Carcinogenesis vol.34 no.7 pp.1536–1542, 2013
doi:10.1093/carcin/bgt094
Advance Access publication March 15, 2013

Genetic variants in DNA repair pathway genes and risk of esophageal squamous cell carcinoma and gastric adenocarcinoma in a Chinese population

Wen-Qing Li[1,*], Nan Hu[1], Paula L.Hyland[1], Ying Gao[1], Zhao-Ming Wang[1,2], Kai Yu[1], Hua Su[1], Chao-Yu Wang[1], Le-Min Wang[1], Stephen J.Chanock[2], Laurie Burdett[2], Ti Ding[3], You-Lin Qiao[4], Jin-Hu Fan[4], Yuan Wang[3], Yi Xu[3], Jian-Xin Shi[1], Fangyi Gu[1], William Wheeler[5], Xiao-Qin Xiong[5], Carol Giffen[5], Margaret A.Tucker[1], Sanford M.Dawsey[1], Neal D.Freedman[1], Christian C.Abnet[1], Alisa M.Goldstein[1] and Philip R.Taylor[1]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institute of Health, Bethesda, MD 20852, USA, [2]Cancer Genomics Research Laboratory, NCI-Frederick, SAIC-Frederick Inc., Frederick, MD, USA, [3]Shanxi Cancer Hospital, Taiyuan, Shanxi, People's Republic of China, [4]Department of Epidemiology, Cancer Institute (Hospital), Chinese Academy of Medical Sciences, Beijing, People's Republic of China and [5]Information Management Services, Inc., Silver Spring, MD, USA

*To whom correspondence should be addressed. Tel: +1 301 402 9437; Fax: +1 301 402 4489;
Email: liw9@mail.nih.gov

The DNA repair pathways help to maintain genomic integrity and therefore genetic variation in the pathways could affect the propensity to develop cancer. Selected germline single nucleotide polymorphisms (SNPs) in the pathways have been associated with esophageal cancer and gastric cancer (GC) but few studies have comprehensively examined the pathway genes. We aimed to investigate associations between DNA repair pathway genes and risk of esophageal squamous cell carcinoma (ESCC) and GC, using data from a genome-wide association study in a Han Chinese population where ESCC and GC are the predominant cancers. In sum, 1942 ESCC cases, 1758 GC cases and 2111 controls from the Shanxi Upper Gastrointestinal Cancer Genetics Project (discovery set) and the Linxian Nutrition Intervention Trials (replication set) were genotyped for 1675 SNPs in 170 DNA repair-related genes. Logistic regression models were applied to evaluate SNP-level associations. Gene- and pathway-level associations were determined using the resampling-based adaptive rank-truncated product approach. The DNA repair pathways overall were significantly associated with risk of ESCC ($P = 6.37 \times 10^{-4}$), but not with GC ($P = 0.20$). The most significant gene in ESCC was *CHEK2* ($P = 2.00 \times 10^{-6}$) and in GC was *CLK2* ($P = 3.02 \times 10^{-4}$). We observed several other genes significantly associated with either ESCC (*SMUG1*, *TDG*, *TP53*, *GTF2H3*, *FEN1*, *POLQ*, *HEL308*, *RAD54B*, *MPG*, *FANCE* and *BRCA1*) or GC risk (*MRE11A*, *RAD54L* and *POLE*) ($P < 0.05$). We provide evidence for an association between specific genes in the DNA repair pathways and the risk of ESCC and GC. Further studies are warranted to validate these associations and to investigate underlying mechanisms.

Introduction

Gastric cancer (GC) and esophageal cancer represent the second and sixth most frequent causes of cancer-related deaths worldwide, respectively (1,2). People living around the Taihang Mountains of north central China have a high risk of esophageal squamous cell carcinoma (ESCC) and GC (3,4). Numerous studies have evaluated environmental risk factors for ESCC and GC, but the molecular mechanisms underlying carcinogenesis remain ill defined (5–7). The increased risk of non-cardia GC associated with *Helicobacter pylori* infection has been well described, but only a small proportion of infected subjects develop GC (8). In Western populations, smoking is an established risk factor for ESCC and GC and heavy alcohol intake is a risk factor for ESCC (9). In contrast, smoking and alcohol intake are not major contributing factors for ESCC and GC in high-risk populations (6,7). These findings suggest the likely significance of genetic or other lifestyle contributions.

Genomic instability due to DNA damage by carcinogens has been implicated in the development of cancer (10–12). DNA damage response and repair counteract the threats to genomic integrity, and variations in DNA repair capacity resulting from genetic polymorphisms could therefore correlate with cancer predisposition (10,11,13). Polymorphisms in candidate DNA repair genes from small-scale studies have been associated with risk of ESCC or GC, but the findings have been inconsistent and coverage of genes limited (14–20). One review of candidate gene association studies provided only sparse evidence for an association between DNA repair-related genes and cancer (21). Prior genome-wide association studies (GWAS) have identified a number of genetic loci linked to risk of ESCC or GC, but information on DNA repair genes is limited (5,22–29). Instead of one-by-one single nucleotide polymorphism (SNP) analysis, the analysis of pathways offers the opportunity to combine evidence from multiple potentially related genetic variants and may provide additional insight about the genetic architecture of complex diseases (30,31). We, therefore, sought to comprehensively examine associations between DNA repair pathway genes and the risk of ESCC and GC in ethnic Chinese subjects in a combined analysis of 1942 ESCC cases, 1758 GC cases and 2111 controls drawn from the Shanxi Upper Gastrointestinal (UGI) Cancer Genetics Project and the Linxian Nutrition Intervention Trials (NITs).

Materials and methods

Study populations

Our study contained two sets of populations: a discovery set, with participants drawn from the Shanxi UGI Cancer Genetics Project in the western part of the Taihang Mountain area and a replication set, with participants drawn from the NITs in the southern part of the Taihang Mountain area. The Shanxi study was conducted between 1997 and 2007, which had a case–control portion and a case only portion. We identified newly diagnosed, histologically confirmed ESCC and GC cases (7). Controls were matched on age (±5 years), sex and neighborhood for the case–control portion (7). Blood samples were collected at enrollment for all cases and controls. The NITs were initiated in Linxian in 1985 and tested the effect of multiple vitamin and mineral combinations taken for up to 6 years on the incidence and mortality of esophageal cancer and gastric cardia cancer (4). Following a blood survey conducted in 1999 and 2000, all newly diagnosed, histologically confirmed ESCC and GC cases documented during the follow-up through 31 December 2007, along with controls from an age- and gender-stratified randomly sampled subcohort, were included in the current analysis. All examined esophageal cancers were identified as ESCC, and all GCs were adenocarcinomas. Cardia cancers were defined as those located in the proximal 3 cm of the stomach, whereas non-cardia cancers were those in the remainder of the stomach.

Both the Shanxi and NIT studies obtained informed consent from subjects and were approved by their Institutional Review Boards. The NCI Special Studies Institutional Review Board approved both the Shanxi and NIT studies and the overall GWAS.

Gene and SNP selection

A complete list of the 173 genes included can be found on the website of MD Anderson Cancer Center (http://sciencepark.mdanderson.org/labs/wood/

Abbreviations: ARTP, adaptive rank-truncated product; CI, confidence interval; ESCC, esophageal squamous cell carcinoma; GC, gastric cancer; GWAS, genome-wide association study; LD, linkage disequilibrium; NIT, Nutrition Intervention Trial; OR, odds ratio; SNP, single nucleotide polymorphism; UGI, upper gastrointestinal.

DNA_Repair_Genes.html, updated on 3 October 2011, retrieved 30 January 2012) (32). These genes include (i) those responsible for base excision repair, direct reversal of damage, repair of DNA–protein cross-links, mismatch excision repair, nucleotide excision repair, homologous recombination, Fanconi anemia repair, non-homologous end-joining, modulation of nucleotide pools, ubiquitination and modification; (ii) chromatin structure; (iii) genes encoding DNA polymerases (catalytic subunits) and editing and processing nucleases; (iv) genes defective in diseases associated with sensitivity to DNA-damaging agents; and (v) other identified genes with known or suspected DNA repair function. SNPs mapping to three genes were not found in the GWAS database after quality control filters (*GIYD1*, *GIYD2* and *GTF2H2*), leaving 170 genes in our analysis (Supplementary Table S1, available at *Carcinogenesis* Online). Using these 170 genes, we designated the major subpathways based on investigators' knowledge and an extensive literature search (Supplementary Table S2, available at *Carcinogenesis* Online). A total of 1675 SNPs located within these 170 genes and their flanking areas (20 kb upstream and 10 kb downstream), with a minor allele frequency of >1% (in cases and controls combined), were included in the final analysis.

Genotyping and quality control

Our published GWAS detailed the genome-wide scan that used the Illumina 660W array (5). After that initial report, additional subjects were scanned on the same platform at the same facility. Both the initial and additional scan data underwent similar processing and quality control metrics. We excluded SNPs with missing rates >5%; we excluded subjects whose SNP completion rates were <94%, who had abnormal mean heterozygosity values (>30 or <25%), were gender discordant or were an unexpected duplicate pair. Following all subject exclusions, data on a total of 5811 subjects were included in this analysis, including 1942 ESCCs, 1758 GCs (1126 cardia and 632 non-cardia cancers) and 2111 controls (Table I).

Statistical analysis

Gene-level analysis is our primary analysis. To conduct gene-level analysis, we first carried out SNP-level analysis in the Shanxi, NIT, and the pooled Shanxi and NIT study populations. We performed gene and SNP discovery in the Shanxi population; genes and SNPs with $P < 0.05$ were then replicated in the NIT population.

For each SNP, an unconditional logistic regression model was fitted to calculate the odds ratios (ORs) and 95% confidence intervals (CIs) for one minor allele. We used an additive model adjusted for age (<40, 40–49, 50–59, 60–69 or ≥70 years), gender and study (for the combined data). Because there was no evidence for significant problems with population substructure within studies, we did not consider population stratification within each study but we adjusted for study in our analyses of the combined participants (5). When the expected number of subjects with heterozygous genotype or homozygous for the minor allele was less than five, we combined these two genotypes and used a dominant model. After excluding SNPs with pairwise linkage disequilibrium (LD) $r^2 \geq 0.80$ in controls, a Bonferroni-corrected significance threshold was calculated from 1046 SNPs ($P = 4.78 \times 10^{-5}$, 0.05/1046 SNPs). The most significant SNPs in *CHEK2*, *SMUG1* and *CLK2* were plotted along with their proxies as a function of genomic location (based on Hapmap CHBJPT), annotated by the recombination rate across the locus and nearby genes (33). In secondary analyses, we evaluated associations after additionally adjusting for cigarette smoking (ever or never), or for smoking, alcohol intake (ever or seldom/never) and family history of UGI cancer (yes or no). The results from these secondary models were very similar to that in the primary model; therefore, we present the analyses using the primary model.

We calculated gene-level associations using rank-truncated test statistics and a permutation-based sampling procedure (1 000 000 resamplings). We applied this adaptive rank-truncated product (ARTP) approach, which combines association signals over a set of SNPs within a gene while accounting for SNP LD structures and multiple comparisons (31). Statistical significance for gene-based analyses was defined as $P < 0.05$. In addition, a more stringent Bonferroni-corrected significance threshold for gene-based analysis was performed to account for testing 170 genes ($P = 2.94 \times 10^{-4}$, 0.05/170 genes).

Pathway-level analyses were conducted for the overall DNA repair pathways and several subpathways in the combined population. Each pathway-level analysis was a global test for the association between the outcome and genes within the pathway in which we applied the ARTP method with 1 000 000 resamplings to the gene-level *P*-values and obtained a single summary pathway-level *P*-value.

For SNPs with association *P*-value < 0.01, we tested the association with ESCC and GC by subgroups of smoking or other characteristics (age, gender, alcohol intake or family history of UGI cancer) in the combined population. The *P* for interactions (P_{int}) between SNPs and these variables were examined using likelihood ratio tests.

Statistical analyses were performed using the SAS (V9.2; SAS Institute, Cary, NC) and R language. We evaluated the LD between SNPs across specific gene regions with Haploview version 4.1.

Results

Characteristics of this study

A total of 1942 ESCCs, 1758 GCs and 2111 controls were drawn from the Shanxi (1421 ESCCs, 1395 GCs and 1660 controls) and NIT studies (521 ESCCs, 363 GCs and 451 controls). Characteristics of participants are shown in Table I. Shanxi participants were older and less likely to be female or have a family history of UGI cancer than NIT participants. In both studies, there were more cases with cardia than non-cardia cancer.

Association of ESCC and GC with individual genes

Among Shanxi participants, we identified 20 genes that were significantly associated with ESCC risk ($P < 0.05$) (Table II and Supplementary Table S1, available at *Carcinogenesis* Online). *CHEK2* showed the strongest association with ESCC in the combined data ($P = 2.00 \times 10^{-6}$), with a significance level that exceeded the Bonferroni-corrected threshold. Another 11 genes, including *SMUG1*, *TDG*, *TP53*, *GTF2H3*, *FEN1*, *POLQ*, *HEL308*, *RAD54B*, *MPG*, *FANCE* and *BRCA1*, were significant in the combined data ($P < 0.05$) and had ORs of the most significant SNPs in the same direction in both the Shanxi and NIT data. *SMUG1* was the only gene that replicated in NIT. In addition, the most significant SNP (rs4135054) in *TDG* was replicated in NIT (Table II and Supplementary Table S1, available at *Carcinogenesis* Online). Additionally adjusting for smoking (Supplementary Table S1, available at *Carcinogenesis* Online) and other variables for SNP-level analysis (data not shown) did not appreciably change the gene-level *P*-values from the primary model that adjusted for only age, gender and study.

For the 11 genes associated with risk of GC in Shanxi, 4 remained significant after combination with NIT, including *CLK2*, *MRE11A*, *RAD54L* and *POLE* (Table III and Supplementary Table S1, available at *Carcinogenesis* Online). *CLK2* showed significance ($P = 3.02 \times$

Table I. Characteristics of the study participants in Shanxi UGI Cancer Genetics Project and Linxian NITs

	Shanxi UGI Cancer Genetics Project (Shanxi)					Linxian NITs (NIT)				
	Controls	ESCC cases	GC cases			Controls	ESCC cases	GC cases		
			Total	Cardia	Non-cardia			Total	Cardia	Non-cardia
n	1660	1421	1395	864	531	451	521	363	262	101
Age, mean (SD), year	57.8 (9.2)	58.1 (8.1)	57.8 (9.4)	59.3 (8.4)	55.4 (10.4)	49.5 (7.4)	50.0 (6.9)	50.4 (6.9)	50.3 (7.1)	50.6 (6.6)
Female (yes, %)	434 (26.1)	923 (35.1)	264 (18.9)	133 (15.4)	131 (24.7)	243 (53.9)	280 (53.7)	152 (41.9)	109 (41.6)	43 (42.6)
Smoking (ever, %)	1079 (65.0)	861 (60.6)	990 (71.0)	628 (72.7)	362 (68.2)	144 (31.9)	147 (28.2)	146 (40.2)	110 (42.0)	36 (35.6)
Alcohol intake (ever, %)	290 (17.5)	309 (21.8)	329 (23.6)	206 (23.8)	123 (23.2)	13 (2.9)	6 (1.15)	10 (2.8)	8 (3.05)	2 (1.98)
Family history of UGI cancer (yes, %)	338 (20.4)	343 (24.1)	298 (21.4)	204 (23.6)	94 (17.7)	140 (31.0)	198 (38.0)	129 (35.5)	93 (35.5)	36 (35.6)

Table II. The association between DNA repair pathway genes and risk of ESCC in the Shanxi, NIT and combined populations[a]

Gene	Location	Gene-level analysis			No. of SNP	Most significant SNP analysis[b]						
		Shanxi	NIT	Combined		dbSNP id (major, minor allele)	Shanxi		NIT		Combined	
		P	*P*	*P*			OR	*P*	OR	*P*	OR	*P*
CHEK2	22q12	4.00×10^{-6}	0.34	2.00×10^{-6}	9	rs738722 (C, T)	1.34	4.61×10^{-7}	1.17	0.13	1.30	1.65×10^{-7}
SMUG1	12q13	0.040	0.037	1.72×10^{-3}	7	rs2029166 (T, C)	1.15	8.01×10^{-3}	1.27	7.55×10^{-3}	1.18	3.04×10^{-4}
TDG	12q23	0.036	0.17	5.69×10^{-3}	11	rs4135054 (C, T)	1.19	6.54×10^{-3}	1.30	0.036	1.21	9.08×10^{-4}
TP53	17p13	7.32×10^{-3}	0.91	8.88×10^{-3}	6	rs12951053 (A, C)	1.18	1.60×10^{-3}	1.07	0.48	1.15	1.98×10^{-3}
GTF2H3	12q24	0.033	0.57	0.014	6	rs4930737 (T, C)	1.33	0.019	1.41	0.18	1.35	6.21×10^{-3}
FEN1	11q12	0.010	0.68	0.014	5	rs174537 (G, T)	0.85	5.99×10^{-3}	0.94	0.58	0.88	8.61×10^{-3}
POLQ	3q13	0.021	0.85	0.015	11	rs7632907 (C, T)	0.76	3.52×10^{-3}	0.87	0.42	0.78	2.45×10^{-3}
HEL308	4q21	0.018	0.49	0.023	5	rs13115704 (T, C)	0.86	6.45×10^{-3}	0.95	0.58	0.88	8.07×10^{-3}
RAD54B	8q22	0.033	0.27	0.024	10	rs2930961 (T, C)	0.89	0.019	0.84	0.059	0.88	4.08×10^{-3}
MPG	16p13	0.031	0.83	0.025	6	rs216606 (G, A)	0.78	6.88×10^{-3}	0.85	0.36	0.80	5.51×10^{-3}
FANCE	6p21	0.035	0.72	0.029	11	rs4713859 (T, C)	1.18	4.79×10^{-3}	1.10	0.37	1.16	3.80×10^{-3}
BRCA1	17q21	0.036	0.12	0.032	12	rs8176257 (C, A)	1.14	0.017	1.07	0.49	1.12	0.015
BRIP1	17q23	0.025	0.23	0.023	16	rs2191249 (C, A)	0.82	2.46×10^{-3}	1.13	0.27	0.89	0.046
RAD54L	1p34	0.031	0.93	0.11	6	rs17102086 (T, C)	1.15	7.38×10^{-3}	0.97	0.71	1.11	0.028
POLN	4p16	0.029	0.89	0.13	21	rs3117813 (A, G)	0.86	5.03×10^{-3}	1.07	0.49	0.90	0.036
XPC	3p25	0.025	0.95	0.14	9	rs1106087 (G, T)	0.85	4.39×10^{-4}	1.07	0.50	0.90	0.029
MUTYH	1p34	0.031	0.74	0.17	3	rs3219487 (G, A)	0.84	0.012	1.10	0.40	0.90	0.07
POLB	8p11	0.050	0.24	0.18	3	rs10958713 (T, C)	0.89	0.024	1.06	0.55	0.93	0.094
BRCA2	13q13	8.51×10^{-3}	0.63	0.22	29	rs9590896 (G, A)	1.27	4.53×10^{-4}	0.81	0.081	1.15	0.017
RNF4	4p16	0.018	0.81	0.23	8	rs704352 (C, T)	0.85	2.61×10^{-3}	1.12	0.25	0.91	0.041

[a]Genes were ranked by the gene-level *P*-value except *BRIP1* for which the directionality of the OR of the most significant SNP was different. Gene-level *P*-values were calculated using ARTP approach. The *P*-values and ORs for the SNPs were calculated from unconditional logistic regression models using genotype-trend tests adjusted for age (10 year categories), sex and study (for the combined population).
[b]The analysis was based on the most significant SNP in the discovery set (the Shanxi data).

10^{-4}) close to the threshold after adjustment for multiple comparisons. Results differed by anatomic subsite of GC. *CLK2* was significantly associated with both cardia and non-cardia cancers. In contrast, *RAD54L* and *MRE11A* were associated only with cardia cancer and *POLE* only with non-cardia cancer. *MRE11A* was significant in both the Shanxi and NIT data, but only among cardia cancers. Analysis of cardia cancer alone identified a significant association with *BRCA2*, whereas non-cardia cancer alone was associated with both *ERCC2* and *XPA*.

Association of ESCC and GC with individual SNPs

We found three SNPs in *CHEK2* (rs738722, rs1547014 and rs6005863) significantly associated with ESCC risk, exceeding Bonferroni-corrected threshold, and one SNP each in *CHEK2* (rs2073327), *SMUG1* (rs2029166), *TDG* (rs4135054) and *POLM* (rs4236374) with $P < 0.001$ (Supplementary Table S3, available at *Carcinogenesis* Online). rs1052176 (*CLK2*) and rs13447720 (*MRE11A*) were associated with risk of GC (both $P < 0.001$). Other analyses indicated associations of rs9590896 (*BRCA2*) and rs7135624 (*GTF2H3*) with cardia cancer risk (Supplementary Table S4, available at *Carcinogenesis* Online). rs738722, rs2029166 and rs1052176 were further plotted along with their proxies (based on Hapmap CHBJPT) as a function of genomic location. The regional LD plots showed associations of rs738722 only in *CHEK2* and rs2029166 only in *SMUG1*, whereas the most significant SNP in *CLK2* (rs1052176) was also in LD with portions of *HCN3*, *PKLR* and *FDPS*, and several other genes (Supplementary Figure S1, available at *Carcinogenesis* Online).

Association of ESCC and GC with the overall pathways and subpathways

The overall DNA repair pathways were associated with ESCC risk ($P = 6.37 \times 10^{-4}$), but not GC ($P = 0.20$). We evaluated subpathways and observed significant associations only with homologous recombination ($P = 1.27 \times 10^{-4}$), which harbors *CHEK2*, and base excision repair ($P = 0.012$) (Supplementary Table S2, available at *Carcinogenesis* Online). For GC, we did not observe significant subpathways (data not shown).

Stratified analysis and interaction analysis

We observed an interaction between rs9590896 (*BRCA2*) and sex on ESCC ($P_{int} = 4.96 \times 10^{-4}$) such that the SNP was significantly associated with risk only in men (OR = 1.33, 95% CI = 1.15–1.54), but not in women (OR = 0.87, 95% CI = 0.72–1.06). We did not observe significant interactions between SNPs and smoking or other covariates with $P_{int} < 0.001$. However, at the threshold of 0.01, we observed an interaction between smoking and rs2029166 (*SMUG1*) on both ESCC and GC risk, with a significant association with ESCC and GC risk only observed among non-smokers; the *P*-value for association with ESCC risk exceeded the Bonferroni-corrected threshold ($P = 1.43 \times 10^{-5}$). Finally, an interaction was found between smoking and rs11614717 (*POLE*) on GC risk, with the significant association of the SNP only among smokers (Supplementary Table S5, available at *Carcinogenesis* Online).

Discussion

There have been a number of previous reports relating genetic variants in DNA repair genes and risk of UGI cancer, but results have been inconsistent (14–21). Prior studies have had very limited coverage of DNA repair-related genes, typically evaluating only a handful of genes (e.g. *XRCC1*, *OGG1* and *APE1*) (14–21). As a complex network, the significance of the DNA repair pathways in UGI cancer is not yet well understood. In our study, we observed that the DNA repair pathways overall, as well as numerous DNA repair-related genes, were associated with risk of ESCC. We also found evidence for associations between several DNA repair-related genes and GC risk, although the overall pathway itself was not significantly associated with GC. To our knowledge, this is the first study to comprehensively evaluate the association between the variants in DNA repair pathway genes and risk of UGI cancer.

The DNA repair pathways can be divided into functionally interwoven subpathways that are distinguished by the type of DNA lesion

Table III. The association between DNA repair pathway genes and risk of gastric adenocarcinoma in the Shanxi, NIT and combined populations[a]

Gene	Location	Gene-level analysis			No. of SNP	Most significant SNP analysis						
		Shanxi	NIT	Combined		dbSNP id (major, minor allele)	Shanxi		NIT		Combined	
		P	*P*	*P*			OR	*P*	OR	*P*	OR	*P*
Total												
CLK2	1q21	3.83×10^{-4}	0.41	3.02×10^{-4}	2	rs1052176 (A, C)	0.79	3.02×10^{-4}	0.89	0.36	0.81	2.59×10^{-4}
MRE11A	11q21	0.049	0.10	6.06×10^{-3}	12	rs13447720 (A, G)	1.18	8.32×10^{-3}	1.31	0.032	1.20	8.86×10^{-4}
RAD54L	1p34	0.021	0.29	0.010	6	rs17102087 (C, T)	0.80	4.87×10^{-3}	0.83	0.25	0.80	2.27×10^{-3}
POLE	12q24	0.014	0.15	0.033	10	rs11614717 (G, T)	1.22	3.59×10^{-3}	1.02	0.86	1.17	8.46×10^{-3}
RECQL5	17q25	8.61×10^{-3}	0.14	0.050	4	rs1093992 (G, A)	1.26	2.55×10^{-3}	0.95	0.74	1.18	0.016
POLB	8p11	0.022	0.77	0.054	3	rs10958713 (T, C)	0.88	0.010	1.01	0.91	0.90	0.026
MSH2	2p21	0.013	0.66	0.087	16	rs1421 (A, G)	0.81	1.28×10^{-3}	1.09	0.50	0.86	9.61×10^{-3}
HEL308	4q21	0.028	0.80	0.094	5	rs1494961 (T, C)	0.87	9.98×10^{-3}	1.06	0.59	0.90	0.035
SPO11	20q13	0.041	1.00	0.094	4	rs2865367 (C, A)	0.61	0.014	1.06	0.89	0.69	0.035
MAD2L2	1p36	0.039	0.62	0.24	4	rs2336031 (G, A)	1.26	0.012	0.80	0.24	1.15	0.088
REV1	2q11	0.023	0.74	0.44	12	rs3087395 (A, G)	0.87	0.047	1.05	0.72	0.90	0.10
Cardia cancer												
CLK2	1q21	2.87×10^{-3}	0.30	1.57×10^{-3}	2	rs1052176 (A, C)	0.79	2.25×10^{-3}	0.85	0.26	0.80	1.23×10^{-3}
MRE11A	11q21	0.019	0.043	7.63×10^{-4}	12	rs13447720 (A, G)	1.23	3.47×10^{-3}	1.41	0.013	1.27	1.62×10^{-4}
RAD54L	1p34	0.063	0.31	0.021	6	rs17102087 (C, T)	0.80	0.016	0.77	0.18	0.79	4.67×10^{-3}
POLE	12q24	0.11	0.089	0.25	10	rs11614717 (G, T)	1.18	0.032	0.96	0.75	1.13	0.080
Other significant genes for cardia cancer, but not significant for the GC overall												
BRCA2	13q13	0.013	0.79	7.12×10^{-3}	29	rs9590896 (G, A)	1.30	7.45×10^{-4}	1.20	0.18	1.27	3.82×10^{-4}
Non-cardia cancer												
CLK2	1q21	6.84×10^{-3}	0.87	0.010	2	rs1052176 (A, C)	0.78	5.96×10^{-3}	0.89	0.60	0.80	8.23×10^{-3}
MRE11A	11q21	0.16	0.75	0.11	12	rs13447720 (A, G)	1.08	0.35	1.14	0.56	1.08	0.31
RAD54L	1p34	0.20	0.71	0.25	6	rs17102087 (C, T)	0.81	0.057	1.08	0.76	0.84	0.093
POLE	12q24	0.038	0.46	0.023	10	rs11614717 (G, T)	1.26	9.71×10^{-3}	1.25	0.33	1.25	5.79×10^{-3}
Other significant genes for cardia cancer, but not significant for the GC overall												
ERCC2	19q13	0.020	0.14	0.012	4	rs50871 (T, G)	0.79	5.82×10^{-3}	0.84	0.32	0.80	3.20×10^{-3}
XPA	9q22	0.022	0.83	0.047	7	rs7045160 (T, C)	1.40	8.97×10^{-3}	1.09	0.78	1.33	0.015

[a]Genes were ranked by the gene-level *P*-value for GC overall. Four significant SNPs for GC overall were evaluated by subsite with the same order as in GC overall. Gene-level *P*-values were calculated using ARTP approach. The *P*-values and ORs for the SNPs were calculated from unconditional logistic regression models using genotype-trend tests adjusted for age (10 year categories), sex and study (for the combined population).

they process (10). Base excision repair plays a key role in repairing subtle DNA damage due to cellular metabolisms (32). Nucleotide excision repair mainly removes bulky single-strand adducts (32). Homologous recombination and non-homologous end-joining cope with double-strand breaks, where the former only acts in the S and G_2 phases of the cell cycle (34). Mismatch excision repair is crucial to DNA repair during replication (10). Although previous studies have provided some insight into possible effects of DNA repair genes in cancer risk, consistent evidence has been sparse (14,21). A systematic review of meta-analyses relating variants in candidate genes to esophageal and GCs identified few robust relations with DNA repair genes (21). GWAS have advanced our understanding regarding the contribution of genetic susceptibility in UGI cancers (5,22–29); however, only *CHEK2* among the DNA repair-related genes has been identified in GWAS to date (5,29). Further review of a continually updated Cancer Genome-wide Association and Meta Analyses database (http://www.hugenavigator.net/CancerGEMKB/caIntegratorStartPage.do, updated on 3 May 2012, retrieved 2 September 2012) did not find other noteworthy associations for DNA repair genes and UGI cancer risk (35). Realizing the limitations of conventional single marker association analyses in identifying true gene associations, pathway analyses have been developed to complement GWAS (30). We, therefore, sought to establish an overall understanding of possible relationships between DNA repair pathway genes and the risk of UGI cancer.

In our study, we observed a significant association for the overall DNA repair pathways with ESCC. Among the 12 individual genes associated with risk of ESCC in the combined data, *CHEK2* on 22q12, which encodes the checkpoint kinase CHK2, was the most significant. Central to transduction of the DNA damage signal, *CHEK2* has been proposed to act as a cancer susceptibility gene (36,37). In our previous GWAS, *CHEK2* was the only DNA repair-related gene that had SNPs significantly associated with risk of ESCC (5), a result confirmed recently by another GWAS (29). The present study extends the SNP-based GWAS results and provides evidence for the role of *CHEK2* from gene-based analyses. *SMUG1* on 12q13 showed significance in both the discovery and replication study sets. The uracil-DNA glycosylase *SMUG1* coordinates the initial steps of base excision repair (BER) and is involved in the removal of 5-fluorouracil incorporated into DNA (38,39). Downregulation of *SMUG1* was indicated as a typical feature of GC with microsatellite instability (40). The most significant SNP in *SMUG1* in our study (rs2029166) was correlated previously with breast cancer risk (41). It remains to be seen whether this association is due to the function of *SMUG1* or other as yet unannotated genes, although a regional LD plot did indicate that the LD block centered on *SMUG1*. *TDG* located on 12q23 is another gene associated with ESCC risk in Shanxi for which the most significant SNP (rs4135054) was replicated in NIT. *TDG* encodes the thymine-DNA glycosylase and is located in a region of high loss of heterozygosity in GC (42). Previous GWAS have correlated SNPs in *ACAD10* (rs11066015), *RPL6* (rs11066280), *C12orf51* (rs2074356), *BRAP* (rs3782886) and *FZD10* (rs12580487) in this region with risk of ESCC (22,24,26). Our online pairwise LD search did not find close linkage between these previously reported SNPs and rs4135054 (*TDG*) (33).

Our analysis highlighted several other ESCC susceptibility genes. A prior study from our group revealed frequent somatic inactivation of *TP53* in ESCC tissues in this population (43), and our observation of associations between SNPs in *TP53* and ESCC provide additional evidence for this gene. We also showed associations for several other genes, including *GTF2H3*, *FEN1*, *POLQ*, *HEL308*, *RAD54B*, *MPG*, *FANCE* and *BRCA1*, with risk of ESCC. Further studies are warranted to validate these results.

We observed four significant genes for GC. *CLK2* located at 1q21 was significantly associated with risk of GC overall as well as in both cardia and non-cardia subsites. CLK2 protein has been identified as a component of hepatic insulin signaling and glucose metabolism (44). The most significant SNP rs1052176 of *CLK2* is also in LD with regions of *HCN3*, *PKLR*, *FDPS* and several other genes (33); a deeper investigation into this region appears warranted. *MRE11A* was significantly associated with risk of GC, and the most significant SNP (rs13447720) replicated in the NIT. *MRE11A* at 11q21 encodes a nuclease involved in homologous recombination and telomere length maintenance, and it is reported that *MRE11A* expression is impaired in GC with microsatellite instability (45). We observed significant associations for *MRE11A* and risk of GC overall as well as for cardia cancers, but not for non-cardia cancers. It is worth noting that *MRE11A* was the only gene significantly associated with GC in both discovery and replication data sets, a finding limited to cardia cancer. Two other genes, *RAD54L* (1p34) and *POLE* (12q24), were also associated with GC in the combined data sets. However, *RAD54L* was only associated with cardia cancer, and *POLE* was only with non-cardia cancer.

Three other DNA repair genes that were not associated with GC risk overall were significantly related to risk in subsite analyses, suggesting possible different genetic susceptibility genes for cardia cancer and non-cardia cancer. *BRCA2*, a key breast cancer susceptibility gene that cooperates with *RAD51* in homologous recombination (32), was associated only with risk of cardia cancer, whereas *ERCC2* and *XPA*, which both act in nucleotide excision repair (32), were associated only with risk of non-cardia cancer.

The causes of the geographic correlation of UGI tumors, particularly ESCC and gastric cardia cancer, in the Taihang Mountain area have not been determined. In our study, we did not identify additional shared genes beyond *PLCE1*, which was reported previously as associated with both ESCC and gastric cardia cancer (5). *BRCA2* showed some common association. It was significantly associated with both ESCC and cardia cancer in Shanxi data. Stratified analysis observed a significant interaction between rs9590896 in *BRCA2* and gender on ESCC risk, where the significant association only existed in men. This might contribute to the observed difference in effect of this gene between Shanxi and NIT because the Shanxi study has a higher proportion of men. A previous study from our group found more frequent *BRCA2* germline mutations in ESCC cases with a family history of UGI cancer, suggesting a possible role for *BRCA2* in genetic susceptibility to familial ESCC (46). However, the present study failed to observe an interaction between rs9590896 and family history of UGI cancer for either ESCC or GC.

Although risk factors for ESCC remain poorly understood in very high-risk regions such as north central China, nutritional inadequacies and thermal damage have been associated with risk of ESCC (6,7). The roles of predominantly *H.pylori* infection are well established in the development of GC (6–8). In esophageal and gastric carcinogenesis, exposure to carcinogens with consequent accumulation of DNA damage accompanied by deficits in DNA repair capacity from genetic polymorphisms presumably leads to genetic instability and carcinogenesis (10–13,47,48). In contrast to the findings in Western populations, smoking was only moderately associated with ESCC and GC risk in the Taihang Mountain area (6,7). To further evaluate smoking, we examined the possible interactions between smoking and selected SNPs. The results indicated a possible interaction between smoking and rs2029166 (*SMUG1*) on risk of ESCC and GC where the significant association with this SNP was only observed among non-smokers. The other possible effect modification by smoking was observed for rs11614717 (*POLE*) on GC risk. As this study was an association study, further functional studies are warranted to elucidate the mechanisms underlying the observed associations between the genetic variants in DNA repair pathways and risk of ESCC and GC and possible effect modifications by smoking.

Using a pathway-based approach, we comprehensively investigated associations between DNA repair genes and the risk of ESCC and GC in a high-risk population in north central China. We sought to replicate significant associations observed in the discovery population (Shanxi) in a second replication population (NIT). For the gene- and pathway-level analyses, we applied the resampling-based ARTP method to reduce the false positive rate from multiple comparisons. We also acknowledge limitations. Sample size in the NIT was only modest, which limited power in the replication analyses. Generalizability to other populations requires caution since our study was conducted among high-risk Han Chinese.

In conclusion, by taking into account prior biological knowledge, our pathway-based analysis identified significant associations between several DNA repair pathway genes and the risk of ESCC and

GC, providing further evidence for the role of genetics in the etiology of UGI cancer. Confirmation of these findings in other populations, combined with advances in our understanding of the molecular mechanisms underlying these associations, is needed to solidify our understanding of the role of DNA repair genes in UGI carcinogenesis.

Supplementary material

Supplementary Tables S1–S5 and Figure S1 can be found at http://carcin.oxfordjournals.org/

Funding

The Shanxi Upper Gastrointestinal Cancer Genetics Project was supported by the National Cancer Institute contract NO2-SC-66211 with the Shanxi Cancer Hospital and Institute. The Nutrition Intervention Trials were supported by National Cancer Institute contracts NO1-SC-91030 and HHSN261200477001C with the Cancer Institute of the Chinese Academy of Medical Sciences. The current analysis was supported by the Intramural Research Program of the National Institute of Health, National Cancer Institute and the Division of Cancer Epidemiology and Genetics.

Acknowledgements

We acknowledge the editorial assistance of the NIH Fellows Editorial Board. W.-Q.L., A.M.G. and P.R.T. conceived and designed the study. W.-Q.L., K.Y., J.X.S., W.W. and X.-Q.X. conducted data analysis. W.-Q.L. and P.R.T. wrote the manuscript. N.H., Z.-M.W., H.S., C.-Y.W., L.-M.W., S.J.C., L.B., T.D., Y.-L.Q., J.-H.F., Y.W., Y.X. and C.G. contributed significantly to genotyping and data acquisition. N.H., P.L.H., Y.G., F.G., M.A.T., S.M.D., N.D.F., C.C.A., A.M.G. and P.R.T. critically revised the manuscript for important intellectual content. P.R.T. supervised this study.

Conflicts of interest statement: None declared.

References

1. Jemal,A. *et al.* (2011) Global cancer statistics. *CA Cancer J. Clin*., **61**, 69–90.
2. Ferlay,J. *et al.* (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*, **127**, 2893–2917.
3. Li,J.Y. *et al.* (1981) Atlas of cancer mortality in the People's Republic of China. An aid for cancer control and research. *Int. J. Epidemiol*., **10**, 127–133.
4. Li,B. *et al.* (1993) Linxian nutrition intervention trials. Design, methods, participant characteristics, and compliance. *Ann. Epidemiol*., **3**, 577–585.
5. Abnet,C.C. *et al.* (2010) A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat. Genet*., **42**, 764–767.
6. Tran,G.D. *et al.* (2005) Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int. J. Cancer*, **113**, 456–463.
7. Gao,Y. *et al.* (2011) Risk factors for esophageal and gastric cancers in Shanxi Province, China: a case control study. *Cancer Epidemiol*., **35**, e91–e99.
8. Nyren,O. *et al.* (2006) *Helicobacter pylori* infection: mainly foe but also friend? *J. Natl Cancer Inst*., **98**, 1432–1434.
9. Freedman,N.D. *et al.* (2007) A prospective study of tobacco, alcohol, and the risk of esophageal and gastric cancer subtypes. *Am. J. Epidemiol*., **165**, 1424–1433.
10. Lord,C.J. *et al.* (2012) The DNA damage response and cancer therapy. *Nature*, **481**, 287–294.
11. Halazonetis,T.D. *et al.* (2008) An oncogene-induced DNA damage model for cancer development. *Science*, **319**, 1352–1355.
12. Toller,I.M. *et al.* (2011) Carcinogenic bacterial pathogen *Helicobacter pylori* triggers DNA double-strand breaks and a DNA damage response in its host cells. *Proc. Natl Acad. Sci. U.S.A*., **108**, 14944–14949.
13. Bonn,D. (1998) How DNA-repair pathways may affect cancer risk. *Lancet*, **351**, 42.
14. Goode,E.L. *et al.* (2002) Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol. Biomarkers Prev*., **11**, 1513–1530.
15. Li,W.Q. *et al.* (2009) Association between genetic polymorphisms of DNA base excision repair genes and evolution of precancerous gastric lesions in a Chinese population. *Carcinogenesis*, **30**, 500–505.
16. Capella,G. *et al.* (2008) DNA repair polymorphisms and the risk of stomach adenocarcinoma and severe chronic gastritis in the EPIC-EURGAST study. *Int. J. Epidemiol*., **37**, 1316–1325.
17. He,J. *et al.* (2012) Polymorphisms in the XPG gene and risk of gastric cancer in Chinese populations. *Hum. Genet*., **131**, 1235–1244.
18. Liu,L. *et al.* (2012) Functional FEN1 genetic variants contribute to risk of hepatocellular carcinoma, esophageal cancer, gastric cancer and colorectal cancer. *Carcinogenesis*, **33**, 119–123.
19. Hao,B. *et al.* (2004) Identification of genetic variants in base excision repair pathway and their associations with risk of esophageal squamous cell carcinoma. *Cancer Res*., **64**, 4378–4384.
20. Pan,J. *et al.* (2009) Genetic susceptibility to esophageal cancer: the role of the nucleotide excision repair pathway. *Carcinogenesis*, **30**, 785–792.
21. Dong,L.M. *et al.* (2008) Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. *JAMA*, **299**, 2423–2436.
22. Wang,L.D. *et al.* (2010) Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and C20orf54. *Nat. Genet*., **42**, 759–763.
23. Shi,Y. *et al.* (2011) A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat. Genet*., **43**, 1215–1218.
24. Wu,C. *et al.* (2011) Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. *Nat. Genet*., **43**, 679–684.
25. Sakamoto,H. *et al.* (2008) Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.*, **40**, 730–740.
26. Cui,R. *et al.* (2009) Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. *Gastroenterology*, **137**, 1768–1775.
27. Abnet,C.C. *et al.* (2012) Genotypic variants at 2q33 and risk of esophageal squamous cell carcinoma in China: a meta-analysis of genome-wide association studies. *Hum. Mol. Genet*., **21**, 2132–2141.
28. Tanaka,F. *et al.* (2010) Strong interaction between the effects of alcohol consumption and smoking on oesophageal squamous cell carcinoma among individuals with ADH1B and/or ALDH2 risk alleles. *Gut*, **59**, 1457–1464.
29. Wu,C. *et al.* (2012) Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat. Genet*., **44**, 1090–1097.
30. Wang,K. *et al.* (2010) Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet*., **11**, 843–854.
31. Yu,K. *et al.* (2009) Pathway analysis by adaptive combination of P-values. *Genet. Epidemiol*., **33**, 700–709.
32. Wood,R.D. *et al.* (2001) Human DNA repair genes. *Science*, **291**, 1284–1289.
33. Johnson,A.D. *et al.* (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938–2939.
34. O'Driscoll,M. *et al.* (2006) The role of double-strand break repair—insights from human genetics. *Nat. Rev. Genet*., **7**, 45–54.
35. Schully,S.D. *et al.* (2011) Cancer GAMAdb: database of cancer genetic associations from meta-analyses and genome-wide association studies. *Eur. J. Hum. Genet*., **19**, 928–930.
36. Antoni,L. *et al.* (2007) CHK2 kinase: cancer susceptibility and cancer therapy—two sides of the same coin? *Nat. Rev. Cancer*, **7**, 925–936.
37. Zhang,B. *et al.* (2011) Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol*., **12**, 477–488.
38. Pettersen,H.S. *et al.* (2007) Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res*., **35**, 3879–3892.
39. An,Q. *et al.* (2007) 5-Fluorouracil incorporated into DNA is excised by the Smug1 DNA glycosylase to reduce drug cytotoxicity. *Cancer Res*., **67**, 940–945.
40. D'Errico,M. *et al.* (2009) Genome-wide expression profile of sporadic gastric cancers with microsatellite instability. *Eur. J. Cancer*, **45**, 461–469.
41. Marian,C. *et al.* (2011) Single nucleotide polymorphisms in uracil-processing genes, intake of one-carbon nutrients and breast cancer risk. *Eur. J. Clin. Nutr*., **65**, 683–689.
42. Schmutte,C. *et al.* (1997) Human thymine-DNA glycosylase maps at chromosome 12q22-q24.1: a region of high loss of heterozygosity in gastric cancer. *Cancer Res*., **57**, 3010–3015.

43. Hu,N. *et al.* (2001) Frequent inactivation of the TP53 gene in esophageal squamous cell carcinoma from a high-risk population in China. *Clin. Cancer Res*., **7**, 883–891.
44. Rodgers,J.T. *et al.* (2010) Cdc2-like kinase 2 is an insulin-regulated suppressor of hepatic gluconeogenesis. *Cell Metab*., **11**, 23–34.
45. Ottini,L. *et al.* (2004) MRE11 expression is impaired in gastric cancer with microsatellite instability. *Carcinogenesis*, **25**, 2337–2343.
46. Hu,N. *et al.* (2004) Evaluation of BRCA2 in the genetic susceptibility of familial esophageal cancer. *Oncogene*, **23**, 852–858.
47. Farinati,F. *et al.* (1998) Oxidative DNA damage accumulation in gastric carcinogenesis. *Gut*, **42**, 351–356.
48. Shao,L. *et al.* (2006) Deficiency of cell cycle checkpoints and DNA repair system predispose individuals to esophageal cancer. *Mutat. Res*., **602**, 143–150.

Received December 27, 2012; revised February 21, 2013; accepted March 12, 2013