA, Hamard PJ, Tonnessen C, et al (2012). E2F7, a novel target, is up-regulated by p53 and mediates DNA damage-dependent transcriptional repression. *Genes Dev*, **26**, 1533-45.

Chen HZ, Tsai SY, Leone G (2009). Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat Rev Cancer*, **9**, 785-97.

Chen W, Zheng R, Baade PD, et al (2016). Cancer statistics in China, 2015. *CA Cancer J Clin*, **66**, 115-32.

de Bruin A, Maiti B, Jakoi L, et al (2003). Identification and characterization of E2F7, a novel mammalian E2F family member capable of blocking cellular proliferation. *J Biol Chem*, **278**, 42041-9.

Di Stefano L, Jensen MR, Helin K (2003). E2F7, a novel E2F featuring DP-independent repression of a subset of E2F-regulated genes. *EMBO J*, **22**, 6289-98.

Endo-Munoz L, Dahler A, Teakle N, et al (2009). E2F7 can regulate proliferation, differentiation, and apoptotic responses in human keratinocytes: implications for cutaneous squamous cell carcinoma formation. *Cancer Res*, **69**, 1800-8.

Jia WH, Zhang B, Matsuo K, et al (2013). Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat Genet*, **45**, 191-6.

Li J, Ran C, Li E, et al (2008). Synergistic function of E2F7 and E2F8 is essential for cell survival and embryonic development. *Dev Cell*, **14**, 62-75.

Lichtenstein P, Holm NV, Verkasalo PK, et al (2000). Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*, **343**, 78-85.

Liu B, Shats I, Angus SP, et al (2013). Interaction of E2F7 transcription factor with E2F1 and C-terminal-binding protein (CtBP) provides a mechanism for E2F7-dependent transcription repression. *J Biol Chem*, **288**, 24581-9.

Logan N, Delavaine L, Graham A, et al (2004). E2F-7: a distinctive E2F family member with an unusual organization of DNA-binding domains. *Oncogene*, **23**, 5138-50.

Mitxelena J, Apraiz A, Vallejo-Rodriguez J, Malumbres M, Zubiaga AM (2016). E2F7 regulates transcription and maturation of multiple microRNAs to restrain cell proliferation. Nucleic Acids Res. *Nucleic Acids Res*, **146**.

Polager S, Ginsberg D (2009). p53 and E2f: partners in life and death. *Nat Rev Cancer*, **9**, 738-48.

Siegel R, Desantis C, Jemal A (2014). Colorectal cancer statistics, 2014. *CA Cancer J Clin*, **64**, 104-17.

Thurlings I, Martinez-Lopez LM, Westendorp B, et al (2016). Synergistic functions of E2F7 and E2F8 are critical to suppress stress-induced skin cancer. *Oncogene*, **36**, 829-39

Torre LA, Bray F, Siegel RL, et al (2015). Global cancer statistics, 2012. *CA Cancer J Clin*, **65**, 87-108.

Wang M, Gu D, Du M, et al (2016). Common genetic variation in ETV6 is associated with colorectal cancer susceptibility. *Nat Commun*, **7**, 11478.

Zalmas LP, Zhao X, Graham AL, et al (2008). DNA-damage response control of E2F7 and E2F8. *EMBO Rep*, **9**, 252-9.

Zeng C, Matsuda K, Jia WH, et al (2016). Identification of susceptibility loci and genes for colorectal cancer risk. *Gastroenterology*, **150**, 1633-45.

Zhang B, Jia WH, Matsuda K, et al (2014). Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet*, **46**, 533-42.