Thoracic Cancer ISSN 1759-7706

ORIGINAL ARTICLE

Polymorphisms in DNA repair genes of XRCC1, XPA, XPC, XPD and associations with lung cancer risk in Chinese people

Chaorong Mei[1,2], Mei Hou[3], Shanxian Guo[1], Feng Hua[1], Dejie Zheng[1], Feng Xu[3], Yong Jiang[4], Lu Li[3], Youlin Qiao[4], Yaguang Fan[1] & Qinghua Zhou[1]

1 Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, China
2 Tibet Chengdu branch of West China Hospital, Sichuan University, Changdu, China
3 Cancer Center, West China Hospital, Sichuan University, Chengdu, China
4 Department of Cancer Epidemiology, Cancer Institute, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Keywords
DNA repair gene; haplotype; lung cancer susceptibility; multifactor dimensionality reduction; tagging SNP.

Correspondence
Qinghua Zhou, Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin 300052, China.
Tel: +86 22 60363020
Fax: +86 22 60363013
Email: zhouqh135@163.com

Youlin Qiao, Department of Cancer Epidemiology, Cancer Institute, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China.
Tel: +86 10 87788489
Fax: +86 10 67713648
Email: qiaoy@cicams.ac.cn

Yaguang Fan, Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin 300052, China.
Tel: +86 22 603622047
Fax: +86 22 60363013
Email: fanyaguang75@163.com

Received: 28 August 2013;
accepted 1 September 2013.

doi: 10.1111/1759-7714.12073

Abstract

Background: The carcinogenic chemicals and reactive oxygen species in tobacco can result in DNA damage. DNA repair genes play an important role in maintaining genome integrity. Genetic polymorphisms of DNA repair genes and smoking may contribute to susceptibility of lung cancer.

Methods: In this hospital-based case-control study, we investigated the relationship between 13 tagging single nucleotide polymorphisms (SNPs) in base excision repair pathway and nucleotide excision repair pathway genes, smoking, and lung cancer susceptibility. Thirteen tag SNPs were genotyped in 265 lung cancer patients and 301 healthy controls. Logistic regression and multifactor dimensionality reduction method were applied to explore the association and high-order gene-gene and gene-smoking interaction.

Results: In single tag SNP analysis, XPA rs2808668, XPC rs2733533, and XPD rs1799787 were significantly associated with lung cancer susceptibility. Joint effects analysis of XPA rs2808668, XPC rs2733533 and XPD rs1799787 showed that there was an increased risk of lung cancer with increasing numbers of risk alleles. Haplotype analysis showed that XRCC1 (rs25487, rs1799782, rs3213334) GCC had a positive association with lung cancer. Analysis of gene-gene and gene-smoking interaction by multifactor dimensionality reduction showed that a positive interaction existed between the four genes and smoking. The two-factor model, including XPC rs2755333 and smoking, had the best prediction ability for lung cancer. Compared with the C/C genotype of XPC rs2733533 and no smoking, the combination of genotype A carriers with XPC rs2733533 and heavy smokers (≥30 pack-year) had a 13.32-fold risk of lung cancer.

Conclusion: Our results suggest multiple genetic variants in multiple DNA repair genes may jointly contribute to lung cancer risk through gene-gene and gene-smoking interactions.

Introduction

Lung cancer is responsible for the most cancer-related deaths in the world among both men and women. Eighty-five to 90% of lung cancers are attributable to cigarette smoking.[1–5] Although cigarette smoking is the main cause of lung cancer, not all smokers develop lung cancer.[6,7] Genetic susceptibility to carcinogenesis, which includes epigenetic factors and gene-environment interaction, is also an important determinant of lung cancer risk.[8–11] Tobacco smoke contains many carcinogens and reactive oxygen species that produce DNA adducts, cross-links, DNA damage, and DNA strand breaks

© 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

Table 1 Tag single nucleotide polymorphisms (SNPs) selected from the HapMap database

Repair pathway and genes	Gene location	SNP (rs no.)	Location	Base change	Rare allele frequency in HapMap HCB
Base excision repair					
XRCC1	19q13.2	rs25487	Exon6	A/G	0.274
		rs1799782	Exon10	C/T	0.244
		rs3213334	Intron3	C/T	0.102
		rs3213255	Intron2	C/T	0.144
Nucleotide excision repair					
XPA	9q22.3	rs3176720	Intron5	A/C	0.100
XPC	3p25	rs2808668	Intron2	C/T	0.487
		rs2229090	3'UTR	C/G	0.282
		rs2228001	Exon16	A/C	0.378
		rs2733533	Intron15	A/C	0.089
XPD	19q13.3	rs3729584	Intron10	A/G	0.227
		rs1799787	Intron19	C/T	0.068
		rs238415	Intron17	C/G	0.475
		rs238406	Exon6	G/T	0.407

HCB, Hapmap-Han Chinese in Beijing.

requiring repair through multiple pathways.[12] DNA repair is critical to maintaining the integrity of the genome and repairing the damage from exposure to exogenous environmental xenobiotics, as well as to endogenous damage (e.g. from oxidative metabolism) or spontaneous disintegration of chemical bonds in DNA.[13–15] There are five DNA repair pathways: direct repair,[16] base excision repair (BER),[17] nucleotide excision repair (NER),[18] mismatch repair,[19] and double-strand breaks repair.[20,21] NER is the most versatile in terms of lesion recognition.[22] PAH-induced bulky DNA adducts, such as benzo(a)pyrene diol epoxide-DNA adducts,[23,24] which are the most potent premutagenic adducts, are mainly repaired by NER. A variety of reactive oxygen species, such as hydroxyl radical and hydrogen peroxide, are generated during enzymatic oxidation of PAHs.[25] These reactive oxygen species can lead to DNA damages, which may be quantitatively predominant PAH-induced DNA damage. Oxidative DNA damages are primarily removed via BER.[26,27] BER is the main guardian against damage as a result of cellular metabolism, including reactive oxygen species, methylation, deamination, and hydroxylation.[22,28]

Many studies have suggested that polymorphisms in DNA repair genes are associated with lung cancer.[29–31] However, most analyses focus on single-candidate polymorphisms and the results are not consistent. Lung cancer, as a complex disease, most likely results from genetic variants in multiple genes of different DNA repair pathways. Single-locus effects hardly detect small genetic effects on lung cancer risk. Analysis of multiple genetic variants within a gene, even multiple genes within an entire pathway, should be considered in association studies.[32] The International HapMap Project described the common patterns of variation, including associations between single nucleotide polymorphisms (SNPs), and contained the tag SNPs selected to most efficiently and comprehensively capture this information.[33] In this study, we selected tag SNPs of four DNA repair genes from the HapMap database using the Tagger program,[34] with a threshold of minor allele frequency ≥0.05 and $r^2 \geq 0.8$ in samples of Han Chinese in China. We also examined the heterozygosity of these tag SNPs in Chinese patients and predicted the functional effects of them using the F-SNP database. SNP rs1799782 has been selected as an important polymorphism of XRCC1. Finally, 13 tag SNPs in four DNA repair genes involved in the BER and NER pathway were selected (Table 1). We studied their association with lung cancer risk and estimated haplotypes for SNPs in the four genes. The multifactor dimensionality reduction (MDR) method was used to examine the high-order gene-gene and gene-smoking interactions in these DNA repair genes.

Materials and methods

Study subject

Two hundred and sixty-five patients with lung cancer were consecutively recruited from the Tianjin Medical University General Hospital from July 2008 to July 2009, with no gender, age, histology or cancer stage restrictions. These patients were genetically unrelated ethnic Han Chinese from Northern China. All patients with lung cancer were newly diagnosed and histologically confirmed. None had been treated by chemotherapy or radiotherapy at the inception of the study. Trained abstractors reviewed the medical records. Control subjects were recruited by selecting 301 healthy and genetically unrelated individuals from the same geographic area who had visited the hospital for a routine check-up. The

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

Table 2 Genotyping primers and probes for thirteen tag single nucleotide polymorphisms (SNPs)

Polymorphisms	Primer sequences	Probe sequences	Design strand
Rs25487	F: GCAGGGTTGGCGTGTGA	VIC: CCCTCCCGGAGGTAA	Reverse
CT	R: GAGTGGGTGCTGGACTGT	FAM: CCCTCCCAGAGGTAA	
Rs1799782	F: AGGATGAGAGCGCCAACTC	VIC: CTTGTTGATCCGGCTGAA	Reverse
CT	R: ACTCAGGACCCACGTTGTC	FAM: CTTGTTGATCCAGCTGAA	
Rs3213334	F: CTCCCAAAGTGCTAGGATTACACA	VIC: ACACAGCGGCTCACA	Reverse
CT	R: TGACAAAGTGAGACCTCGTTTCAAA	FAM: ACACAGCAGCTCACA	
Rs3213255	F: TCAGCAAGGGCCTTAAATGCA	VIC: TTGGCTTTTGTGCTCCCAT	Forward
CT	R: CTGGCAAATGTTCTCATGGCATATT	FAM: TTGGCTTTTGTGTTCCCAT	
Rs3176720	F: GTCTTTCACGACATTGACATTTTGCA	VIC: CAGGCCAGCTGCTG	Forward
AC	R: CTGAATGGAGGGACACACTGAA	FAM: AGGCCCGCTGCTG	
Rs2808668	F: CCTCCATCTTCATAGCCAGCAATG	VIC: TGATGCCGTGTGAGAAG	Reverse
CT	R: GTCAGAGGGACATGTGATTATGGAA	FAM: TGATGCCATGTGAGAAG	
Rs2229090	F: GCCCAGCCCCTGGTG	VIC: AGCAGAGAAGCCCCCAC	Reverse
CG	R: GCTGCCTCAGTTTGCCTTCT	FAM: AGCAGAGAACCCCCCAC	
Rs2228001	F: CAGCAGCTTCCCACCTGTT	VIC: CCCATTTGAGAAGCTGT	Forward
AC	R: GTGGGTGCCCCTCTAGTG	FAM: CCCATTTGAGCAGCTGT	
Rs2733533	F: ACAGAAGACTGAGGTGTCCTAACA	VIC: TCTGCCCCATCCTCAA	Reverse
AC	R: GAAAGGCCTGGCCCAGAT	FAM: TGCCCCAGCCTCAA	
Rs3729584	F: CTCTGGGCCCCCTAGGA	VIC: CAGGTTCAGCTACCCTG	Reverse
AG	R: AGCGCAGCCCTGCA	FAM: AGGTTCAGCCACCCTG	
Rs1799787	F: CCCCAACTCAGACACAGCAT	VIC: CCTCACGCGACCCAG	Reverse
CT	R: CTGGTGGGACAGGGACAG	FAM: CTCACGCAACCCAG	
Rs238415	F: CGCGGCGGGAAAGG	VIC: AAGGCACCTGGGCTGT	Reverse
CG	R: GTAGGCAAAGGTGTCTTAAGTAGGA	FAM: AAGGCACCTGCGCTGT	
Rs238406	F: AGCCCTGCCCTCCAGT	VIC: ACCTCATAGAAGCGGCAGT	Forward
GT	R: GCGCAGTACCAGCATGACA	FAM: AACCTCATAGAATCGGCAGT	

control subjects were frequency matched to the case subjects by age (± 5 years) and gender. Once written informed consent was provided, the patients' demographic information and environmental tobacco smoke exposure histories were collected. Approximate 5 mL of venous blood samples were collected from each participant for DNA analysis. In our study, environmental tobacco smoke exposure history was calculated by multiplying the number of packs of cigarettes smoked per day by the number of years that the person had smoked. Non-smokers were defined as persons who had never smoked. Those smoking <30 pack-years were defined as light smokers and those smoking ≥30 pack-years were defined as heavy smokers. The Institutional Review Board of the Tianjin Medical University approved this study.

Genotyping

Genomic DNA was extracted from whole blood with an AxyPrep-96 Blood Genomic DNA Kit. The XRCC1 rs25487, rs1799782, rs3213334, rs3213255, XPA rs3176720, rs2808668, XPC rs2229090, rs2228001, rs2733533, rs3729584, XPD rs1799787, rs238415, and rs238416 polymorphisms were genotyped using TaqMan allelic discrimination assays (Applied Biosystems 7500 Fast Real-Time PCR System, Carlsbad CA, USA. The primers and probes were provided by Applied Biosystems (as shown in Table 2). The polymerase chain reaction (PCR) amplification was performed with 20ng DNA, 5ul Allelic Discrimination PCR Reaction 40X mix (Applied Biosystems), 12.5ul Taqman Universal PCR Master Mix (2X) (Applied Biosystems) and 0.625 uL of the assay mix (primers and probes were included) using 96-well plates on an ABI Prism 7500 Sequence Detection System (Applied Biosystems). The reaction conditions were: 95°C for 10 minutes, followed by 40 cycles of 15 seconds at 92°C, and one minute at 60°C. Genotype was analysed by ABI Prism 7500 SDS software (Applied Biosystems, Carlsbad CA, USA). About 5% of the samples were re-tested for quality control and the concordance was 100%.

Statistical methods

We performed a χ^2 test to discern the differences between the lung cancer and control groups in distribution of gender, age, and smoking exposure (pack-years). To find the mean age, we also performed the t-test. The χ^2 test was also used to assess the differences between the observed and expected genotype frequencies in the control group for Hardy-Weinberg analysis. The linkage disequilibrium analysis was performed using the LDA program.[35] The haplotypes analysis was performed by PHASE 2.0 program,[36] by which haplotypes could be

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

Table 3 Demographic characteristics in lung cancer and control group

Characteristic	Lung cancer *n* (%)	Controls *n* (%)	*P*
Mean age (± SD)	60.51 (9.57)	60.30 (10.10)	0.421
Age			
<60y	113 (45.0)	143 (47.5)	0.559
≥60y	138 (55.0)	158 (52.5)	
Gender			
Male	181 (72.1)	214 (71.1)	0.792
Female	70 (27.9)	87 (28.9)	
Smoking status			
No smoking	89 (35.5)	169 (56.1)	<0.0001
Smoking <30 pack-year	36 (14.3)	91 (30.2)	
Smoking ≥30 pack-year	126 (50.2)	41 (13.6)	
Smoking	162 (64.5)	132 (43.9)	

reconstructed and haplotype frequencies could be inferred from genotype data on the basis of the Bayesian algorithm. Unconditional logistic regression was performed to estimate the odd ratio (OR) and 95% confidence interval (CI) with adjustments made for age, gender, and smoking exposure accordingly. All analyses were performed using the SPSS 11.5 software package (SPSS Company, Chicago, IL). All tests were two-sided and the criteria of statistical significance was $P < 0.05$. The MDR[37] (MDR version 2.0) was used to estimate the combinations of gene-gene and gene-smoking interactions.

Results

Characteristics of the subjects

We recruited 265 lung cancer cases and 301 healthy controls for this study. There were 14 patients whose smoking history was unavailable, and, therefore, their data was excluded from our study. There were no significant differences between the case and control groups in the mean age (60.51 ± 9.57 years vs. 60.30 ± 10.10 years; $P = 0.421$). There were also no significant differences in age and gender distributions between the lung cancer and control groups. However, a significant difference in smoking status was observed between the lung cancer and control groups. The percentage of smokers in the lung cancer group was significantly higher than that of the control group (64.5% vs. 43.9%; $P < 0.001$). The details of demographic characteristics and smoking status are presented in Table 3.

According to World Health Organization (WHO) classifications, the histological types of the 251 lung cancer patients were as follows: squamous cell carcinoma ($n = 120$, 47.8%), adenocarcinoma ($n = 87$, 34.7%), and other non-small cell lung cancer (NSCLC) ($n = 44$, 17.5%).

Genotype of DNA repair genes and associations with lung cancer risk

The genotype distributions of all SNPs in the control group were within the Hardy-Weinberg equilibrium, except for rs3213255 of XRCC1 and rs238406 of XPD ($P < 0.05$), therefore, we excluded these two SNPs from the next analysis. In the single SNP analysis, the genotype frequencies of XPA rs2808668 were significantly different between the lung cancer and control groups ($\chi2 = 9.846$, $P = 0.007$, data not shown); genotypes of XPC rs2733533, C/C, A/C, and A/A in the lung cancer and control groups were 85.7%, 14.3%, 0%, and 94.0%, 6.0%, 0%, respectively. The A/A genotype was not found in our study, so we combined A/C and A/A genotypes ("A carriers," that is any A) and found that a significant difference of genotype frequencies existed between the lung cancer and control groups ($\chi2 = 10.845$, $P = 0.001$, data not shown). However, no significant genotype frequency differences of other SNPs were found between the lung cancer and control groups.

After age, gender, and smoking status (pack-year) adjustment, compared with the common homozygous genotype C/C, for XPA rs2808668 polymorphism, individuals with the heterozygous (C/T) genotype had a significantly increased risk of lung cancer (adjusted OR, 1.77; 95% CI: 1.12–2.80). The presence of any T meant a 66% increase in lung cancer risk (adjusted OR, 1.66; 95% CI: 1.08–2.55). For XPC rs2733533 polymorphism, the presence of any A was associated with a 1.48-fold increased risk of lung cancer, compared with the C/C genotype, (adjusted OR, 2.48; 95% CI: 1.29–4.76). For XPD rs1799787, individuals with the heterozygous (C/T) genotype showed a significantly increased risk (adjusted OR, 1.89; 95% CI: 1.13–3.15) and the presence of any T showed a borderline association with lung cancer risk compared with the C/C genotype (adjusted OR, 1.63; 95% CI: 0.99–2.68). Individuals with the T/T genotype showed a borderline decreased risk of lung cancer (adjusted OR, 0.09; 95% CI: 0.01–1.00). The associations between SNPs of DNA repair genes and lung cancer risk is shown in Table 4. We estimated the associations of these SNPs with lung cancer risk further, stratified by age, gender, smoking status, and histological type. The common homozygous genotype was used as the reference group. Individuals with any T allele of XPA rs2808668 polymorphism showed a more pronounced increase in the risk of developing lung cancer, including patients aged <60 years (adjusted OR, 2.23; 95% CI: 1.21–4.12; data not shown), non-smokers (adjusted OR, 2.17; 95% CI: 1.16–4.06; data not shown), and in squamous cancer (adjusted OR, 2.09; 95% CI: 1.09–4.01; data not shown). Individuals with any A allele of XPC rs2733533 polymorphism showed a more pronounced increase of risk in patients aged <60 years (adjusted OR, 6.44; 95% CI: 2.13–19.52; data not shown), males (adjusted OR, 2.72; 95% CI: 1.24–5.97; data

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

Table 4 Genotype of DNA repair genes and associations with lung cancer risk

SNPs	Lung cancer *n* (%)	Controls *n* (%)	OR†	*P* value
XRCC1				
rs25487				
GG	142 (56.6)	145 (48.2)	Reference	
AG	95 (37.8)	126 (41.9)	0.74 (0.50–1.09)	0.131
AA	14 (5.6)	30 (10.0)	0.52 (0.25–1.07)	0.077
Any A	109 (43.4)	156 (51.8)	0.70 (0.48–1.01)	0.058
A MAF	0.25	0.31		
rs1799782				
CC	138 (55.0)	155 (51.5)	Reference	
CT	90 (35.9)	119 (39.5)	0.82 (0.55–1.22)	0.33
TT	23 (9.2)	27 (9.0)	0.95 (0.49–1.83)	0.87
Any T	113 (45.0)	146 (48.5)	0.85 (0.58–1.22)	0.37
T MAF	0.27	0.29		
.rs3213334				
CC	206 (82.1)	243 (80.7)	Reference	
CT	44 (17.5)	55 (18.3)	0.95 (0.59–1.54)	0.85
TT	1 (0.4)	3 (1.0)	0.89 (0.09–8.68)	0.92
Any T	45 (17.9)	58 (19.3)	0.95 (0.59–1.53)	0.84
T MAF	0.09	0.10		
XPA				
rs3176720				
AA	202 (80.5)	250 (83.1)	Reference	
AC	48 (19.1)	48 (15.9)	1.38 (0.85–2.23)	0.196
CC	1 (0.4)	3 (1.0)	0.36 (0.03–4.02)	0.404
Any C	49 (19.5)	51 (16.9)	1.31 (0.81–2.10)	0.270
C MAF	0.10	0.09		
.rs2808668				
CC	49 (19.5)	93 (30.9)	Reference	
CT	142 (56.6)	139 (46.2)	1.77 (1.12–2.80)	0.014
TT	60 (23.9)	69 (22.9)	1.43 (0.84–2.43)	0.194
Any T	202 (80.5)	208 (69.1)	1.66 (1.08–2.55)	0.022
T MAF	0.52	0.46		
XPC				
.rs2229090				
CC	116 (46.2)	152 (50.5)	Reference	
CG	109 (43.4)	114 (37.9)	1.33 (0.90–1.97)	0.157
GG	26 (10.4)	35 (11.6)	1.06 (0.57–1.99)	0.848
Any G	135 (53.8)	149 (49.5)	1.27 (0.88–1.84)	0.209
G MAF	0.32	0.31		
.rs2228001				
AA	96 (38.2)	113 (37.5)	Reference	
AC	116 (46.2)	137 (45.5)	1.12 (0.74–1.68)	0.595
CC	39 (15.5)	51 (16.9)	0.88 (0.51–1.53)	0.653
Any C	155 (61.8)	188 (62.5)	1.05 (0.72–1.54)	0.805
C MAF	0.39	0.40		
. rs2733533				
CC	215 (85.7)	283 (94.0)	Reference	
AC	36 (14.3)	18 (6.0)	2.48 (1.29–4.76)	0.006
AA	0 (0.0)	0 (0.0)	–	–
Any A	36 (14.3)	18 (6.0)	2.48 (1.29–4.76)	0.006
AA MAF	0.07	0.03		
.rs3729584				
GG	135 (53.8)	147 (48.8)	Reference	
AG	96 (38.2)	126 (41.9)	0.72 (0.48–1.06)	0.099
AA	20 (8.0)	28 (9.3)	0.81 (0.41–1.60)	0.552
Any A	116 (46.2)	154 (51.2)	0.73 (0.51–1.07)	0.104
AA MAF	0.27	0.30		
XPD				
rs1799787				
CC	203 (80.9)	260 (86.4)	Reference	
CT	47 (18.7)	37 (12.3)	1.89 (1.13–3.15)	0.015
TT	1 (0.4)	4 (1.3)	0.09 (0.009–1.00)	0.050
Any T	48 (19.1)	41 (13.6)	1.63 (0.99–2.68)	0.055
T MAF	0.10	0.07		
.rs238415				
CC	76 (30.3)	90 (29.9)	Reference	
CG	127 (50.6)	138 (45.8)	1.07 (0.69–1.64)	0.772
GG	48 (19.1)	73 (24.3)	0.70 (0.42–1.19)	0.192
Any G	175 (69.7)	211 (70.1)	0.94 (0.63–1.41)	0.761
G MAF	0.44	0.47		

†OR: adjusted for age, gender and smoking status (pack-year) using unconditional logistic regression.
SNPs, single nucleotide polymorphisms.

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

Table 5 Joint-effects among XPA rs2808668, XPC rs2733533, and XPD rs1799787

XPA rs2808668	XPC rs2733533	XPD rs1799787	OR (95% CI)†	*P*‡
CC	CC	CC	Reference	
CC	CC	Any T	2.51 (0.65–9.67)	0.18
CC	Any A	CC	3.50 (0.73–16.87)	0.12
CC	Any A	Any T	4.62 (0.21–99.98)	0.33
Any T	CC	CC	1.70 (1.04–2.80)	0.036
Any T	CC	Any T	2.35 (1.13–4.91)	0.022
Any T	Any A	CC	2.72 (0.99–7.49)	0.052
Any T	Any A	Any T	10.80 (1.83–63.70)	0.009

†Adjusted by age, gender, smoking status. ‡After Bonferroni adjustment for multiple comparison, significance level for each individual test is 0.0071. CI, confidence interval; OR, odds ratio.

not shown), non-smokers (adjusted OR, 5.80; 95% CI: 1.72–19.55, $P = 0.005$; data not shown), and in squamous cancer (adjusted OR, 2.61; 95% CI: 1.14–5.97. $P = 0.023$; data not shown). Individuals with the heterozygous (C/T) genotype of XPD rs1799787 showed a significant increase of lung cancer risk in males (OR, 1.89; 95% CI: 1.01–3.55; data not shown), and light smokers (adjusted OR, 2.87; 95% CI: 1.04–7.97; data not shown). Moreover, the common homozygous genotype was used as the reference group for the following SNPs. Individuals with any A allele of XRCC1 rs25487 polymorphism showed a significant decrease of risk in patients aged ≥60 years (adjusted OR, 0.53; 95% CI: 0.31–0.92; data not shown). Women with any G allele of XPC rs2229090 polymorphism showed a significantly increased risk (adjusted OR, 2.53; 95% CI: 1.28–5.00; data not shown). Further, women with any A allele of XPC rs37298584 polymorphism showed a significantly decreased risk of lung cancer (adjusted OR, 0.35; 95%CI: 0.18–0.71; data not shown). Individuals with the G/G genotype of XPD rs238415 showed a significantly decreased risk in squamous cancer (adjusted OR, 0.46; 95% CI: 0.22–0.96; data not shown).

Because significant associations of XPA rs2808668, XPC rs2733533, and XPD rs1799787 polymorphisms with lung cancer risk were found in single SNP analysis, we performed multivariable logistic regression analysis to evaluate the joint effects of these SNPs. The results are shown in Table 5. Compared with the low-risk genotypes of the three polymorphisms, individuals with one or more high-risk genotypes showed a positive association with lung cancer risk. The combined presence of high-risk genotype in all three polymorphisms showed a 9.8-fold increase of lung cancer risk (adjusted OR, 10.80; 95% CI: 1.83–63.70).

Haplotype of DNA repair genes and associations with lung cancer risk

We performed linkage disequilibrium (LD) chi-square tests for LD analysis on these SNPs of the four DNA repair genes, respectively, by LDA. XRCC1 rs25487, rs1799782, and rs3213334 polymorphisms were in LD (D' = 0.96, 1.00, 0.83, respectively; all values of $P < 0.001$). XPA rs3176720, and rs2808668 were in LD (D' = 1.00, $P < 0.001$). XPC rs2229090, rs2228001, rs2733533, and rs3729584 were in LD (D' = 0.63, 1.00, 0.96, 0.58, 1.00, 1.00; all values of $P < 0.05$). XPD rs1799787, and rs238415 were in LD (D' = 1.0, $P < 0.001$). We subsequently constructed haplotypes using PHASE 2.0 (Stephens and Donnelly, 2003) and evaluated their association with lung cancer risk.

The distribution of haplotypes in the lung cancer and control groups and their association with lung cancer are shown in Table 6. (Haplotypes with frequencies of less than 0.10 were categorized into a mixed group named "others.") The most common haplotypes of XRCC1 (rs25487, rs1799782, rs321333), XPA (rs3176720, rs2808668), XPC (rs2229090, rs2228001, rs2733533, rs3729584), and XPD (rs1799787, rs238415) were ACC, AC, CCCG, and CG, respectively.

Using the most common haplotype of the four genes as references, haplotype GCC of XRCC1 was remarkably associated with an increased risk of lung cancer (adjusted OR, 1.63; 95% CI: 1.17–2.28). Moreover, we found that an increased risk of haplotype GCC of XRCC1 was more pronounced in squamous cancer (adjusted OR, 1.89; 95% CI: 1.19–2.99; data not shown).

Gene–gene and gene–environment interactions on lung cancer risk.

We performed the non-parametric MDR approach for the analysis of gene-gene and gene-smoking interactions on lung cancer risk with the 11 SNPs of the four DNA repair genes and smoking status in our study. The models inferred by the method are shown in Table 7.

It is well known that smoking is the major risk factor of lung cancer. MDR results showed that smoking was included in all of the best examples of one or more factor models. The best one-factor model for lung cancer risk predication only included smoking, with the highest cross-validation consis-

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

Table 6 Associations between frequencies of DNA repair gene haplotypes and lung cancer risk

Haplotypes	Lung cancer *n* (%)	Controls *n* (%)	OR†† (95% CI)	*P* value
XRCC1†				
ACC	122 (24.3)	185 (30.7)	Reference	
GCC	198 (39.4)	184 (30.6)	1.63 (1.17–2.28)	0.004
GTC	135 (26.9)	171 (28.4)	1.18 (0.83–1.68)	0.357
Others	47 (9.4)	62 (10.3)	1.23 (0.76–2.00)	0.397
XPA‡				
AC	240 (47.8)	325 (54.0)	Reference	
AT	212 (42.2)	223 (37.0)	0.77 (0.49–1.21)	0.258
CT	50 (10.0)	54 (9.0)	0.92 (0.58–1.45)	0.709
XPC§				
CCCG	171 (34.1)	211 (35.0)	Reference	
CACA	135 (26.9)	181 (30.1)	1.12 (0.81–1.55)	0.483
GACG	134 (26.7)	156 (25.9)	1.32 (0.83–2.08)	0.238
Others	62 (12.4)	54 (9.0)	1.21 (0.86–1.71)	0.284
XPD¶				
CG	223 (44.4)	284 (47.2)	Reference	
CC	230 (45.8)	273 (45.3)	0.89 (0.68–1.16)	0.374
TC	49 (9.8)	45 (7.5)	1.24 (0.77–1.99)	0.379

†XRCC1: rs25487- rs1799782 -rs3213334. ‡XPA: rs3176720-rs2808668. §XPC: rs2229090-rs2228001-rs2733533-rs3729584. ¶XPD: rs1799787-rs238415. ††OR: adjusted for age, gender, smoking status.
CI, confidence interval; OR, odds ratio.

tency (CVC) of 10/10 and testing accuracy of 68.29%. In two-factor models, XPC rs2733533 and smoking were the best two-factor predictors of lung cancer risk, with the highest CVC of 10/10 and testing accuracy of 69.32%, which was higher than that of the one-factor model, and, thus, showed improved capability of prediction than smoking alone. In three-factor models, the combination of XPA rs3176720, XPC rs2733533, and smoking status was the best model with a CVC of 8/10 and the highest testing accuracy of 69.56%. Compared with the best two-factor model, the best three-factor model had slightly improved testing accuracy, but a decrease in CVC. When four factors were considered in the model, XPA rs3176720, XPC rs2228001, XPD rs238415, and smoking status was the strongest model with cross-validation consistency of 6/10 and testing accuracy of 66.08%. Compared with the best two- or three-factor models, the best four-factor model had a decrease in both the testing accuracy and the CVC.

The best two-factor model, consisting of XPC rs2733533 and smoking, was thought to be the fitted model (See Fig 1).

We further evaluated the joint effects of XPC rs2733533 and smoking status by logistic regression. The C/C genotype of XPC rs2733533 and no smoking were used as references; the combination of genotype A carriers of XPC rs2733533 and heavy smokers (≥30P.Y) showed the most maximum positive association with lung cancer risk (adjusted OR, 14.32; 95% CI: 4.46–45.93). This result was consistent with the MDR result (See Table 8).

We further used interaction dendrograms (Fig 2) with MDR to demonstrate the visualized interaction of these SNPs and smoking. The dendrogram placed strongly interacting attributes close together at the leaves of the tree.[38,39] The colors of the branch indicated the degree of interaction. The degrees of interaction from strong to weak were represented by red, orange, gold, green, and blue. Red represented the highest degree of synergy, and blue represented redundancy or no interaction. The hierarchical cluster analysis placed XRCC1 rs1799782, XPC rs2228001, XPD rs238415, and smoking on the same branch, but XPA rs3176720 and XPC rs2733533 on another branch. The distribution of the attributes in the

Table 7 Multifactor dimensionality reduction (MDR) analysis for the lung cancer risk predication (*n* = 552)

Best model	Cross-validation consistency	Testing accuracy	Permutation test *P* value†
One factor: smoking status	10/10	0.6829	<0.001
Two factors: rs2733533; smoking status	10/10	0.6932	<0.001
Three factors: rs3176720; rs2733533; smoking status	8/10	0.6956	<0.001
Four factors: rs1799782; rs2228001; rs238415; smoking status	6/10	0.6608	<0.001

†1000-fold permutation test.

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

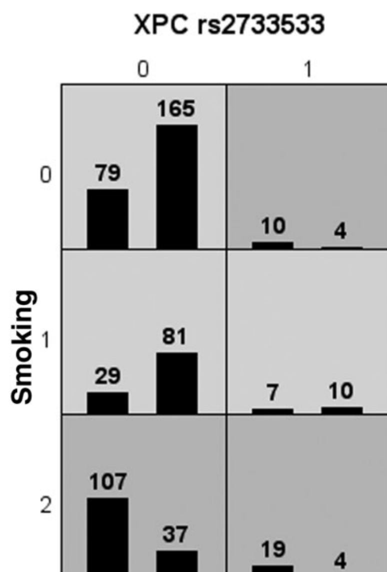

Figure 1 XPC rs2733533 and smoking combined are associated with high and low risks of lung cancer multifactor dimensionality reduction (MDR) analysis with the highest testing accuracy. For each multifactor cell, the number of lung cancer cases is displayed in the left bar and the number of controls is displayed in the right bar. Cells of dark gray indicated high risk combinations and cells of light gray indicated low risk combinations. XPC rs2733533: 0:CC genotype, 1:AC genotype. Smoking: 0: no smoking, 1: light smoking (<30 pack-year), 2: heavy smoking (≥30 pack-year).

dendrogram indicated that the four-factor model consisting of XRCC1 rs1799782, XPCrs 2228001, XPDrs 238415, and smoking had a synergistic effect on increasing the lung cancer risk. No synergistic effect was observed between XPA rs3176720 and XPC rs2733533. The interaction dendrogram also showed the structure of the best model in two, three, and four factors.

Discussion

In this study, 13 tagging SNPs were genotyped to capture a large proportion of common genetic variation in four DNA repair genes belonging to two DNA repair pathways. We investigated the association between these SNPs, their haplotypes, and lung cancer risk. We performed a non-parametric MDR to evaluate potential gene-gene and gene-smoking interactions.

It is a well-established fact that smoking is the main risk factor of lung cancer.[40–42] In our study, smoking was found to increase lung cancer risk by logistic regression analysis, which was also verified by MDR. Single polymorphism analysis showed that XPA rs2808668, XPC rs2733533, and XPD rs1799787 had a significant association with lung cancer risk. In young individuals, non-smokers with squamous cell carcinoma, XPA rs2808668, and XPC rs2733533 showed a more pronounced association with lung cancer risk. XPC rs2733533 also showed a more pronounced association with lung cancer risk in men. In males and light smokers, XPD rs1799787 showed a close association with lung cancer risk. The result of the joint effects of these three SNPs showed an increase risk of lung cancer with increasing numbers of risk alleles. The combined presence of the high-risk genotype in all three polymorphisms showed a 9.8-fold increase in lung cancer risk. Moreover, after stratification by age, gender, and smoking status, XRCC1 rs25487 had a close association with lung cancer risk in elderly individuals, and XPC rs2229090 and XPC rs37298584 had close associations with lung cancer risk in women. XPD rs238415 had a close association with lung cancer risk in squamous cell carcinoma of the lung. When we evaluated the haplotypes derived from the four DNA repair genes, we found that the haplotype of XRCC1 (rs25487, rs1799782, rs3213334) GCC had a positive association with lung cancer, and this association was more pronounced in squamous cell carcinoma.

Further analysis of gene-gene and gene-smoking interaction by MDR showed that a significant interaction existed between the four genes and smoking. The interaction dendrograms showed that there was no significant interaction between XPA rs3176720 and XPC rs2733533. However, both the SNPs had positive interactions with XRCC1 rs1799782, XPC rs2228001, XPD rs238415 and smoking. The strongest synergistic interaction was found between XRCC1 rs1799782 and XPC rs2228001. Consistent with the significant positive association with lung cancer risk in XPC rs2755333, XPC

Table 8 Joint effects between XPC rs2733533 and smoking status for lung cancer risk

XPC rs2733533	Smoking status	Adjusted OR (95%CI)†	*P*‡
CC	No	Reference	
Any A	No	5.8 (1.72–19.55)	0.005
CC	<30 pack-year	1.01 (0.57–1.77)	0.982
Any A	<30 pack-year	1.59 (0.57–4.47)	0.378
CC	≥30 pack-year	7.63 (4.50–12.95)	<0.001
Any A	≥30 pack-year	14.32 (4.46–45.93)	<0.001

†Adjusted by age, gender, smoking status. ‡After Bonferroni adjustment for multiple comparison, significance level for each individual test is 0.01.
CI, confidence interval; OR, odds ratio.

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

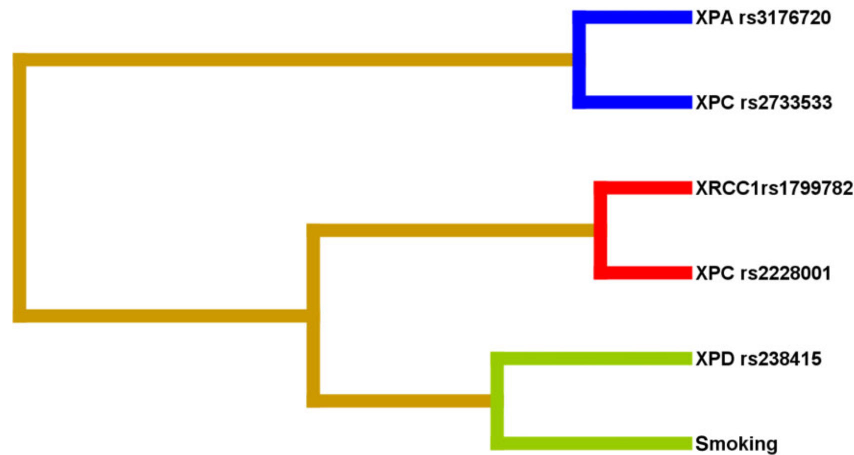

Figure 2 Interaction dendrogram gained from the multifactor dimensionality reduction (MDR) for gene-gene and gene-smoking interactions on lung cancer risk. XRCC1 rs1799782 and XPC rs2228001 had the strongest synergistic interaction, whereas the interaction of XPA rs3176720 and XPC rs2733533 were redundant.

rs2755333 and smoking represented the best two-factor model by MDR. Otherwise, logistic regression analysis further confirmed that the joint-effect of XPC rs2755333 and smoking could remarkably increase lung cancer risk. Compared with the C/C genotype of XPC rs2733533 and no smoking, the combination of genotype A carriers of XPC rs2733533 and heavy smokers (≥30 pack-years) could increase lung cancer risk by 13.32-fold.

In our study, XPC rs2755333 had the strongest interaction with all of the SNPs and lung cancer risk. The results were verified by traditional parametric statistical methods logistic regression and by a non-parametric MDR approach. XPC rs2755333 was located in the intron region and captured 12 SNPs by tagger program. Intronic SNPs are common and may have an indirect functional role, such as affecting RNA splicing, thus influencing the transcription of the gene.[43,44] To date, there has not been any functional research on XPCrs275533. However, using the F-SNP database,[45] which provides information about the functional effects of SNPs obtained from 16 bioinformatics tools and databases, we found this tag SNP might have some function on transcriptional regulation. Another possible reason for the observed interaction is that XPC rs2755333 could be in tight linkage disequilibrium with other ungenotyped SNPs, which have some function to contribute to lung cancer.

Our study has several specialties. Firstly, the DNA samples were available from the Han Chinese residence of the same district in Northern China, therefore, the lung cancer cases and controls used have ethnic and residence homogeneity. Secondly, lung cancer is a complex multifactorial disease, which occurs by multiple gene-gene and multiple gene-environment interplay, thus, the effect of single SNP and a single gene does not adequately represent lung cancer risk.[46,47] Therefore, our study examined multiple SNPs in different DNA repair pathways. The HapMap database can provide wide coverage of common variations. Tagging approaches may substantially improve the cost-effectiveness of association studies by delivering greater power and better genotyping efficiency through the selection of tag SNPs and definition of statistical tests, based on the empirical LD patterns in HapMap.[48,49] We selected 13 tag SNPs from four DNA repair genes in two DNA repair pathways from the HapMap database, instead of commonly selecting single potentially functional SNPs. The tag SNPs were obtained from HapMap Han Chinese in Beijing. Thirdly, we used two methods; a traditional parametric statistical method logistic regression, and a non-parametric MDR approach, to evaluate the relationship between the SNPs and lung cancer risk. As a traditional parametric statistical method, logistic regression is useful for covariate adjustment and to describe relative risks for disease, in association with various combinations of genetic and environmental factors.[50,51] However, it hardly detects complex multifactorial disease interaction because of a combination of factors with no observations, or has limited power to detect clinically relevant interactions because of a low number of events per parameter in the model. The MDR method was proposed as a possible solution in such settings.[52] With MDR, multilocus genotypes are pooled into high-risk and low-risk groups, effectively reducing the genotype predictors from n dimensions to one dimension. The new, one-dimensional multilocus-genotype variable is evaluated for its ability to classify and predict disease status through cross-validation and permutation testing.[53] The use of MDR to identify potential gene combinations or interactions for more efficient testing using traditional logistic regression techniques seems appropriate.[50,51,54]

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

Conclusion

In conclusion, our study suggests that multiply SNPs, from different DNA repair genes in different pathways, and smoking may have a joint contribution to lung cancer genetic susceptibility. However, large sample size studies are warranted for further study.

Acknowledgments

This study was partly supported by the grants from the National Eleveth-Five-Year Key Task Project of China (No.2006BAI02A01, to Qinghua Zhou), the National High Technology Research and Development Program of China (863) (No. 2006AA02A401, 2012AA02A201 and 2012AA02A502) and the China-Sweden International Scientific and Technological Cooperative Project (No.09ZCZDSF04100).

Disclosure

No authors report any conflict of interest.

References

1 Lin IH, Ho ML, Chen HY *et al*. Smoking, green tea consumption, genetic polymorphisms in the insulin-like growth factors and lung cancer risk. *PLoS ONE* 2012; **7**: e30951.
2 Schwartz AG, Ruckdeschel JC. Familial lung cancer: genetic susceptibility and relationship to chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006; **173**: 16–22.
3 Wynder EL, Muscat JE. The changing epidemiology of smoking and lung cancer histology. *Environ Health Perspect* 1995; **103** (Suppl. 8): 143–8.
4 Li D, Zhou Q, Gu Z *et al*. Relationship between genetic polymorphism of CYP1A1 and lung cancer genetic susceptibility. *Chin J Cancer Prev Treat* 2006; **13**: 1765–8. (In Chinese.)
5 Li C, Zhou Q, Wang Y, Chen X, Yang X, Zhu D. [Relationship between LAPTM4B gene polymorphism and susceptibility of lung cancer.] *Chin J Lung Cancer* 2006; **9**: 109–12. (In Chinese.)
6 Hirvonen A, Husgafvel-Pursiainen K, Anttila S, Vainio H. The GSTM1 null genotype as a potential risk modifier for squamous cell carcinoma of the lung. *Carcinogenesis* 1993; **14**: 1479–81.
7 Bryant A, Cerfolio RJ. Differences in epidemiology, histology, and survival between cigarette smokers and never-smokers who develop non-small cell lung cancer. *Chest* 2007; **132**: 185–92.
8 Shen J. Evaluation of environmental and personal susceptibility characteristics that modify genetic risks. *Methods Mol Biol* 2009; **471**: 163–77.
9 Wei Q, Cheng L, Hong WK, Spitz MR. Reduced DNA repair capacity in lung cancer patients. *Cancer Res* 1996; **56**: 4103–7.
10 Taioli E. Gene-environment interaction in tobacco-related cancers. *Carcinogenesis* 2008; **29**: 1467–74.
11 Kiyohara C, Yoshimasu K, Shirakawa T, Hopkin JM. Genetic polymorphisms and environmental risk of lung cancer: a review. *Rev Environ Health* 2004; **19**: 15–38.
12 Schwartz AG, Prysak GM, Bock CH, Cote ML. The molecular epidemiology of lung cancer. *Carcinogenesis* 2007; **28**: 507–18.
13 Spitz MR, Wei Q, Dong Q, Amos CI, Wu X. Genetic susceptibility to lung cancer: the role of DNA damage and repair. *Cancer Epidemiol Biomarkers Prev* 2003; **12**: 689–98.
14 Harms C, Salama SA, Sierra-Torres CH, Cajas-Salazar N, Au WW. Polymorphisms in DNA repair genes, chromosome aberrations, and lung cancer. *Environ Mol Mutagen* 2004; **44**: 74–82.
15 Sakoda LC, Loomis MM, Doherty JA *et al*. Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers. *Int J Mol Epidemiol Genet* 2012; **3**: 1–17.
16 Eker AP, Quayle C, Chaves I, van der Horst GT. DNA repair in mammalian cells: direct DNA damage reversal: elegant solutions for nasty problems. *Cell Mol Life Sci* 2009; **66**: 968–80.
17 Robertson AB, Klungland A, Rognes T, Leiros I. DNA repair in mammalian cells: base excision repair: the long and short of it. *Cell Mol Life Sci* 2009; **66**: 981–93.
18 Diderich K, Alanazi M, Hoeijmakers JH. Premature aging and cancer in nucleotide excision repair-disorders. *DNA Repair (Amst)* 2011; **10**: 772–80.
19 Hsieh P, Yamane K. DNA mismatch repair: molecular mechanism, cancer, and ageing. *Mech Ageing Dev* 2008; **129**: 391–407.
20 Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* 2001; **27**: 247–54.
21 Scott RJ. DNA double strand break repair and its association with inherited predispositions to breast cancer. *Hered Cancer Clin Pract* 2004; **2**: 37–43.
22 Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature* 2001; **411**: 366–74.
23 Gelboin HV. Benzo[alpha]pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes. *Physiol Rev* 1980; **60**: 1107–66.
24 Gelhaus SL, Harvey RG, Penning TM, Blair IA. Regulation of benzo[a]pyrene-mediated DNA- and glutathione-adduct formation by 2,3,7,8-tetrachlorodibenzo-p-dioxin in human lung cells. *Chem Res Toxicol* 2011; **24**: 89–98.
25 Hammel KE, Kalyanaraman B, Kirk TK. Oxidation of polycyclic aromatic hydrocarbons and dibenzo[p]-dioxins by Phanerochaete chrysosporium ligninase. *J Biol Chem* 1986; **261**: 16948–52.
26 Park JY, Lee SY, Jeon HS *et al*. Polymorphism of the DNA repair gene XRCC1 and risk of primary lung cancer. *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 23–7.

 © 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd

27 Hazra TK, Das A, Das S, Choudhury S, Kow YW, Roy R Oxidative DNA damage repair in mammalian cells: a new perspective. *DNA Repair (Amst)* 2007; **6**: 470–80.
28 Madhusudan S, Smart F, Shrimpton P *et al*. Isolation of a small molecule inhibitor of DNA base excision repair. *Nucleic Acids Res* 2005; **33**: 4711–24.
29 Kiyohara C, Yoshimasu K. Genetic polymorphisms in the nucleotide excision repair pathway and lung cancer risk: a meta-analysis. *Int J Med Sci* 2007; **4**: 59–71.
30 Zheng H, Wang Z, Shi X, Wang Z. XRCC1 polymorphisms and lung cancer risk in Chinese populations: a meta-analysis. *Lung Cancer* 2009; **65**: 268–73.
31 Qian B, Zhang H, Zhang L, Zhou X, Yu H, Chen K. Association of genetic polymorphisms in DNA repair pathway genes with non-small cell lung cancer risk. *Lung Cancer* 2011; **73**: 138–46.
32 Michiels S, Danoy P, Dessen P *et al*. Polymorphism discovery in 62 DNA repair genes and haplotype associations with risks for lung and head and neck cancers. *Carcinogenesis* 2007; **28**: 1731–9.
33 International HapMap Consortium. The International HapMap Project. *Nature* 2003; **426**: 789–96.
34 International HapMap Project (public releases up to Jan 2008). Available from URL: http://www.hapmap.org.
35 Ding K, Zhou K, He F, Shen Y. LDA – a java-based linkage disequilibrium analyzer. *Bioinformatics* 2003; **19**: 2147–8.
36 Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978–89.
37 Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* 2003; **19**: 376–82.
38 Moore JH, Gilbert JC, Tsai CT *et al*. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *J Theor Biol* 2006; **241**: 252–61.
39 Chen CC, Schwender H, Keith J, Nunkesser R, Mengersen K, Macrossan P. Methods for identifying SNP interactions: a review on variations of Logic Regression, Random Forest and Bayesian logistic regression. *IEEE/ACM Trans Comput Biol Bioinform* 2011; **8**: 1580–91.
40 Liu Z. Smoking and lung cancer in China: combined analysis of eight case-control studies. *Int J Epidemiol* 1992; **21**: 197–201.
41 Steliga MA, Dresler CM. Epidemiology of lung cancer: smoking, secondhand smoke, and genetics. *Surg Oncol Clin N Am* 2011; **20**: 605–18.
42 Wynder EL, Hoffmann D. Smoking and lung cancer: scientific challenges and opportunities. *Cancer Res* 1994; **54**: 5284–95.
43 Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* 2004; **5**: 389–96.
44 Kurmangaliyev YZ, Gelfand MS. Computational analysis of splicing errors and mutations in human transcripts. *BMC Genomics* 2008; **9**: 13.
45 Queen's university. *F-SNP: a collection of functional SNPs, specifically prioritized for disease association studies.* [Cited 15 Jun 2008.]. Available from URL: http://compbio.cs .queensu.ca/F-SNP/.
46 Ihsan R, Chauhan PS, Mishra AK *et al*. Multiple analytical approaches reveal distinct gene-environment interactions in smokers and non smokers in lung cancer. *PLoS ONE* 2011; **6**: e29431.
47 To-Figueras J, Gené M, Gómez-Catalán J, Piqué E, Borrego N, Corbella J. Lung cancer susceptibility in relation to combined polymorphisms of microsomal epoxide hydrolase and glutathione S-transferase P1. *Cancer Lett* 2001; **173**: 155–62.
48 De Bakker PI, Graham RR, Altshuler D, Henderson BE, Haiman CA. Transferability of tag SNPs to capture common genetic variation in DNA repair genes across multiple populations. *Pac Symp Biocomput* 2006; **11**: 478–86.
49 de Bakker PI, Burtt NP, Graham RR *et al*. Transferability of tag SNPs in genetic association studies in multiple populations. *Nat Genet* 2006; **38**: 1298–303.
50 Duell EJ, Bracci PM, Moore JH, Burk RD, Kelsey KT, Holly EA. Detecting pathway-based gene-gene and gene-environment interactions in pancreatic cancer. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 1470–9.
51 Oki NO, Motsinger-Reif AA. Multifactor dimensionality reduction as a filter-based approach for genome wide association studies. *Front Genet* 2011; **2**: 80.
52 Coffey CS, Hebert PR, Ritchie MD *et al*. An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: the importance of model validation. *BMC Bioinformatics* 2004; **5**: 49.
53 Ritchie MD, Hahn LW, Roodi N *et al*. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* 2001; **69**: 138–47.
54 Ritchie MD, Hahn LW, Moore JH. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genet Epidemiol* 2003; **24**: 150–7.

© 2013 Tianjin Lung Cancer Institute and Wiley Publishing Asia Pty Ltd