Carcinogenesis vol.32 no.3 pp.336–342, 2011
doi:10.1093/carcin/bgq264
Advance Access publication December 10, 2010

Association of candidate genetic variations with gastric cardia adenocarcinoma in Chinese population: a multiple interaction analysis

Li Liu[1,†], Chen Wu[2,†], Ying Wang[1], Rong Zhong[1], Feng Wang[3], Xuemei Zhang[4], Shengyu Duan[1], Jiao Lou[1], Dianke Yu[2], Wen Tan[2], Jing Yuan[3], Tangchun Wu[3], Shaofa Nie[1,†], Xiaoping Miao[1,*] and Dongxin Lin[2]

[1]Department of Epidemiology and Biostatistics and MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China, [2]State Key Laboratory of Molecular Oncology and Department of Etiology & Carcinogenesis, Cancer Institute and Hospital, Chinese Academy of Medical Science and Peking Union Medical College, Beijing 100021, China, [3]Department of Occupational and Environmental Health and MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China and [4]Division of Medical Genetics, College of Medicine, University of Arkansas for Medical Sciences, Arkansas, AR 72205, USA

[*]To whom correspondence should be addressed. Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China.
Tel: +86 27 83650744; Fax: +86 27 83650744;
Email: miaoxp@mail.hust.edu.cn
Correspondence may also be addressed to Dongxin Lin.
Tel: +86 10 87788491; Fax: +86 10 67722460;
Email: lindx72@cicams.ac.cn

Single genetic variation may only have a modest effect on risk of gastric cardia adenocarcinoma (GCA) because this malignancy is believed to result from complex interactions among multiple genetic and environmental factors. However, it has been a challenge to characterize multiple interactions using parametric analytic approaches. This study utilized a multi-analytic strategy combining logistic regression (LR), multifactor dimensionality reduction (MDR) and classification and regression tree (CART) approaches to explore high-order interactions among smoking and 12 polymorphisms involved in different processes of carcinogenesis in 344 GCA patients and 324 controls. LR, MDR and CART analyses consistently suggested *MMP-2* C–1306T polymorphism as the strongest individual factor for GCA risk. Intriguingly, a high-order interaction was consistently identified by MDR, LR and CART analyses. In MDR analysis, the three-factor model including *MMP-2* C–1306T, *FASL* T–844C and *FAS* G–1377A yielded the highest testing accuracy of 0.632. When analysing combined effect of these three polymorphisms by LR, a significant gene dose effect was observed with the odds ratios (ORs) being increased with increasing numbers of risk genotypes ($P_{trend} = 4.736 \times 10^{-12}$). In CART analysis, individuals carrying the combined genotypes of *MMP-2* –1306CC, *FASL*–844TT or TC and *FAS* –1377AA had the highest risk for GCA (OR = 4.58; 95% confidence interval, 2.07–10.14) compared with the lowest risk carriers of the *MMP-2* –1306CT or TT genotype. These results suggest that *MMP-2* C–1306T polymorphism is an important risk factor for GCA and the multifactor interactions among polymorphisms in *MMP-2*, *FASL* and *FAS* play more important role in the development of GCA.

Abbreviations: CART, classification and regression tree; CI, confidence interval; CVC, cross-validation consistency; GCA, gastric cardia adenocarcinoma; LR, logistic regression; MDR, multifactor dimensionality reduction; MMP, matrix metalloproteinase; OR, odds ratio; SNP, single-nucleotide polymorphism.

[†]These authors contributed equally to this work.

Introduction

Gastric cardia adenocarcinoma (GCA) represents the second leading cause of cancer-related death worldwide, with >700 000 deaths each year (1). In the last two decades, although incidence rate of non-cardia gastric cancer has slightly declined, GCA has shown a significantly increased trend in China (2). Since most patients in early stage of GCA have no obvious symptoms, the early diagnosis of GCA is challenging. Therefore, despite the advances in treatment strategies, GCA still has a dismal prognosis, with much lower 5 years survival compared with cancer at the pyloric antrum (3). The rapid increase in incidence and poor prognosis of GCA highlight the importance of prevention against the disease. Epidemiological studies have suggested several environmental factors that are involved in the development of GCA, including cigarette smoking, alcohol consumption, pathogenic infections and dietary carcinogen exposure (4,5). However, only a fraction of exposed individuals actually developed GCA during their lifespan, suggesting that this malignancy may result from interactions of multiple environmental and genetic factors.

The development of GCA, among other cancers, has been thought to be of multisteps (5). The initiation of carcinogenesis is probably caused by metabolic activation of carcinogens via phase I enzymes that convert carcinogens into DNA-damaging metabolites (6). Competing with this is the metabolic detoxification of carcinogens to harmless excreted products primarily by phase II enzymes (7). The balance between metabolic activation and detoxification of carcinogens by various enzymes varies among individuals and has been shown to modulate cancer susceptibility (5). If cellular DNA damages can be fixed by repair system, the cell is returned to its normal state. On the other hand, the cell with permanently damaged DNA may be arrested by cell cycle control and removed by apoptosis or programmed cell death. However, if DNA damages escape from cellular repair mechanisms and persist, permanent mutation and malignant transformation may occur to the cell, resulting in cancer under the circumstance of the loss of normal growth control (8). Additionally, many studies have shown that aberrant extracellular proteolysis might also be implicated in cancer development. For example, elevated expression of matrix metalloproteinases (MMPs) has been recognized as a critical modulator for cancer initiation and development by regulating various signaling pathways involved in cell growth, differentiation, apoptosis, angiogenesis and immune surveillance (9,10). Previous studies have shown that functional variations in genes involved in carcinogen metabolism (e.g. *CYP2E1* and *SULT1A1*), DNA repair (e.g. *XPC* and *XRCC1*), cell cycle control (e.g. *STK15*), apoptosis (e.g. *P53*, *FAS* and *FASL*) and proteolysis (e.g. *MMP-2*) may be individually associated with susceptibility to GCA (6,11–17).

Although certain genetic polymorphisms have been associated with GCA susceptibility, the previous studies, which usually used single gene- or polymorphism-based design, often yield conflicting results across different studies. Furthermore, since cancers are complex diseases involving multiple genetic variations and gene–environment interactions, not a single locus can fully explain their genetic susceptibility. Some previous studies suggested the possibility of gene–gene and gene–environment interactions; however, these interactions are difficult to fully characterize using traditional analytic strategies such as logistic regression (LR) since sparseness of the data in high dimensions would occur to result in inaccuracy parameter estimates for identifying interactions. Additionally, statistic power would decrease and type II errors would increase when detecting interactions by LR in relatively small sample size (18,19). Recently, two non-parametric data mining approaches, multifactor dimensionality reduction (MDR) and classification and regression tree (CART) have been

documented to overcome the inaccuracy parameter estimates and low power of LR for detecting interactions and have good power for identifying high-order interactions (18,20). Given the statistical advantages in identifying complex interactions, MDR and CART have been applied to explore high-order gene–gene and gene–environment interactions in modulating risk of various types of cancer (21–23).

In this study, we investigated a panel of 12 well-studied polymorphisms in nine genes involved in carcinogen metabolism, DNA repair, cell cycle control, apoptosis and proteolysis (6,11–17) in a case–control study of 344 GCA patients and 324 cancer-free controls. We examined the individual and combined effects of these 12 polymorphisms by traditional unconditional LR model and high-order gene–gene and gene–smoking interactions in modulating GCA risk using MDR and CART analyses.

Patients and methods

Study subjects

This study consisted of 344 patients with newly diagnosed GCA and 324 cancer-free controls. All subjects were unrelated ethnic Han Chinese. Patients were consecutively recruited between January 1997 and July 2003 at the Cancer hospital, Chinese Academy of Medical Sciences (Beijing). GCA was defined as tumor arising at the gastric cardia and/or gastroesophageal junction with or without involvement of other esophageal and/or gastric subsites. Tumors confined entirely to the esophageal or gastric subsites were excluded. All patients with histopathologically confirmed GCA were enrolled, yielding a response rate of 94%. Controls were cancer-free individuals living in Beijing region, who were randomly selected from an endoscopic screening campaign in the same time period as the patients were enrolled, with a response rate of 96%. Controls were frequency matched to patients for age (±5 years) and sex. At recruitment, informed consent was obtained from each subject and the information on demographic characteristics, such as sex, age and smoking habit, was collected by questionnaire. Subjects who had never smoked or smoked <1 cigarette per day and <1 year in their lifetime were defined as non-smokers; otherwise, they were considered as smokers (including current smokers and ex smokers). For smokers, the number of pack-years smoked was calculated to indicate the cumulative smoking dose [pack-years = (cigarettes per day/20) × (years smoked)]. Light and heavy smokers were categorized by using the median pack-year value of the controls as the cut points. This study was conducted under the approval of the institutional review board of the Chinese Academy of Medical Science Cancer Institute.

Genotype determination

On the basis of previous functional and epidemiological studies and common frequency in the Chinese population (6,11–17), we selected a total of 12 literature-defined functional polymorphisms in nine important genes involved in different processes of carcinogenesis. These included two single-nucleotide polymorphisms (SNPs) in carcinogen metabolism genes (*CYP2E1* G–1293C rs3813867 and *SULT1A1* Arg213His rs9282861), five SNPs in the DNA repair genes (*XRCC1* T–77C rs3213245, Arg194Trp rs1799782, Arg280His rs25489, Arg399Gln rs25487 and *XPC* PAT+/–), one SNP in cell cycle control gene (*STK15* Phe31Ile rs2273535), three SNPs in apoptotic genes (*P53* Arg72Pro rs1042522, *FAS* G–1377A rs2234767 and *FASL* T–844C rs763110) and one SNP in extracellular proteolytic gene (*MMP-2* C–1306T rs243865) (supplementary Table 1 is available at *Carcinogenesis* Online). Genomic DNA was extracted from 5 ml blood sample that was collected from each participant at recruitment. Genotypes of all polymorphisms were determined by polymerase chain reaction–restriction fragment length polymorphism assays except for the *XPC* poly AT insertion/deletion polymorphism (PAT+/–) and *MMP-2* C–1306T polymorphism, which were analyzed by polymerase chain reaction–amplified fragment length polymorphism and polymerase chain reaction–denatured high performance liquid chromatography, respectively (24,25). Genotyping was performed without knowledge of case/control status of the subjects. A 10% masked, random sample of cases and controls were tested twice by different people and the reproducibility was 100%.

Statistical analysis

Difference in the distribution of demographic characteristics and genotype frequencies between cases and controls were evaluated using the χ^2-test, Fisher's exact test and *t*-test, where appropriate. Hardy–Weinberg equilibrium for genotypes was tested in controls by a goodness-of-fit χ^2-test. Unconditional multivariate LR was used to estimate odds ratios (ORs) and their 95% confidence intervals (CIs) adjusting for age, sex, smoking status or pack-years, where appropriate. The potential gene–environment interaction between polymorphism and cigarette smoking was evaluated by genotype-smoking combined effect and multiplicative interaction term that considered cumulative smoking dose as discrete variables: non-smokers, light smokers (≤27 pack-years) and heavy smokers (>27 pack-years). For multiple testing, a powerful bootstrapping method was applied to reduce the potential spurious findings. All statistical analyses were conducted using SPSS software (version 18.0).

MDR analysis

The MDR software (version 2.0 beta 8) and MDR permutation testing software (version 1.0 beta 2) were applied to identify possible high-order interactions associated with GCA risk. The MDR is a non-parametric, genetic model-free method for overcoming some of the limitations of LR (i.e. sample size limitations) for the detection and characterization of gene–gene and gene–environment interactions (18). MDR collapses high-dimensional data into a single dimensional variable with two levels (high and low risk) using the ratio of the number of patients to the number of controls, thereby reducing the high-dimensional data to a single variable and permitting detection of interactions in relatively small sample sizes. The one-dimensional multifactor variable was evaluated for its ability to classify and predict disease status through cross-validation and permutation testing scheme. MDR was utilized to generate a single model that maximized the number of individuals with the proper risk assignment. In this study, the best candidate interaction model was selected across all multilocus models that maximized testing accuracy and the cross-validation consistency (CVC). Furthermore, validation of models as effective predictors of disease status was derived empirically from 1000 permutations, which accounted for multiple comparison testing as long as the entire model fitting procedure was repeated for each randomized dataset to provide an opportunity to identify false positives. The MDR permutation results were considered to be statistically significant at the 0.05 level. All the variables identified in the best model were combined and dichotomized according to the MDR software and their ORs and 95% CIs in relation to GCA risk were calculated. Finally, combined effect of the variables in the best model by the number of risk genotypes was evaluated using LR analysis.

CART analysis

CART analysis was performed using the SPSS software to build a decision tree via recursive partitioning. A CART is constructed by splitting a node into two child nodes repeatedly, beginning with the root node that contains the whole learning sample. Before growing a tree, we choose measure for goodness of split using Gini criteria, by which splits were found that maximize the homogeneity of child nodes with respect to the value of the target variable. After the tree is grown to its full depth, a pruning procedure was performed to avoid overfitting the model. Subgroups of individuals with differential risk associations with GCA were identified in the different terminal nodes of the tree, indicating potential presence of interactions. Finally, the risk of these subgroups was evaluated by using the LR analysis. ORs and 95% CIs were adjusted for age and sex, with treating the least percentage of cases as the reference.

Results

Characteristics of study subjects

Age and sex were frequency matched among 344 GCA cases and 324 cancer-free controls. Distributions of smokers and non-smokers in cases and controls were similar and light or heavy smokers who had smoked ≤27 or >27 pack-years were also not significantly different among cases and controls (Table I).

Table I. Distribution of select characteristics among patients and controls

Variable	Cases (n = 344)	Controls (n = 324)	P
	N (%)	*N* (%)	
Age, years (mean ± SD)	60.25 ± 9.82	59.80 ± 7.83	0.514
Sex			
Male	297 (86.3)	268 (82.7)	0.195
Female	47 (13.7)	56 (17.3)	
Smoking status			
Non-smokers	159 (46.2)	137 (42.3)	0.306
Smokers	185 (53.8)	187 (57.7)	
Smoking level, pack-years			
Light smokers, ≤27	84 (45.4)	98 (52.4)	0.177
Heavy smokers, >27	101 (54.6)	89 (47.6)	

Associations of individual or pairwise factors with GCA risk by LR analysis

Genotype distributions of the selected polymorphisms are shown in Table II and supplementary Table 2 (available at *Carcinogenesis* Online). Genotype distributions of all SNPs in our control subjects conformed to those expected by Hardy–Weinberg equilibrium (all $P > 0.05$) and were similar to those in HapMap CHB database and reported in the previously published studies in Chinese Han population (supplementary Table 1 is available at *Carcinogenesis* Online). The genotype frequencies of five polymorphisms (*SULT1A1* Arg213His, *XRCC1* Arg194Trp, *XPC* PAT+/–, *FASL* T–844C and *MMP-2* C–1306T) in cases were significantly different from those in controls ($P = 3.756 \times 10^{-5}$, $P = 0.033$, $P = 0.022$, $P = 4.726 \times 10^{-5}$ and $P = 4.374 \times 10^{-8}$, respectively). Multivariate LR analysis showed that after adjustment for sex, age and smoking status, increased risk of GCA was significantly associated with the heterozygote (OR = 1.92; 95% CI, 1.33–2.78) or combined variant genotypes (OR = 2.04; 95% CI, 1.41–2.94) of *SULT1A1* Arg213His and the variant homozygote of *XRCC1* Arg194Trp (OR = 2.09; 95% CI, 1.19–3.66) compared with the corresponding wild-type homozygote. In the recessive model, assuming that only the variant homozygote had an increased risk for GCA, significantly higher risk of GCA was presented in individuals with the homozygous variant genotype of *XPC* PAT (OR = 1.79, 95% CI, 1.13–2.82), *FASL* T–844C (OR = 1.93, 95% CI, 1.41–2.63) or *P53* Arg72Pro (OR = 1.44, 95% CI, 1.01–2.06) compared with the respective wild-type homozygous and heterozygous genotype carriers. Individuals carrying the wild-type homozygote of *MMP-2* C–1306T showed a significantly increased risk of GCA (OR = 3.07, 95% CI, 2.09–4.50) compared with those with the combined variant genotypes. Furthermore, the ORs calculated by bootstrapping were all identical to the presented adjusted ORs (Table II).

We further assessed the combined effect or interaction between aforementioned six polymorphisms and trichotomized cumulative smoking dose. As shown in Table III, there was a significant two-way multiplicative interaction between *SULT1A1* Arg213His and smoking ($P = 0.020$), and the interaction was still significant when adjusting for multiple comparisons by bootstrapping ($P = 0.026$). Analysis of combined effect revealed that heavy smokers carrying at least one *SULT1A1* 213His allele had significantly increased GCA risk (OR = 2.19, 95% CI, 1.27–3.76) compared with non-smokers without the *SULT1A1* 213His allele.

Association of high-order interactions with GCA risk by MDR analysis

Table IV shows the best interaction model by MDR analysis. The best one-factor model for predicting GCA risk was *MMP-2* C–1306T SNP (testing accuracy = 0.595, CVC = 10/10, permutation $P < 0.005$). The best two-factor model of *MMP-2* C–1306T and *FASL* T–844C had an improved testing accuracy of 0.616 (CVC = 10/10, permutation $P < 0.005$). The best interaction model was the three-factor model including *MMP-2* C–1306T, *FASL* T–844C and *FAS* G–1377A SNPs, which yielded the highest testing accuracy of 0.632 and the maximal CVC of 10/10 (permutation $P < 0.001$). The four-factor model consisting of *MMP-2* C–1306T, *FASL* T–844C, *FAS* G–1377A and *SULT1A1* Arg213His and five-factor model consisting of *MMP-2* C–1306T, *FASL* T–844C, *FAS* G–1377A, *SULT1A1* Arg213His and *XRCC1* Arg399Gln also improved testing accuracy compared with the one-factor model (permutation $P = 0.006$ and $P < 0.013$, respectively); however, the CVCs were decreased (9/10 and 6/10, respectively).

For the three SNPs identified in the best interaction model, *MMP-2* C–1306T, *FASL* T–844C and *FAS* G–1377A were combined and dichotomized according to the MDR software. Individuals carrying the combined risk stratum had a 2.99-fold increased risk for GCA

Table II. Association between candidate variants and GCA risk

Genotype	Cases	Controls	P^a	OR[b] (95% CI)	OR[c] (95% CI)	*P* for bootstrap
	n (%)	*n* (%)				
SULT1A1 Arg213His						
Arg/Arg	237 (68.9)	266 (82.1)	3.76×10^{-5}	1.00	1.00	0.001
Arg/His + His/His[d]	107 (31.1)	58 (17.9)		2.04 (1.41–2.94)	2.04 (1.40–2.98)	
XRCC1 Arg194Trp						
Arg/Arg	155 (45.1)	160 (49.4)	0.033	1.00	1.00	
Arg/Trp	145 (42.2)	142 (43.8)		1.06 (0.77–1.47)	1.06 (0.78–1.49)	0.706
Trp/Trp	44 (12.7)	22 (6.8)		2.09 (1.19–3.66)	2.08 (1.23–3.83)	0.010
XPC PAT+/–						
–/–	145 (42.2)	133 (41.0)	0.022	1.00	1.00	
–/+	140 (40.7)	157 (48.5)		0.83 (0.60–1.15)	0.83 (0.58–1.15)	0.262
+/+	59 (17.1)	34 (10.5)		1.62 (1.00–2.64)	1.62 (1.01–2.78)	0.050
P53 Arg72Pro						
Arg/Arg	94 (27.3)	96 (29.6)	0.119	1.00	1.00	
Arg/Pro	153 (44.5)	159 (49.1)		0.98 (0.69–1.41)	1.00 (0.70–1.48)	0.991
Pro/Pro	97 (28.2)	69 (21.3)		1.43 (0.94–2.18)	1.44 (0.97–2.27)	0.097
FASL T–844C						
TT	30 (8.7)	31 (9.6)	4.73×10^{-5}	1.00	1.00	
TC	115 (33.4)	160 (49.4)		0.73 (0.42–1.27)	0.72 (0.40–1.32)	0.291
CC	199 (57.9)	133 (41.0)		1.48 (0.85–2.57)	1.48 (0.85–2.73)	0.174
MMP-2 C–1306T						
TT + CT[d]	49 (14.2)	108 (33.3)	4.37×10^{-8}	1.00	1.00	
CC	295 (85.8)	216 (66.7)		3.07 (2.09–4.50)	3.06 (2.06–4.72)	0.001

[a]*P* values were calculated by χ^2-test or Fisher's exact test.
[b]Data were calculated by LR model and adjusted for sex, age and smoking status.
[c]Data were calculated by bootstrap with 1000 replications.
[d]There were six cases and zero controls with the *SULT1A1* His/His genotype and three cases and eight controls with the *MMP-2* TT genotype.

Table III. Analyses of interaction between genotypes and cumulative smoking dose

Genotype	OR[a] (95% CI)			P[b] for interaction
	Non-smokers	Light smokers	Heavy smokers	
SULT1A1 Arg213His				
Arg/Arg	1.00	0.75 (0.48–1.17)	0.92 (0.59–1.43	0.020
Arg/His + His/His	1.96 (0.97–3.99)	1.28 (0.64–2.55)	2.19 (1.27–3.76)	
XRCC1 Arg194Trp				
Arg/Arg	1.00	0.89 (0.51–1.55)	1.02 (0.59–1.74)	0.651
Arg/Trp + Trp/Trp	1.53 (0.96–2.44)	0.80 (0.47–1.36)	1.14 (0.66–1.95)	
XPC PAT+/–				
(–/–) + (–/+)	1.00	0.70 (0.46–1.07)	0.85 (0.56–1.30)	0.726
(+/+)	2.11 (1.02–4.39)	0.93 (0.38–2.29)	1.51 (0.71–3.22)	
P53 Arg72Pro				
Arg/Arg + Arg/Pro	1.00	0.77 (0.49–1.20)	0.99 (0.63–1.55)	0.713
Pro/Pro	1.77 (1.05–2.98)	0.93 (0.44–1.98)	1.10 (0.58–2.09)	
FASL T–844C				
TT + TC	1.00	0.65 (0.37–1.14)	0.88 (0.51–1.50)	0.472
CC	1.85 (1.16–2.95)	1.32 (0.76–2.28)	1.77 (1.01–3.13)	
MMP-2 C–1306T				
CT + TT	1.00	0.78 (0.34–1.82)	0.87 (0.38–1.99)	0.562
CC	3.33 (1.91–5.82)	2.03 (1.11–3.71)	2.69 (1.48–4.92)	

[a]Data calculated by LR, adjusted for sex, age and smoking status.
[b]P values for gene–smoking interaction were calculated using the multiplicative interaction term in SPSS software.

Table IV. Interaction models by MDR analysis

Number of risk factors	Best interaction model[a]	Testing accuracy	CVC	P for permutation test
1	*MMP-2*	0.595	10/10	0.020
2	*MMP-2*, *FASL*	0.616	10/10	<0.001
3	*MMP-2*[a], *FASL* and *FAS*	0.632	10/10	<0.001
4	*MMP-2*, *FASL*, *FAS* and *SULT1A1*	0.623	9/10	<0.001
5	*MMP-2*, *FASL*, *FAS*, *SULT1A1* and *XRCC1-4*	0.628	6/10	<0.001

XRCC1-4, *XRCC1* Arg399Gln polymorphism.
[a]The best model was selected as the onewith the maximum testing accuracy and maximum CVC. In this study, the best interaction model was the three-factor model including *MMP-2* C–1306T, *FASL* T–844C and *FAS* G–1377A polymorphisms.

Table V. Combined effect of FAS G-1377A, FASL T-844C and MMP-2 C-1306T on GCA risk

Number of risk genotypes[a]	Cases	Controls	OR[b] (95% CI)	P	OR[c] (95% CI)	P
	N (%)	N (%)				
0 (Group 1)	20 (5.8)	54 (16.7)	1.00		1.00	
1 (Group 2)	122 (35.5)	162 (50.0)	2.03 (1.16–3.58)	0.014	2.07 (1.18–3.65)	0.012
2 (Group 3)	175 (50.9)	97 (29.9)	4.87 (2.76–8.61)	5.167×10^{-8}	4.85 (2.74–8.60)	6.109×10^{-8}
3 (Group 4)	27 (7.8)	11 (3.4)	6.63 (2.78–15.80)	1.984×10^{-5}	6.60 (2.76–15.76)	2.210×10^{-5}
P_{trend}					4.736×10^{-12}	

[a]The *MMP-2* –1306CC, *FASL* –844CC and *FAS* –1377AA genotypes were considered as risk genotypes. Individuals in group 1 had no risk genotypes; in the next three groups, we pooled all individuals carrying risk genotype in any one gene (Group 1), in any two genes (Group 3) and all the three genes (Group 4).
[b]Unadjusted OR.
[c]OR adjusted for sex, age and smoking status.

(95% CI, 2.18–4.11; $P < 0.001$). Furthermore, a combined effect of *MMP-2* C–1306T, *FASL* T–844C and *FAS* G–1377A was evaluated by LR analysis (Table V) with the *MMP-2* –1306CC, *FASL* –844CC and *FAS* –1377AA genotypes as risk genotypes. Subjects were categorized into four groups based on the number of risk genotypes they carried and those without any risk genotype were designated as the reference group. We found that the OR of GCA for individuals carrying one, two or three risk genotypes was 2.07 (95% CI, 1.18–3.65), 4.85 (95% CI, 2.74–8.60) or 6.60 (95% CI, 2.76–15.76), suggesting a significant gene dose effect ($P_{trend} = 4.736 \times 10^{-12}$).

Analysis of gene–environment interaction was also performed using the 12 SNPs and the trichotomized cumulative smoking dose by MDR analysis. However, no significant interaction between smoking and any SNP was found.

Association of high-order interactions with GCA risk by CART analysis

The final resulting tree generated by the CART analysis is shown in supplementary Figure 1 (available at *Carcinogenesis* Online) and Table VI. Consistent with the MDR best one-factor model, the initial

Table VI. Risk estimates of CART terminal nodes

Node	Genotype	Cases	Controls	Case rate[a] (%)	OR[b] (95% CI)	LR *P*	Bootstrap *P*
1	*MMP-2* (HV)	49	108	31.2	1.00		
4	*MMP-2* (W)-*FASL* (V)	175	90	66.0	4.27 (2.80–6.52)	1.789×10^{-11}	0.001
6	*MMP-2* (W)-*FASL* (WH)-*FAS* (V)	23	11	67.6	4.58 (2.07–10.14)	1.747×10^{-4}	0.001
7	*MMP-2* (W)-*FASL* (WH)-*FAS* (WH)-*SULT1A1* (W)	67	93	41.9	1.60 (1.01–2.54)	0.045	0.052
8	*MMP-2* (W)-*FASL* (WH)-*FAS* (WH)-*SULT1A1* (HV)	30	22	57.5	3.03 (1.59–5.79)	0.001	0.001

H, heterozygote; V, variant homozygote; W, wild-type homozygote.
[a]Case rate is the percentage of cancer patients among all individuals in each node.
[b]ORs of terminal nodes were calculated by LR analysis adjusted for age and gender.

split of the root node on the decision tree was *MMP-2* C–1306T SNP, suggesting that this SNP is the strongest risk factor for GCA among the polymorphisms examined. Further inspection of the tree structure revealed distinct interaction patterns between individuals carrying the *MMP-2* –1306CT or TT and those with the *MMP-2* –1306CC genotype. Individuals carrying at least one *MMP-2* –1306T allele (–1306CT or TT genotype) had the lowest risk for GCA with a rate of 31.2% cases. Using the terminal node comprising the *MMP-2* –1306CT or TT genotype carriers as the reference, individuals carrying both the *MMP-2* –1306CC and *FASL* –844CC genotypes exhibited a significantly higher risk for GCA (adjusted OR = 4.27; 95% CI, 2.80–6.52), whereas individuals with the combined genotypes of *MMP-2* –1306CC, *FASL* –844TT or TC and *FAS* –1377AA had the highest risk for GCA (adjusted OR = 4.58; 95% CI, 2.07–10.14). Furthermore, individuals carrying combined genotypes of *MMP-2* –1306CC, *FASL* –844TT or TC, *FAS* –1377GG or GA and *SULT1A1* 213Arg/His or His/His also had significantly increased risk (OR = 3.03, 95% CI, 1.59–5.79).

Discussion

In this study, we applied a multi-analytic strategy combining LR, MDR and CART approaches to systematically examine the associations between GCA risk and a panel of genetic polymorphisms involved in carcinogen metabolism, DNA repair, cell cycle control, apoptosis and extracellular proteolysis. In the single-locus analysis, six polymorphisms showed significant association with GCA risk. Results from LR, MDR and CART analyses consistently suggested that *MMP-2* C–1306T polymorphism was the most important single susceptibility factor for GCA development. Furthermore, LR, MDR and CART analyses also consistently revealed the prediction value of gene–gene interaction among *MMP-2* C–1306T, *FASL* T–844C and *FAS* G–1377A polymorphisms in GCA risk. To the best of our knowledge, this is the first study showing that complex gene–gene interactions may significantly contribute to GCA susceptibility.

Although MMP-2 was traditionally considered to be important in late steps of cancer progression (invasion and metastasis) due to its proteolysis of extracellular matrix, current understanding of the effects of MMP-2 is particularly relevant to both cancer initiation and progression because this enzyme cleaves a diverse non-extracellular matrix substrates to modulate cell proliferation, apoptosis, angiogenesis and immune surveillance (10). For instance, it has been shown that MMP-2 is capable of cleaving insulin-like growth factor binding proteins and releasing insulin-like factors, which is a stimulator of malignant cell proliferation (9). MMP-2 can convert protumor necrosis factor-α to its soluble cytokinetic form, promoting cancer survival (26). MMP-2 may also increase the bioavailability of vascular endothelial growth factor to trigger angiogenesis (27). Additionally, MMP-2 can cleave cell adhesion molecule E-cadherin and liberate transforming growth factor-β to promote epithelial–mesenchymal transition, which is associated with malignant cell invasion (28). Moreover, MMP-2 has been shown to be overexpressed in a wide variety of human cancers, strongly supporting its importance in cancer development (29,30). In our LR single-locus analysis, the *MMP-2* C–1306T SNP showed the strongest association with GCA risk and consistent results were also obtained by both MDR and CART analysis, indicating the *MMP-2* C–1306T variant as the predominant risk factor for GCA. The consistent results across three different analytic approaches demonstrate a true association between this polymorphism and susceptibility to GCA development. The –1306C→T change is located within the *MMP-2* promoter region and abolishes a transcriptional factor Sp1-binding site. Previous study has shown that reporter gene expression driven by the T allelic *MMP-2* promoter was significantly lower than that driven by the C allelic promoter counterpart both in epithelial cells and macrophages (31). Intriguingly, it was found that *Mmp-2*-deficient mice developed fewer tumors than wild-type mice induced by carcinogen stimulus (32); conversely, the induced MMP-2 expression by oncogene-mediated cellular transformation caused increased capacity for malignant progression (33). These data strongly supported that high expression of MMP-2 might increase cancer susceptibility. Because MMP-2 plays an important role in multiple steps of cancer development and high expression of MMP-2 has been shown to be risk factor for cancer, one might expect that individuals carrying the *MMP-2* –1306CC genotype are in higher susceptibility to GCA.

By LR analysis we also found that other five polymorphisms (*SULT1A1* Arg213His, *XRCC1* Arg194Trp, *XPC* PAT+/–, *P53* Arg72Pro and *FASL* T–844C) were significantly associated with GCA risk and the associations remained significant after correction by bootstrapping for multiple comparisons. These results are biologically plausible because all these polymorphisms appear to be of functional significance. It has been shown that SULT1A1 encoded by the *SULT1A1* 213His allele has 2-fold lower catalytic activity and thermo stability than SULT1A1 produced by the *SULT1A1* 213Arg allele, thereby reducing this enzyme activity in metabolic detoxification of carcinogens (34). Due to the neoconservative amino acid change, the *XRCC1* 194Trp variant may have changed XRCC1 function and therefore, resulting in altered DNA repair capacity (35). The *XPC* PAT+ allele has been shown to be in strong linkage disequilibrium with the A allele of an SNP comprising a C to A transition in intron 11 splice acceptor, and the A allele represents lower information content than the C allele and may trigger a deletion that entails the skipping of exon 12, resulting in diminished DNA repair capacity (36). For the *P53* polymorphism, the G to C change at codon 72 causes Arg to Pro amino acid substitution resulting in altered P53 function for apoptosis (37). The T to C change at position –844 in promoter region of *FASL* creates a binding site for the CAAT/enhancer-binding protein β transcription factor, leading to a significantly higher basal *FASL* expression that has been shown to be a common feature of malignant transformation (38).

The most significant finding in the current study was the multiple gene–gene interactions consistently identified by three different statistical approaches. MDR analysis reported the highest prediction accuracy of combined *MMP-2*, *FASL* and *FAS* polymorphisms for GCA risk and consistent results were also obtained by CART analysis, showing a significant interaction of these polymorphisms in the genes. Furthermore, LR analysis indicated GAC risk associated with *MMP-2*, *FASL* and *FAS* polymorphisms in a gene dose dependent manner,

suggesting that the three-factor model consisting of *MMP-2*, *FASL* and *FAS* polymorphisms is the best model for assessment of GAC risk. These results are biologically plausible since FAS and FASL are the receptor-ligand system. Previous studies have shown that these *FAS* and *FASL* polymorphisms have synergistic effect on other types of cancer (39–41). The evidence linking FAS/FASL signaling to MMP-2 activity comes from the study showing that inhibition of FAS/FASL signaling reduces MMP-2 level or proteolytic activity via nuclear factor-kappaB (42). On the other hand, MMP-2 may also play a role in regulating FAS/FASL signaling in tumor cells (10) since MMP-7, another member of MMPs, has been shown to be able to cleave membrane forms of FAS and FASL, reducing the effectiveness of FAS/FASL in triggering apoptosis (43). Taken together, this functional relationship among MMP-2, FASL and FAS might explain why these genes polymorphisms have statistical interaction in enhancing GCA risk. In addition, we also found a significant interaction of SNPs in *MMP-2*, *FASL*, *FAS* and *SULT1A1* in increasing GCA risk by MDR and CART analysis, the underlying mechanism for *SULT1A1* in the interaction remains to be elucidated.

Interestingly, a gene–environment interaction between *SULT1A1* polymorphism and smoking was also revealed in this study. Because SULT1A1 plays an important role in the metabolism of cigarette carcinogens that may promote GCA development (44), a synergistic effect between functional *SULT1A1* variant and smoking is expected. Previous study has shown reduced ability of gastric tissues carrying the *SULT1A1* 213His allele to detoxify cigarette carcinogens (45).

This study might have some limitations. Because it was a hospital-based case–control study, selection bias cannot be fully excluded. It would therefore be important to confirm these findings in a population-based prospective study. The sample size of our study is relatively small although we had >80% power to detect a significant pairwise interaction using LR and MDR analysis based on genotype frequencies and the prevalence of smoking in our study population (46,47). In addition, our study lacked information on exposures to other environmental risk factors such as alcohol use, dietary consumption, pathogenic infections and further studies are needed to investigate whether these factors also interact with genetic variants in GCA risk.

In conclusion, our study highlights the contribution of the complex gene–gene and gene–environment interactions among polymorphisms involved in different processes of carcinogenesis and smoking to GCA susceptibility. Furthermore, population-based or cohort studies are required to confirm these results and more functional analyses are warranted to elucidate the biological plausibility of the complex gene–gene interactions identified in this study.

Supplementary material

Supplementary Figure 1 and Tables 1 and 2 can be found at http://carcin.oxfordjournals.org/

Funding

State Key Basic Research Program (2004CB518701); Program for New Century Excellent Talents in University (NCET-10-0388).

Acknowledgements

We are grateful to all patients and the healthy volunteers for agreeing to participate in the study, as well as all the people who helped us successfully complete the research.

Conflict of Interest Statement: None declared.

References

1. Parkin,D.M. *et al*. (2005) Global cancer statistics, 2002. *CA Cancer J. Clin.*, **55**, 74–108.
2. He,Y.T. *et al*. (2008) Trends in incidence of esophageal and gastric cardia cancer in high-risk areas in China. *Eur. J. Cancer Prev.*, **17**, 71–76.
3. Crew,K.D. *et al*. (2006) Epidemiology of gastric cancer. *World J. Gastroenterol.*, **12**, 354–362.
4. Kelley,J.R. *et al*. (2003) Gastric cancer epidemiology and risk factors. *J. Clin. Epidemiol.*, **56**, 1–9.
5. Wogan,G.N. *et al*. (2004) Environmental and chemical carcinogenesis. *Semin. Cancer Biol.*, **14**, 473–486.
6. Trafalis,D.T. *et al*. (2010) CYP2E1 and risk of chemically mediated cancers. *Expert Opin. Drug Metab. Toxicol.*, **6**, 307–319.
7. Dalhoff,K. *et al*. (2005) Cancer and molecular biomarkers of phase 2. *Methods Enzymol.*, **400**, 618–627.
8. Hecht,S.S. (2003) Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer*, **3**, 733–744.
9. Egeblad,M. *et al*. (2002) New functions for the matrix metalloproteinases in cancer progression. *Nat. Rev. Cancer*, **2**, 161–174.
10. Kessenbrock,K. *et al*. (2010) Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell*, **141**, 52–67.
11. Miao,X. *et al*. (2004) Functional STK15 Phe31Ile polymorphism is associated with the occurrence and advanced disease status of esophageal squamous cell carcinoma. *Cancer Res.*, **64**, 2680–2683.
12. Kotnis,A. *et al*. (2008) Case-control study and meta-analysis of SULT1A1 Arg213His polymorphism for gene, ethnicity and environment interaction for cancer risk. *Br. J. Cancer*, **99**, 1340–1347.
13. Qiu,L. *et al*. (2008) Associations between XPC polymorphisms and risk of cancers: a meta-analysis. *Eur. J. Cancer*, **44**, 2241–2253.
14. Miao,X. *et al*. (2006) Adenosine diphosphate ribosyl transferase and x-ray repair cross-complementing 1 polymorphisms in gastric cardia cancer. *Gastroenterology*, **131**, 420–427.
15. 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.
16. Dai,S. *et al*. (2009) P53 polymorphism and lung cancer susceptibility: a pooled analysis of 32 case-control studies. *Hum. Genet.*, **125**, 633–638.
17. Zhou,R.M. *et al*. (2010) Polymorphisms in promoter region of FAS and FASL gene and risk of gastric cardiac adenocarcinoma. *J. Gastroenterol. Hepatol.*, **25**, 555–561.
18. Hahn,L.W. *et al*. (2003) Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics*, **19**, 376–382.
19. Hosmer,D.W. *et al*. (2000) Applied Logistic Regression. Wiley, New York, NY.
20. Garcia-Magarinos,M. *et al*. (2009) Evaluating the ability of tree-based methods and logistic regression for the detection of SNP-SNP interaction. *Ann. Hum. Genet.*, **73**, 360–369.
21. Wang,W. *et al*. (2007) Genetic variants in cell cycle control pathway confer susceptibility to lung cancer. *Clin. Cancer Res.*, **13**, 5974–5981.
22. Zhai,R. *et al*. (2010) Interactions among genetic variants in apoptosis pathway genes, reflux symptoms, body mass index, and smoking indicate two distinct etiologic patterns of esophageal adenocarcinoma. *J. Clin. Oncol.*, **28**, 2445–2451.
23. Andrew,A.S. *et al*. (2008) DNA repair polymorphisms modify bladder cancer risk: a multi-factor analytic strategy. *Hum. Hered.*, **65**, 105–118.
24. Wang,Y.G. *et al*. (2003) [Poly(AT) polymorphism in DNA repair gene XPC and lung cancer risk]. *Zhonghua Zhong Liu Za Zhi*, **25**, 555–557.
25. Miao,X. *et al*. (2003) A functional polymorphism in the matrix metalloproteinase-2 gene promoter (–1306C/T) is associated with risk of development but not metastasis of gastric cardia adenocarcinoma. *Cancer Res.*, **63**, 3987–3990.
26. Manicone,A.M. *et al*. (2008) Matrix metalloproteinases as modulators of inflammation. *Semin. Cell Dev. Biol.*, **19**, 34–41.
27. Yoon,S.O. *et al*. (2003) Roles of matrix metalloproteinases in tumor metastasis and angiogenesis. *J. Biochem. Mol. Biol.*, **36**, 128–137.
28. Yokoyama,K. *et al*. (2003) Increased invasion and matrix metalloproteinase-2 expression by Snail-induced mesenchymal transition in squamous cell carcinomas. *Int. J. Oncol.*, **22**, 891–898.
29. Li,Y. *et al*. (2009) Overexpression of MMP-2 and MMP-9 in esophageal squamous cell carcinoma. *Dis. Esophagus*, **22**, 664–667.
30. Karadag,N. *et al*. (2008) Expression of matrix metalloproteinases in gallbladder carcinoma and their significance in carcinogenesis. *Appl. Immunohistochem. Mol. Morphol.*, **16**, 148–152.
31. Price,S.J. *et al*. (2001) Identification of novel, functional genetic variants in the human matrix metalloproteinase-2 gene: role of Sp1 in allele-specific transcriptional regulation. *J. Biol. Chem.*, **276**, 7549–7558.
32. Bergers,G. *et al*. (2000) Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nat. Cell Biol.*, **2**, 737–744.
33. Baruch,R.R. *et al*. (2001) Altered matrix metalloproteinase expression associated with oncogene-mediated cellular transformation and metastasis formation. *Cell Biol. Int.*, **25**, 411–420.

34. Nagar,S. *et al*. (2006) Sulfotransferase (SULT) 1A1 polymorphic variants *1, *2, and *3 are associated with altered enzymatic activity, cellular phenotype, and protein degradation. *Mol. Pharmacol.*, **69**, 2084–2092.
35. Au,W.W. *et al*. (2003) Functional characterization of polymorphisms in DNA repair genes using cytogenetic challenge assays. *Environ. Health Perspect.*, **111**, 1843–1850.
36. Khan,S.G. *et al*. (2002) The human XPC DNA repair gene: arrangement, splice site information content and influence of a single nucleotide polymorphism in a splice acceptor site on alternative splicing and function. *Nucleic Acids Res.*, **30**, 3624–3631.
37. Baptiste,N. *et al*. (2002) The proline-rich domain of p53 is required for cooperation with anti-neoplastic agents to promote apoptosis of tumor cells. *Oncogene*, **21**, 9–21.
38. Wu,J. *et al*. (2003) A novel polymorphic CAAT/enhancer-binding protein beta element in the FasL gene promoter alters Fas ligand expression: a candidate background gene in African American systemic lupus erythematosus patients. *J. Immunol.*, **170**, 132–138.
39. Yang,M. *et al*. (2008) Functional variants in cell death pathway genes and risk of pancreatic cancer. *Clin. Cancer Res.*, **14**, 3230–3236.
40. Zhang,B. *et al*. (2007) Functional polymorphisms in FAS and FASL contribute to increased apoptosis of tumor infiltration lymphocytes and risk of breast cancer. *Carcinogenesis*, **28**, 1067–1073.
41. Zhang,X. *et al*. (2005) Functional polymorphisms in cell death pathway genes FAS and FASL contribute to risk of lung cancer. *J. Med. Genet.*, **42**, 479–484.
42. Wisniewski,P. *et al*. (2010) Non-apoptotic Fas signaling regulates invasiveness of glioma cells and modulates MMP-2 activity via NFkappaB-TIMP-2 pathway. *Cell. Signal.*, **22**, 212–220.
43. Almendro,V. *et al*. (2009) The role of MMP7 and its cross-talk with the FAS/FASL system during the acquisition of chemoresistance to oxaliplatin. *PLoS One*, **4**, e4728.
44. Toh,Y. *et al*. (2010) Alcohol drinking, cigarette smoking, and the development of squamous cell carcinoma of the esophagus: molecular mechanisms of carcinogenesis. *Int. J. Clin. Oncol.*, **15**, 135–144.
45. Boccia,S. *et al*. (2005) Sulfotransferase 1A1 polymorphism and gastric cancer risk: a pilot case-control study. *Cancer Lett.*, **229**, 235–243.
46. Guaderman,W. (2002) Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.*, **155**, 478–484.
47. Ritchie,M. *et al*. (2003) Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeity. *Genet. Epidemiol.*, **24**, 150–157.

Received October 8, 2010; revised November 25, 2010; accepted December 4, 2010