Int J Clin Exp Med 2019;12(3):2710-2719
www.ijcem.com /ISSN:1940-5901/IJCEM0084842

Original Article

Association between intronic polymorphisms of *XRCC1*, *ERCC2* and *LIG1* genes and risk of esophageal squamous cell carcinoma in a Chinese Han population

Limeng Dai[1], Huanshen Tao[1], Gang Xiong[2], Xingying Guan[1], Yun Bai[1], Xiaofeng Xu[1]

[1]*Department of Medical Genetics, College of Basic Medical Science, Army Medical University (Third Military Medical University), Chongqing 400038, PR China;* [2]*Department of Cardiothoracic Surgery, Southwest Hospital, Army Medical University (Third Military Medical University), Chongqing 400038, PR China*

Received August 30, 2018; Accepted January 9, 2019; Epub March 15, 2019; Published March 30, 2019

Abstract: To further understand the susceptibility of DNA repair genes in esophageal squamous cell carcinomas (ESCC), the present genetic association study, involving intronic polymorphisms of X-ray repair cross-complementing protein 1 (XRCC1), excision repair cross-complementation group 2 (ERCC2), and DNA ligase 1 (LIG1) genes, was performed. Six tag single nucleotide polymorphisms (SNPs) in the introns of *XRCC1* (rs1799778, rs762507), *ERCC2* (rs238415, rs3916840), and *LIG1* (rs20580, rs7246512) genes were genotyped in a Chinese Han population, including 430 ESCC cases and 386 healthy controls. Genotyping of SNPs was conducted by PCR-restriction fragment length polymorphism analysis. Odds ratios (OR) with 95% confidence intervals (CI) were used to assess association in genetic models. For rs1799778 G>T polymorphism of *XRCC1* gene, significant association was found in overall analysis (for TT vs. GT+GG, OR=1.81, 95% CI=1.13-2.90) and male subgroup analysis (for TT vs. GT+GG, OR=1.95, 95% CI=1.11-3.45; for T vs. G, OR=1.29, 95% CI=1.01-1.66). For rs762507 C>T polymorphism of *XRCC1* gene, no significant association was found in overall analysis, but a strong association was identified in the female subgroup (for TT+CT vs. CC, OR=2.43, 95% CI=1.19-4.96; for T vs. C allele, OR=2.52, 95% CI=1.29-4.92). For polymorphisms of *ERCC2* and *LIG1* genes, no significant association was found in overall and subgroup analyses, based on gender. In analysis based on TNM stage, a moderate association was found in the recessive model for rs7246512 G>A polymorphism of *LIG1* gene (OR=2.51, 95% CI=1.07-5.90). The AA genotype increased the risk of I-IIa stages, compared to IIb-IV, in patients. No association was identified in other genetic models. In conclusion, present findings indicate that intronic polymorphisms of *XRCC1* (rs1799778, rs762507) genes are associated with ESCC risk, with association related to gender. Intronic polymorphisms of *ERCC2* (rs238415, rs3916840) and *LIG1* (rs20580, rs7246512) genes are not associated with ESCC risk, but the rs7246512 of *LIG1* gene may be associated with TNM stage in patients with ESCC.

Keywords: Esophageal squamous cell carcinoma, DNA repair gene, genetic polymorphism, susceptibility, XRCC1, ERCC2, LIG1

Introduction

Esophageal squamous cell carcinomas (ESCC) is the most common type in esophageal cancer patients, accounting for about 90% of all cases, with a higher incidence in China [1]. Occurrence of ESCC is an extremely complex process, involving various internal and external environmental exposure, such as physical, chemical, and biological carcinogens. These may lead to cellular stress, inflammation reactions, and even DNA damage of esophageal epithelial cells. Many studies have indicated that a decrease of DNA repair capacity may cause formation of genetic mutations, further leading to genomic instability and malignant transformation of normal cells [2, 3]. Because of the essential roles of the DNA repair system, increasing studies have focused on the genes of DNA repair pathways, investigating its roles on cancer susceptibility. However, these roles have not been fully elucidated in ESCC.

DNA repair pathways mainly contain single-strand damage repair (SSD) and double-strand breaks systems (DSBs), involving many proteins

Table 1. Designs of PCR-FRLP for genotyping

GENE	Tag SNP	Location (hg19)	Sense and antisense primers for PCR	Annealing	Endonuclease	Genotype and its expected bands (bp)
XRCC1	rs762507	chr19: 44057848	5'AGCACTCCCTCCCATTCCAC3'	63°C	*EcoR* I	C/C: 480
			5'CTAAGGTCCCGCAAGGTCAG3'			C/T: 480+434+46
						T/T: 434+46
	rs1799778	chr19: 44058891	5'ATTCCCCTTGCCTTTCCGC3'	63°C	*Hae* III	T/T: 224+75+40+31
			5'GAAGCCACAGTGCATGAGAACC3'			T/G: 224+195+29+75+40+31
						G/G: 195+29+75+40+31
ERCC2	rs238415	chr19: 45857235	5'GCTTCAATAGGCGGCTTCCA3'	63°C	*EcoR* II	G/G: 234+157
			5'AGAGGCTGCCCTGAGACTTCC3'			G/C: 234+157+106+51
						C/C: 234+106+51
	rs3916840	chr19: 45862047	5'GCCCACTCTCAGTCATCAC3'	63°C	*Hinf* I	G/G: 382
			5'CCCCAGCCACCTTTTCTAC3'			G/A: 382+327+55
						A/A: 327+55
LIG1	rs20580	chr19: 48654553	5'CTTTACAGGCACATTGGATTGG3'	58°C	*Hae* III	T/T: 413+58+47
			5'AAACTCACTGGAGGTCTTTAGGG3'			T/G: 413+340+73+58+47
						G/G: 340+73+58+47
	rs7246512	chr19: 48651752	5'AAAGCAGGGAGGAATGACAG3'	58°C	*Hae* III	A/A: 413+20
			5'GGATGAACCATAGAGAGCCG3'			A/G: 413+255+158+20
						G/G: 255+158+20

associated with base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), non-homologic end joining (NHEJ) mechanisms, and homologic repair (HR) [2]. Genes of DNA repair pathways, such as X-ray repair cross-complementing protein 1 (XRCC1) of BER system, excision repair cross-complementation group 2 (ERCC2) of the NER system, and DNA ligase 1 (LIG1) of NEHJ system, participate in the protection of cells from DNA damage, maintaining genomic integrity [3]. In previous studies, epidemiological data has suggested that genetic factors, such as polymorphisms, copy number variation, and *de novo* mutation may be associated with susceptibility of ESCC [4, 5]. To date, association studies between DNA repair genes (*XRCC1*, *ERCC2,* and *LIG1*) and ESCC susceptibility have mainly focused on SNPs in coding region, while the roles of intronic polymorphisms have not been fully clarified [6, 7]. There is not enough evidence to fully understand the association between intronic polymorphisms of these DNA repair genes and ESCC risk. Moreover, increasing studies have showed that non-coding regions of genes also play an essential role in regulation of gene expression [8, 9]. Thus, genetic association studies, involving intronic polymorphisms from *XRCC1*, *ERCC2* and *LIG1* genes, have been performed, aiming to further clarify the roles of these DNA repair genes in ESCC.

In the present study, six tag SNPs from introns of *XRCC1* (rs1799778, rs762507), *ERCC2* (rs238415, rs3916840), and *LIG1* (The rs20580, rs7246512) genes were evaluated for association with susceptibility and clinic stage of ESCC in a Chinese Han population.

Materials and methods

Study populations

This case-control study was approved by the Ethics Committee of Army Military Medical University (Third Military Medical University). All participants provided informed consent. There were 430 cases and 386 healthy controls in this study. Patients with ESCC and healthy controls were Chinese Han individuals from Southwest China. All patients had a defined pathological diagnosis from surgical excision tissues at Southwest Hospital, between 2007 to 2016. Postoperative TNM stages of patients were performed according to the standard of American Joint Committee on Cancer (AJCC). Healthy controls were defined from physical examinations of the population at Southwest Hospital, between 2007 to 2016. They were matched to the patients concerning age, gender, and residential area.

Tag SNPs

Six tag SNPs, with restriction sites in *XRCC1* (rs1799778, rs762507), *ERCC2* (rs238415, rs-

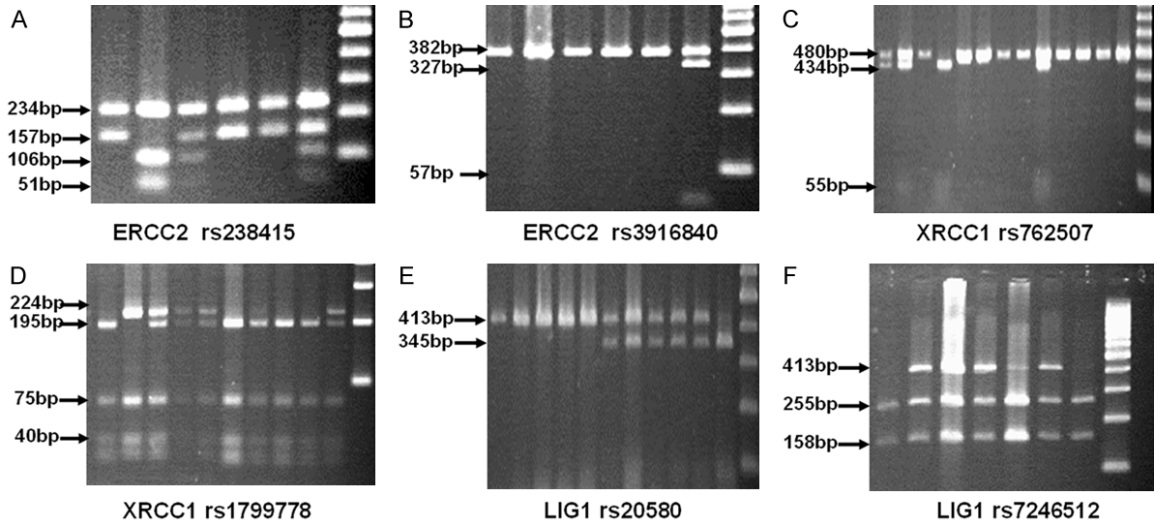

Figure 1. Typical PCR-RFLP determination of the samples. A. The genotypes of rs238415 of *ERCC2* gene. B. The genotypes of rs3916840 of *ERCC2* gene. C. The genotypes of rs762507 of *XRCC1* gene. D. The genotypes of rs1799778 of *XRCC1* gene. E. The genotypes of rs20580 of *LIG1* gene. F. The genotypes rs7246512 of *LIG1* gene.

Table 2. General characteristic of subjects

	ESCC cases	Controls	*P* value
Total	430	386	
Age (years)	55.23±6.68	55.31±5.18	0.85
Gender			
Male (N, %)	323	270	0.06
Female (N, %)	104	116	
TNM stage			
I (N, %)	53 (12.33%)		
IIa (N, %)	214 (49.77%)		
IIb (N, %)	91 (21.16%)		
III (N, %)	46 (10.70%)		
IV (N, %)	26 (6.05%)		

3916840), and *LIG1* (The rs20580, rs72465-12) genes, were included. The minor allele frequency (MAF) of these 6 tag SNPs was more than 0.05. Moreover, rs1799778 and rs762507 were found in intron 3 and 4 of *XRCC1* genes, respectively. The rs238415 was located in the region of intron 4, while rs3916840 was located in intron 12 of *ERCC2* gene. Finally, rs20580 and rs7246512 were found in the region of intron 5 and 7 of *LIG1* gene, respectively.

Genotyping

Genotyping of SNPs was conducted using PCR-restriction fragment length polymorphism (PCR-RFLP) analysis. Genomic DNA was isolated from peripheral blood. Concentration levels needed to be greater than 50 ng/μL. The PCR amplification system (30 μL) contained 1 μL genomic DNA (50-100 ng), 1.5 μL primer pairs (concentration 10 μM), 15 μL GoTaq@ PCR Master Mix (Promega), and 12.5 μL ddH_2O. The reaction program consisted of 95°C for 5 minutes, followed by 36 cycles at 95°C for 30 seconds, 57-63°C for 30 seconds, 72°C for 30 seconds, and 72°C for 5 minutes. PCR products (20 μL) were digested in a 37°C water bath with 6 U restriction endonucleases (*EcoR* I, *EcoR* II, *Hinf* I, or *Hae* III) overnight. They were detected by agarose gel eletrophoresis. Sequences of primers, annealing temperatures of PCR reactions, and expected digestion fragments are listed in **Table 1**. Typical genotype fragments of PCR-RFLP detection that indicated corresponding genotypes are shown in **Figure 1**. Moreover, 10% of the samples were randomly selected for sequencing verification.

Statistical analysis

Hardy-Weinberg equilibrium (HWE) in the healthy control population was evaluated using Chi-squared test. Five genetic models for association analyses were performed: dominant model (AA+AB versus BB), recessive model (AA versus AB+BB), homozygote comparison (AA versus BB), heterozygote comparison (AB versus BB), and allele comparison (A allele versus

Table 3. Genotype and allelic frequencies of SNPs in ESCC cases and controls

Gene	SNPs	Genotype	Case N=430 (%)		Control N=386 (%)		HWE	OR [95% CI]	*P* value
XRCC1	rs762507 C>T	CC	342	79.53%	319	82.64%	0.33	Reference	
		CT	82	19.07%	62	16.06%		1.23 [0.86, 1.77]	0.26
		TT	6	1.40%	5	1.30%		1.12 [0.34, 3.70]	0.85
		TT+CT†	88	20.47%	67	17.36%		1.23 [0.86, 1.74]	0.26
		CC+CT‡	424	98.60%	381	98.70%		1.08 [0.33, 3.56]	0.91
		C	766	89.07%	700	90.67%		Reference	
		T	94	10.93%	72	9.33%		1.19 [0.86, 1.65]	0.29
	rs1799778 G>T	GG	212	49.30%	194	50.26%	0.51	Reference	
		GT	163	37.91%	163	42.23%		0.92 [0.68, 1.22]	0.55
		TT	55	12.79%	29	7.51%		1.74 [1.06, 2.83]	0.03*
		TT+GT†	218	50.70%	192	49.74%		1.04 [0.79, 1.37]	0.78
		GG+GT‡	375	87.21%	357	92.49%		1.81 [1.13, 2.90]	0.01*
		G	587	68.26%	551	71.37%			
		T	273	31.74%	221	28.63%		1.16 [0.94, 1.43]	0.17
ERCC2	rs238415 G>C	GG	121	28.14%	96	24.87%	0.08	Reference	
		GC	206	47.91%	210	54.40%		0.78 [0.56, 1.08]	0.14
		CC	103	23.95%	80	20.73%		1.02 [0.69, 1.52]	0.92
		CC+GC†	309	71.86%	290	75.13%		0.85 [0.62, 1.16]	0.29
		GG+GC‡	327	76.05%	306	79.27%		1.20 [0.87, 1.68]	0.27
		G	448	52.09%	402	52.07%		Reference	
		C	412	47.91%	370	47.93%		1.00 [0.82, 1.21]	0.99
	rs3916840 G>A	GG	380	88.37%	343	88.86%	0.25	Reference	
		GA	49	11.40%	43	11.14%		1.03 [0.67, 1.59]	0.91
		AA	1	0.23%	0	0.00%		2.71 [0.11, 66.70]	0.54
		AA+GA†	50	11.63%	43	11.14%		1.05 [0.68, 1.62]	0.83
		GG+GA‡	429	99.77%	386	100.00%		2.70 [0.11, 66.47]	0.54
		G	809	94.07%	729	94.43%		Reference	
		A	51	5.93%	43	5.57%		1.07 [0.70, 1.62]	0.76
LIG1	rs20580 T>G	TT	211	49.07%	185	47.93%	0.93	Reference	
		TG	181	42.09%	165	42.75%		0.96 [0.72, 1.28]	0.79
		GG	38	8.84%	36	9.33%		0.93 [0.56, 1.52]	0.76
		GG+TG†	219	50.93%	201	52.07%		0.96 [0.73, 1.26]	0.74
		TT+TG‡	392	91.16%	350	90.67%		0.94 [0.58, 1.52]	0.81
		T	603	70.12%	535	69.30%			
		G	257	29.88%	237	30.70%		0.96 [0.78, 1.19]	0.72
	rs7246512 G>A	GG	218	50.70%	196	50.78%	0.61	Reference	
		GA	178	41.40%	161	41.71%		0.99 [0.75, 1.33]	0.97
		AA	34	7.91%	29	7.51%		1.05 [0.62, 1.79]	0.85
		AA+GA†	212	49.30%	190	49.22%		1.00 [0.76, 1.32]	0.98
		GG+GA‡	396	92.09%	357	92.49%		1.06 [0.63, 1.77]	0.83
		G	614	71.40%	553	71.63%		Reference	
		A	246	28.60%	219	28.37%		1.01 [0.82, 1.25]	0.92

†: dominant model (AA+AB versus BB); ‡: recessive model (AA versus AB+BB). The A allele represents the minor allele. *: P<0.05.

B allele). The A allele represents the minor allele. Odds ratios (ORs) and 95% confidence intervals (CIs) were used to assess the effects of any differences between patients and con-

trols or between TNM I-IIa and TNM IIb-IV in patients. Statistical significance is set at P< 0.05. All data were analyzed using SPSS software (version 13.0).

Results

Study characteristics

This study included 430 ESCC patients, with a mean age of 55.25±6.68 years, and 386 healthy controls, with a mean age of 55.31±5.18 years. No statistically significant differences were found between cases and controls in frequency distribution of sex (P=0.06) and age (P=0.85). Most cases were at the stage of TNM-IIa (49.77%) and TNM-IIb (21.16%). Cases of TNM-I, TNM-III and TNM-IV stages accounted for 12.33%, 10.7% and 6.05%, respectively. Data of these subjects are listed in **Table 2**.

For all tag SNPs in this study, as shown in **Table 3**, the observed genotype frequencies of controls (overall, male, or female) were in HWE (P>0.05). They were in the frequency range of CHB (Chinese Han in Beijing) population in the dbSNP database. Present data suggests that the population of the current study was a genetic equilibrium population, with no deviation of artificial selection.

Association between XRCC1 intronic polymorphisms and ESCC risk

Significant association was found between *XRCC1* (rs1799778, rs762507) polymorphisms and ESCC risk, according to overall or subgroup analyses by gender in part genetic models.

For rs1799778 G>T polymorphism, significant association was found in overall analysis under the recessive model (TT vs. TG+GG; OR=1.81; 95% CI=1.13-2.90; P=0.01) and model of TT vs. GG (OR=1.74; 95% CI=1.06-2.83), as shown in **Table 3**. In subgroup analysis by gender, significant association was mainly found from males under the model TT vs. TG+GG (OR= 1.95; 95% CI=1.11-3.45; P=0.01), TT vs. GG (OR=2.09; 95% CI=1.18-3.69; P=0.01) and T vs. G allele (OR=1.29; 95% CI=1.01-1.66; P= 0.04). However, there was no significant association in females. Results of subgroup analysis by gender are shown in **Table 4**. According to subgroup analysis based on TNM stage in patients, a moderate association was found in the model of GT vs. GG (OR=0.60; 95%=0.40-0.92; P=0.02). No significant association was found in other genetics models. Results of subgroup analysis by TNM stage are shown in **Table 5**.

For rs762507 C>T polymorphism, there was no significant association in overall analysis, but a strong association was obtained in females, but not in males. In the subgroup of females, significant association was found in the models of TT+CT vs. CC (OR=2.43, 95% CI=1.19, 4.96; P=0.01), CT vs. CC (OR=2.15; 95% CI= 1.04-4.44; P=0.04), and T vs. C allele (OR= 2.52; 95% CI=1.29-4.92; P=0.007). According to subgroup analysis of TNM stage, no significant association was found. Data are shown in **Tables 3**-**5**.

Association between ERCC2 intronic polymorphisms and ESCC risk

For SNPs rs238415 and rs3916840, no significant association was identified between patients and controls for all genetic models, according to overall analysis. In subgroup analyses by gender, there was no significant association in male or female populations. In subgroup analysis by TNM stage in patients, no significant association was found between TNM I-IIa and TNM IIb-IV individuals with ESCC, suggesting that there was no significant association between these two polymorphisms and ESCC risk or progression of tumor stage in patients. Overall analysis data are shown in **Table 3**. Subgroup analysis data by gender are shown in **Table 4**. Subgroup analysis data by TNM stage are shown in **Table 5**.

Association between LIG1 intronic polymorphisms and ESCC risk

Concerning the relationship between SNPs rs-20580 and rs7246512 of *LIG1* gene, no significant association was identified between patients and controls, according to overall analysis. Furthermore, there was no significant association in subgroup analyses carried by gender. Results showed no significant association between these two polymorphisms and ESCC risk. Overall data are shown in **Table 3**. Subgroup analysis data by gender are shown in **Table 4**.

According to subgroup analysis by TNM stage, a moderate association was identified in reces-

Table 4. Stratification analyses of genotype and allelic frequencies in ESCC patients and controls according to gender

Gene/SNPs		Male					Female				
		Case (N=326)	Control (N=270)	HWE	OR [95% CI]	*P* value	Case (N=104)	Control (N=116)	HWE	OR [95% CI]	*P* value
XRCC1											
rs762507 C>T	CC	264	217	0.23	Reference		78	102	0.49	Reference	
	CT	59	48		1.01 [0.66, 1.54]	0.96	23	14		2.15 [1.04, 4.44]	0.04*
	TT	3	5		0.49 [0.12, 2.09]	0.34	3	0		9.14 [0.47, 179.54]	0.15
	TT+CT†	62	53		0.96 [0.64, 1.45]	0.85	26	14		2.43 [1.19, 4.96]	0.01*
	CC+CT‡	323	265		0.49 [0.12, 2.08]	0.33	101	116		8.03 [0.41, 157.41]	0.17
	C	587	482		Reference		179	218		Reference	
	T	65	58		0.92 [0.63, 1.34]	0.66	29	14		2.52 [1.29, 4.92]	0.007*
rs1799778 G>T	GG	158	144	0.88	Reference		54	50	0.31	Reference	
	GT	126	107		1.07 [0.76, 1.51]	0.69	37	56		0.61 [0.35, 1.08]	0.09
	TT	42	19		2.09 [1.18, 3.69]	0.01*	13	10		1.20 [0.48, 2.99]	0.69
	TT+GT†	168	126		1.22 [0.88, 1.68]	0.24	50	66		0.70 [0.41, 1.19]	0.19
	GG+GT‡	284	251		1.95 [1.11, 3.45]	0.02*	91	106		1.51 [0.63, 3.62]	0.53
	G	442	395		Reference		145	156		Reference	
	T	210	145		1.29 [1.01, 1.66]	0.04*	63	76		0.89 [0.60, 1.33]	0.58
ERCC2											
rs238415 G>C	GG	100	74	0.23	Reference		21	22	0.13	Reference	
	GC	152	144		0.78 [0.54, 1.14]	0.2	54	66		0.86 [0.43, 1.72]	0.67
	CC	74	52		1.05 [0.66, 1.68]	0.83	29	28		1.09 [0.49, 2.40]	0.84
	CC+GC†	226	196		0.85 [0.60, 1.22]	0.38	83	94		0.93 [0.47, 1.80]	0.82
	GG+GC‡	252	218		1.23 [0.83, 1.83]	0.31	75	88		1.22 [0.66, 2.22]	0.53
	G	352	292		Reference		96	110		Reference	
	C	300	248		1.00 [0.80, 1.26]	0.98	112	122		1.05 [0.72, 1.53]	0.79
rs3916840 G>A	GG	289	240	0.33	Reference		91	103	0.52	Reference	
	GA	36	30		1.00 [0.60, 1.67]	0.99	13	13		1.13 [0.50, 2.57]	0.77
	AA	1	0		2.49 [0.10, 61.46]	0.58	0	0		-	
	AA+GA†	37	30		1.02 [0.61, 1.71]	0.93	13	13		1.13 [0.50, 2.57]	0.77
	GG+GA‡	325	270		2.49 [0.10, 61.45]	0.58	104	116		-	
	G	614	510		Reference		195	219		Reference	
	A	38	30		1.05 [0.64, 1.72]	0.84	13	13		1.12 [0.51, 2.48]	0.77
LIG1											
rs20580 T>G	TT	159	123	0.77	Reference		52	62	0.84	Reference	
	TG	143	120		0.92 [0.66, 1.29]	0.64	38	45		1.01 [0.57, 1.78]	0.98
	GG	24	27		0.69 [0.38, 1.25]	0.22	14	9		1.85 [0.74, 4.63]	0.19
	GG+TG†	167	147		0.88 [0.64, 1.21]	0.43	52	54		1.15 [0.68, 1.95]	0.61
	TT+TG‡	302	243		0.72 [0.40, 1.27]	0.25	90	107		1.85 [0.76, 4.47]	0.17
	T	461	366		Reference		142	169		Reference	
	G	191	174		0.87 [0.68, 1.12]	0.28	66	63		1.25 [0.83, 1.88]	0.29
rs7246512 G>A	GG	161	138	0.66	Reference		57	58	0.76	Reference	
	GA	143	112		1.09 [0.78, 1.53]	0.6	35	49		0.73 [0.41, 1.28]	0.27
	AA	22	20		0.94 [0.49, 1.80]	0.86	12	9		1.36 [0.53, 3.47]	0.52
	AA+GA†	165	132		1.07 [0.78, 1.48]	0.68	47	58		0.82 [0.49, 1.40]	0.48
	GG+GA‡	304	250		0.90 [0.48, 1.70]	0.75	92	107		1.55 [0.63, 3.85]	0.34
	G	465	388		Reference		149	165		Reference	
	A	187	152		1.03 [0.80, 1.32]	0.84	59	67		0.98 [0.64, 1.48]	0.91

†: dominant model (AA+AB versus BB); ‡: recessive model (AA versus AB+BB). The A allele represents the minor allele. *: P<0.05.

sive models (AA vs. GA+GG) for the rs7246512 G>A polymorphism. Genotype AA increased the risk of I-IIa stages, compared to IIb-IV stages (OR=2.51, 95% CI=1.07-5.90, P=0.04), in the patients with ESCC. No significant association was found in other genetic models. For the

Table 5. Stratification analyses of genotype and allelic frequencies in ESCC patients according to TNM stage

Gene/SNPs	Genotype	TNM I-IIa N=267		TNM IIb-IV N=163		OR [95% CI]	*P* value
XRCC1							
rs762507 C>T	CC	208	77.90%	134	82.21%	Reference	
	CT	55	20.60%	27	16.56%	1.31 [0.79, 2.18]	0.3
	TT	4	1.50%	2	1.23%	1.29 [0.23, 7.13]	0.77
	TT+CT†	59	22.10%	29	17.79%	1.31 [0.80, 2.15]	0.28
	CC+CT‡	263	98.50%	161	98.77%	1.22 [0.22, 6.76]	0.82
	C	471	88.20%	295	90.49%	Reference	
	T	63	11.80%	31	9.51%	1.27 [0.81, 2.00]	0.3
rs1799778 G>T	GG	133	49.81%	79	48.47%	Reference	
	GT	99	37.08%	64	39.26%	0.60 [0.40, 0.92]	0.02*
	TT	35	13.11%	20	12.27%	1.04 [0.56, 1.92]	0.9
	TT+GT†	134	50.19%	84	51.53%	0.95 [0.64, 1.40]	0.79
	GG+GT‡	232	86.89%	143	87.73%	1.08 [0.60, 1.94]	0.8
	G	365	68.35%	222	68.10%	Reference	
	T	169	31.65%	104	31.90%	0.99 [0.74, 1.33]	0.94
ERCC2							
rs238415 G>C	GG	76	28.46%	45	27.61%	Reference	
	GC	125	46.82%	81	49.69%	0.91 [0.58, 1.45]	0.7
	CC	66	24.72%	37	22.70%	1.06 [0.61, 1.82]	0.84
	CC+GC†	191	71.54%	118	72.39%	0.96 [0.62, 1.48]	0.85
	GG+GC‡	201	75.28%	126	77.30%	1.12 [0.71, 1.77]	0.63
	G	277	51.87%	171	52.45%	Reference	
	C	257	48.13%	155	47.55%	1.02 [0.78, 1.35]	0.87
rs3916840 G>A	GG	235	88.01%	145	88.96%	Reference	
	GA	31	11.61%	18	11.04%	1.06 [0.57, 1.97]	0.85
	AA	1	0.37%	0	0.00%	1.86 [0.08, 46.00]	0.7
	AA+GA†	32	11.99%	18	11.04%	1.10 [0.59, 2.03]	0.77
	GG+GA‡	266	99.63%	163	100.00%	1.84 [0.07, 45.45]	0.71
	G	501	93.82%	308	94.48%	Reference	
	A	33	6.18%	18	5.52%	1.13 [0.62, 2.04]	0.69
LIG1							
rs20580 T>G	TT	135	50.56%	76	46.63%	Reference	
	TG	108	40.45%	73	44.79%	0.83 [0.55, 1.25]	0.38
	GG	24	8.99%	14	8.59%	0.97 [0.47, 1.98]	0.92
	GG+TG†	132	49.44%	87	53.37%	0.85 [0.58, 1.26]	0.43
	TT+TG‡	243	91.01%	149	91.41%	1.05 [0.53, 2.10]	0.89
	T	378	70.79%	225	69.02%	Reference	
	G	156	29.21%	101	30.98%	0.92 [0.68, 1.24]	0.58
rs7246512 G>A	GG	136	50.94%	82	50.31%	Reference	
	GA	104	38.95%	74	45.40%	0.85 [0.57, 1.27]	0.42
	AA	27	10.11%	7	4.29%	2.33 [0.97, 5.58]	0.06
	AA+GA†	131	49.06%	81	49.69%	0.98 [0.66, 1.44]	0.9
	GG+GA‡	240	89.89%	156	95.71%	2.51 [1.07, 5.90]	0.04*
	G	376	70.41%	238	73.01%	Reference	
	A	158	29.59%	88	26.99%	1.14 [0.84, 1.55]	0.41

†: dominant model (AA+AB versus BB); ‡: recessive model (AA versus AB+BB). The A allele represents the minor allele. *: P<0.05.

rs20580 polymorphism, there were no significant association in all genetic models, suggesting that it may have no significant effects on progression of tumor stage. Data are shown in **Table 5**.

Discussion

To the best of our knowledge, there are no other reports regarding the association between intronic polymorphisms of *XRCC1*, *ERCC2* and *LIG1* genes and risk of ESCC. The present study gives new genetic evidence, aiming to understand the roles of polymorphisms located in the intron regions of DNA repair pathway genes concerning ESCC risk.

XRCC1 gene encoded X-ray repair cross-complementing protein 1 is an important protein involving single-strand break repair, base excision repair, and nucleotide excision repair [10]. The current study first explicated the association between the intronic polymorphism of *XRCC1* gene and ESCC risk. Interestingly, the effects of SNPs had gender differences on ESCC risk. A strong association was found in rs1799778 G>T polymorphism. Genotype of TT increased the ESCC risk in the overall and male population, while rs762507 C>T polymorphism only increased the ESCC risk in a female population. Previous studies have mainly focused on the coding regions of Arg194Trp, Arg280His, and Arg399Gln polymorphisms and non-coding region polymorphisms of -77 T>C. For the polymorphism of -77 T>C, there was no association with ESCC risk [11], but it contributed to diminished promoter activity in lung cancer [9]. A recent meta-analysis showed that a strong significant association was identified in Arg194Trp polymorphism [12]. A boundary association was found in Arg399Gln polymorphism [13], while no significant differences were found in Arg280His polymorphism [12]. Notably, the rs1799778 polymorphism was about 3416 bp from tagged rs25487 (Arg399Gln; r^2=1) coding SNP, suggesting that these tag polymorphisms were associated with risk of ESCC. Moreover, rs1799778 polymorphism had a potential association with TNM stage in patients with ESCC, indicating that it may have stronger effects, compared to rs25487 polymorphism, because significant association was found in the present population. Moreover, the location of rs762507 site was between rs1799778 and rs25487. Association was found in the female population. The rs1799778 and rs762507 showed statistically significant association in different gender populations, which was in linkage disequilibrium with rs25487 that was associated with many cancer types, including esophageal cancer risk [13]. However, it lacked related functional studies to clarify the mechanisms of these coding/non-coding regions polymorphisms in ESCC. Thus, more evidence is necessary to confirm it in *XRCC1* gene.

ERCC2 gene encoded excision repair cross-complementation group 2 protein is also known as xeroderma pigmentosum complementary group D (XPD). ERCC2 plays important roles in NER pathways, a part of human transcriptional initiation factor TFIIH and ATP-dependent helicase activity [14]. In the current study, no significant association was found in the rs238415 and rs3916840 polymorphisms, suggesting that it was not the potentially functional SNPs in *ERCC2* genes contributing to ESCC risk. In previous studies, early studies focused on Lys751Gln (rs13181) and Asp312Asn (rs1799793) polymorphisms, but uncertain association was found. With increasing repeated studies performed, as well as sample size accumulation, some meta-analyses have suggested a positive association with ESCC risk for rs13181 and rs1799793 polymorphisms [6, 15]. A polymorphism of synonymous mutation Arg156Arg (rs238406) was associated with ESCC risk, interacting with smoking [16]. Moreover, some polymorphisms were associated with *ERCC2* expression, such as the Arg156Arg (rs238406) and Lys751Gln (rs13181) polymorphism [3]. Haplotypes G-C-G-G-G and G-G-A-G-C of rs3916874 -rs238415 -rs1618536 -rs1799793 -rsl3181 polymorphisms of *ERCC2* gene increased OR values of ESCC risk [17]. Therefore, these cumulative results suggest a moderate association between *XRCC1* gene and ESCC risk, with mechanisms involving functional SNPs, impact of haplotypes, and various environmental factors.

LIG1 gene encoded DNA ligase 1 enzyme functions in DNA replication and the base excision in repair process when DNA damage occurs [18]. The present study was the first association study between *LIG1* polymorphisms and ESCC risk. No significant association was found between *LIG1* rs20580 and rs7246512 and ESCC risk, but rs7246512 G allele was highly prevalent in patients with early TNM stage. In other cancers, the non-coding SNP of rs156641 increased the risk of lung cancer, according to

a meta-analysis [19]. Results suggest a potential association between the *LIG1* gene and ESCC risk, but more evidence is needed results.

In the current study, association was found between intronic polymorphism of DNA repair genes and ESCC risk. Some mechanisms can explain these results. DNA repair is a reaction of cells to DNA damage, which relies on a variety of enzymes involving SSD or DSBs repair system. The SSD system consists of BER, NER and MMR. BER system works when alkylative or oxidative base products emerge by radiation, reactive oxygen species (ROS), and alkylating agents. NER system works when DNA damage is caused by UV light or oxygen radicals, whereas replicative errors were removed by MMR mechanism [3]. DSBs system depends on the phase of replication process, that NHEJ repair was enabled in G1 stage, whereas HR worked in G2/M stage [3]. *XRCC1*, *ERCC2* and *LIG1* genes mainly involve BER, NER and NEHJ systems, respectively. These DNA repair proteins work together when DNA is damaged by various factors, such as oxygen radicals, anti-tumor agents, UV light, and radiation. DNA repair capacity relies on the amount and function of these DNA repair proteins. The coding region of functional polymorphisms of DNA repair genes could lead to changes in doses of protein products via affecting the binding with some transcription factors, affecting coded amino acids and producing attenuated protein [20]. The promoter region of polymorphisms of DNA repair genes may affect promoter activity, thereby changing gene expression. Intronic regions of DNA repair genes may involve the splicing sites or splicing regulatory sites, thereby affecting the amount or sequence of splicing products. Moreover, some polymorphisms may have different effects with gender, family history, and other environmental factors, such as smoking. In this study, gene-sex interactions were found in rs1799778 and rs762507 polymorphisms of *XRCC1* gene. Moreover, rs762507 increased ESCC risk in females, while rs1799778 was mainly associated with the male population. Results showed the complexity between SNPs and gender, involving interaction with sex-related pathway genes, such as estrogen receptors [21, 22].

There were several limitations to the current study. First, significant associations were from a population with a small sample size. More studies with larger samples are necessary to confirm results in different populations with ESCC or other cancers. Second, many factors participate in the progression of ESCC besides gender. The present study did not perform subgroup analyses based on other factors, such as smoking and drinking. Third, a quick look into the dbSNP database showed about 2,772 intron variants in *XRCC1* gene, while the MAF of 108 polymorphisms was more than 5%. There were about 2,590 intronic SNPs, in which the MAF of 53 SNPs was more than 5% in *ERCC2* gene. Corresponding numbers were 5,691 and 832 in *LIG1* gene, respectively. The present experiment focused on few restriction sites due to experimental conditions and economic reasons. Thus, there was potential for SNP selection bias. Moreover, ESCC is a complex disease and interactions such as gene-environment and haplotype-gene should be considered. The DNA repair system involves many proteins, but only three (*XRCC1*, *ERCC2* and *LIG1*) were included. No further analysis of their interactions was performed. Despite these limitations, evidence from the current association analysis was adequate. Some positive associations were found between intronic SNPs and ESCC risk. However, functional assays were not conducted. Thus, present conclusions only represent relevance.

In summary, the present study suggests that intronic polymorphisms of *XRCC1* (rs1799778, rs762507) gene are associated with ESCC risk. The association is related to gender. Intronic polymorphisms of *ERCC2* (rs238415, rs3916840) and *LIG1* (rs20580, rs7246512) genes are not associated with ESCC risk, but the rs7246512 of *LIG1* gene may be associated with TNM stage in ESCC patients.

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (No. 81700222) and Natural Science Foundation Project of CQ CSTC (No. cstc2017jcyjAX0442). We would like to acknowledge the blood samples of cases given by the Department of Cardiothoracic Surgery, the First Affiliated Hospital of Army Medical University (Third Military Medical University). We wish to acknowledge patients supporting the present study.

Disclosure of conflict of interest

None.

Address correspondence to: Xiaofeng Xu and Yun Bai, Department of Medical Genetics, College of Basic Medical Science, Army Medical University (Third Military Medical University), Chongqing 400038, PR China. Tel: +86 023 68771367; E-mail: xfxu98@126.com (XFX); baiyungene@gmail.com (YB)

References

[1] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87-108.

[2] Turgeon MO, Perry NJS and Poulogiannis G. DNA damage, repair, and cancer metabolism. Front Oncol 2018; 8: 15.

[3] Dylawerska A, Barczak W, Wegner A, Golusinski W and Suchorska WM. Association of DNA repair genes polymorphisms and mutations with increased risk of head and neck cancer: a review. Med Oncol 2017; 34: 197.

[4] Wang K, Liu B, Li J, Xiong G, Guan X, Yang K and Bai Y. Association between SNPs in P53 binding regions and risk of esophageal squamous cell carcinoma. Int J Biol Markers 2014; 29: e160-168.

[5] Wu Y, Liu X, Hu L, Tao H, Guan X, Zhang K, Bai Y and Yang K. Copy number loss of variation_91720 in PIK3CA predicts risk of esophageal squamous cell carcinoma. Int J Clin Exp Pathol 2015; 8: 14479-14485.

[6] Guo XF, Wang J, Lei XF, Zeng YP and Dong WG. XPD Lys751Gln polymorphisms and the risk of esophageal cancer: an updated meta-analysis. Intern Med 2015; 54: 251-259.

[7] Findlay JM, Middleton MR and Tomlinson I. A systematic review and meta-analysis of somatic and germline DNA sequence biomarkers of esophageal cancer survival, therapy response and stage. Ann Oncol 2015; 26: 624-644.

[8] Deng N, Zhou H, Fan H and Yuan Y. Single nucleotide polymorphisms and cancer susceptibility. Oncotarget 2017; 8: 110635-110649.

[9] Hao B, Miao X, Li Y, Zhang X, Sun T, Liang G, Zhao Y, Zhou Y, Wang H, Chen X, Zhang L, Tan W, Wei Q, Lin D and He F. A novel T-77C polymorphism in DNA repair gene XRCC1 contributes to diminished promoter activity and increased risk of non-small cell lung cancer. Oncogene 2006; 25: 3613-3620.

[10] London RE. The structural basis of XRCC1-mediated DNA repair. DNA Repair (Amst) 2015; 30: 90-103.

[11] Yun YX, Dai LP, Wang P, Wang KJ, Zhang JY and Xie W. Association of polymorphisms in X-ray repair cross complementing 1 gene and risk of esophageal squamous cell carcinoma in a Chinese population. Biomed Res Int 2015; 2015: 509215.

[12] Li S, Deng Y, You JP, Chen ZP, Peng QL, Huang XM, Lu QH, Huang XL, Zhao JM and Qin X. XRCC1 Arg399Gln, Arg194Trp, and Arg280His polymorphisms in esophageal cancer risk: a meta-analysis. Dig Dis Sci 2013; 58: 1880-1890.

[13] Zhang ZY, Xuan Y, Jin XY, Tian X and Wu R. Meta-analysis demonstrates association of XRCC1 genetic polymorphism Arg399Gln with esophageal cancer risk in the Chinese population. Genet Mol Res 2013; 12: 2567-2577.

[14] Fuss JO and Tainer JA. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. DNA Repair (Amst) 2011; 10: 697-713.

[15] Xiao F, Pu J, Wen Q, Huang Q, Zhang Q, Huang B, Huang S, Lan A, Zhang Y, Li J, Zhao D, Shen J, Wu H, He Y, Li H and Yang X. Association between the ERCC2 Asp312Asn polymorphism and risk of cancer. Oncotarget 2017; 8: 48488-48506.

[16] Zhu ML, He J, Wang M, Sun MH, Jin L, Wang X, Yang YJ, Wang JC, Zheng L, Xiang JQ and Wei QY. Potentially functional polymorphisms in the ERCC2 gene and risk of esophageal squamous cell carcinoma in Chinese populations. Sci Rep 2014; 4: 6281.

[17] Zhang Y, Wang L, Wang P, Song C, Wang K, Zhang J and Dai L. Association of single nucleotide polymorphisms in ERCC2 gene and their haplotypes with esophageal squamous cell carcinoma. Tumour Biol 2014; 35: 4225-4231.

[18] Paul-Konietzko K, Thomale J, Arakawa H and Iliakis G. DNA ligases I and III support nucleotide excision repair in DT40 cells with similar efficiency. Photochem Photobiol 2015; 91: 1173-1180.

[19] Li D, Li R, Zhang J, Li K and Wu Y. Association between the LIG1 polymorphisms and lung cancer risk: a meta-analysis of case-control studies. Cell Biochem Biophys 2015; 73: 381-387.

[20] Dai LM, Fu LY, Liang Y, Guan XY, Bai Y and Xiong G. A functional polymorphism rs2257440 in the gene DcR3 regulates its expression via MTF-1 in esophageal squamous cell carcinoma. Int J Clin Exp Pathol 2017; 10: 11006-11013.

[21] Kim HI, Lim H and Moon A. Sex differences in cancer: epidemiology, genetics and therapy. Biomol Ther (Seoul) 2018; 26: 335-342.

[22] Bugni JM, Han J, Tsai MS, Hunter DJ, Samson LD. Genetic association and functional studies of major polymorphic variants of MGMT. DNA Repair (Amst) 2007; 6: 1116-1126.