RESEARCH ARTICLE

Polymorphisms of *XRCC1* and *XRCC2* DNA Repair Genes and Interaction with Environmental Factors Influence the Risk of Nasopharyngeal Carcinoma in Northeast India

Seram Anil Singh, Sankar Kumar Ghosh*

Abstract

Multiple genetic and environmental factors have been reported to play key role in the development of nasopharyngeal carcinoma (NPC). Here, we investigated interactions of *XRCC1* Arg399Gln and *XRCC2* Arg188His polymorphisms and environmental factors in modulating susceptibility to NPC in Northeast India. One-hundred NPC patients, 90 first-degree relatives of patients and 120 controls were enrolled in the study. *XRCC1* Arg399Gln and *XRCC2* Arg188His polymorphisms were determined using PCR-RFLP, and the results were confirmed by DNA sequencing. Logistic regression (LR) and multifactor dimensionality reduction (MDR) approaches were applied for statistical analysis. The *XRCC1* Gln/Gln genotype showed increased risk (OR=2.76; P<0.024) of NPC. However, individuals with both *XRCC1* and *XRCC2* polymorphic variants had 3.2 fold elevated risk (P<0.041). An enhanced risk of NPC was also observed in smoked meat (OR=4.07; P=0.004) and fermented fish consumers (OR=4.34, P=0.001), and tobacco-betel quid chewers (OR=7.00; P=0.0001) carrying *XRCC1* polymorphic variants. However, smokers carrying defective *XRCC1* gene showed the highest risk (OR = 7.47; P<0.0001). On MDR analysis, the best model for NPC risk was the five-factor model combination of *XRCC1* variant genotype, fermented fish, smoked meat, smoking and chewing (CVC=10/10; TBA=0.636; P<0.0001); whereas in interaction entropy graphs, smoked meat and tobacco chewing showed synergistic interactions with *XRCC1*. These findings suggest that interaction of genetic and environmental factors might increase susceptibility to NPC in Northeast Indian populations.

Keywords: Nasopharyngeal carcinoma - *XRCC1* - *XRCC2* - MDR analysis - first-degree relatives

Asian Pac J Cancer Prev, **17 (6)**, 2811-2819

Introduction

Nasopharyngeal carcinoma (NPC) is a malignant tumor arising in the epithelial lining of the nasopharynx. It is a rare disease with unusual racial and geographical distributions. According to GLOBOCAN 2012, it contributes to about 0.6% of global cancer burden, and is the 19th most commonly diagnosed cancer among males and 21st among females. Higher incidences of NPC have been reported in Southern China, North Africa, and parts of the Mediterranean basin, where it can reach upto 30/100,000 person-years (Chang et al., 2006). In India it contributes to 0.6% of total cancers in male and 0.2% in female, and is mostly prevalent in the north-eastern region of the sub-continent (Kataki et al., 2011).

The National cancer registry programme (NCRP), 2009-11 has reported high age-adjusted incidence rates (AARs) in Nagaland (21/100,000), Mizoram (4.9/100,000), Manipur (4.5/100,000) and Sikkim (3/100,000) in males. It is a complex disease caused by the chronic infection of Epstein Barr virus (EBV), salted fish, smoked meat, tobacco, alcohol, and inherited factors (Jia et al., 2010; Ruan et al., 2010; Bei et al., 2012; Xue et al., 2013; Ghosh et al., 2014a; Tsao et al., 2014; Singh et al., 2015). However, even in endemic areas only few people are susceptible to NPC, suggesting a strong interaction between the environmental and genetic factors in NPC carcinogenesis.

Error prone cellular replication process, exposure to toxic compounds or other environmental carcinogens such as nitrosamines and acetaldehydes etc., may damage the nucleotide residues, which is monitored by the DNA repair system. Any defect in the DNA repair pathway will lead to gene mutations and chromosomal instability, known to play major role in tumorigenesis (Shi et al., 2004; Metsola et al., 2005; Forat-Yazdi et al., 2015). Several studies have been identified genetic polymorphisms in DNA repair genes that show significant association with cancers (Guo et al., 2013; Choudhury et al., 2014). *XRCC1* (x-ray repair cross-complementation group 1) gene is present on chromosome 19q13.2. It is one of the molecules involved in the restoration phase of the base excision repair (BER)

*Department of Biotechnology, Assam University Silchar, Pin, Assam, India *For correspondence: drsankarghosh@gmail.com*

pathway. It encodes protein that interacts with poly-ADP-ribose polymerase (PARP), DNA ligase III (LIG3), DNA polymerase-beta (POLB). Together these proteins are involved in the efficient repair of single-strand breaks induced by ionising agents and/or by reactive oxidising species (Lamerdin et al., 1995; Dianova et al., 2004). Three common polymorphisms (Arg194Trp, Arg280His and Arg399Gln) that lead to amino acids substitutions have been identified in *XRCC1* gene. Carriers of the *XRCC1* Arg399Gln variant allele have altered enzyme functions, higher levels of DNA adduct and a stronger association with cancer (Yu et al., 2004; Choudhury et al., 2014). *XRCC2* (x-ray repair cross-complementing group 2) gene is located on chromosome 7q36.1. The encoded protein is involved in homologous recombination repair (HRR) of double stranded DNA, is a member of the RecA/Rad51- related protein family that maintain chromosome stability and repair DNA damage (Johnson et al., 1999). A common variants within *XRCC2* gene, particularly at exon 3 (Arg188His, rs3218536 or 31479 G > A), has been associated with a significantly increased risk of cancer (Jiao et al., 2008; Choudhury et al., 2014).

In the present case-control study we investigate the interaction of *XRCC1* (Arg399Gln) and *XRCC2* (Arg188His) polymorphisms in NPC and their role in modulating the relationship between tobacco (smoking and betel-quid chewing) and dietary (smoked meat and fermented fish intake) habits, and NPC risk. Furthermore, we also used multifactor dimensionality reduction (MDR) approach to investigate the high-degree gene-environmental interaction in NPC carcinogenesis in the northeast Indian population.

Materials and Methods

Study subjects

The case-control study consisted of 100 histologically confirmed, untreated NPC cases (diagnosed between 2012 and 2014), 90 first degree relative (FDRs) of the patients and 120 controls without family history of cancer. All subjects belong to ethnic community (viz., Manipuri, Naga and Mizo) living in the northeastern states of India. The oral swab and/or peripheral blood of the participating subjects were collected upon written consent. Controls were individually matched to cases in sex, age, ethnicity and neighborhood. The study was approved by the Institutional Ethics Committee, Assam University, Silchar. Precautions were taken to avoid contaminations while handling the samples.

Data collection

A standard pre-designed questionnaire was used to collect general and exposure information of the subjects. Each subject was requested to report information on socio demographic characteristics, tobacco (smoking and betel quid chewing) status and dietary (smoked meat and ferment fish) habits. In this study, smokers and chewers were defined as having smoked or chewed at least 1/day to six months. Those who had not smoked or chewed betel quid were defined as non smokers and chewers. They were further categorized based on their frequency of consumption. Heavy smokers were those who smoked at least ≥ 20 bidis/cigarettes per day for ≥ 20 years, and light smokers where those who smoked < 20 bidis/cigarettes per day for < 20 years. Whereas, heavy chewers where those who chewed ≥ 10 doses per day for ≥ 20 years, and those who chewed < 10 doses per day for < 20 years were defined as light chewers. For dietary habits, subjects were divided to never (who do not consume), regularly (who consumed weekly or more) and occasionally (consuming monthly or biweekly).

DNA extraction and genotyping

The genomic DNA was isolation from the collected blood and oral swaps using the standardized phenol/chloroform/ isoamylalcohol (Ghosh and Mondal, 2012) method, and then stored in TE buffer (10Mm Tris-HCl, pH 0.8, 1 Mm EDTA) at -20ºC before genotyping.

In the *XRCC1* (Arg399Gln) genotype analysis, a 615 bp fragment amplified using forward and reverse primers; 5'-TTGTGCTTTCTCTGTGTCCA-3' and 5'-TCCTCCAGCCTTTTCTGATA-3' (Choudhury et al., 2014), was digested with HpaII restriction enzyme (New England BioLabs, USA); two fragments of 240 and 375 bp represents the wild-type allele GG (Arg/Arg), three fragments of 615, 375 and 240 bp indicates for heterozygous GA (Arg/Gln) and a single 615 bp fragment for the variant allele AA (Gln/Gln).

The *XRCC2* (Arg188His) genotype was determined by amplifying a 205 bp fragment using forward and reverse primers; 5'-TCACCCATCTCTCTGCCTTTTG-3' and 5'-TTCTGATGAGCTCGAGGCTTTC-3' (Choudhury et al., 2014). The resulting fragment was digested with HphI restriction enzyme (New England BioLabs, USA); a single 205 bp fragment represents the wild-type allele GG (Arg/Arg), three fragments of 205, 137 and 68 bp indicates for the heterozygous GA (Arg/Gln) and two fragments of 137 and 68 bp for the variant allele AA (His/His). The RFLP results were confirmed by sequencing 10% of the randomly selected samples from both cases and controls by Sanger sequencing using Genetic Analyzer 3500, Applied BioSystems (Molecular Medicine Lab, Department of Biotechnology, Assam University, Silchar, India).

Statistical analysis

Statistically significant differences of demographic characteristics in the study populations were assessed by $\chi 2$ test. Logistic regression (LR) method were used to analyse the risk association between the dietary habits, tobacco status, *XRCC1* Arg399Gln and *XRCC2* Arg188His genotypes and NPC in terms of odds ratios (ORs), 95% confidence intervals (95% CI), and their corresponding p-values. A p-value of less than 0.05 was considered to be statistically significant. Departures from Hardy-Weinberg equilibrium were evaluated by comparing the expected to observed genotype frequencies using χ^2 tests.

MDR analysis

The multifactorial dimensionality reduction analysis (MDR) software package (www.multifactordimensionalityreduction.org) was used to

detect the gene-environment interactions (Hahn et al., 2003). MDR is a new advanced and non-parametric approach that simultaneously detects and characterizes the combined effects of multiple disease factors. The advantage of using MDR is it overcomes the sample size limitations often encountered by parametric approaches (example: logistic regression analysis) while studying high-level interaction. The best prediction model was selected by maximum testing balance accuracy (TBA) and cross validation consistency (CVC). The best model with highest CVC and TBA was tested by 1000 fold permutation testing and χ^2 test at 0.05% significance levels during MDR analysis.

Interaction entropy graphs

The interaction entropy graphs were constructed based on MDR results to determine synergistic and non-synergistic interactions among the variables (Lavender et al., 2009; Yu et al., 2014). The entropy graphs comprise of nodes containing percentage entropy of each individual variables and connections joining them pairwise showing entropy of interaction between them. Positive entropy signifies synergy and negative entropy indicate redundancy, whereas, zero entropy indicates independence.

Results

Relationship between dietary and tobacco habits and the risk for NPC

The present study includes 100 cases, 90 FDRs of the patient and 120 controls. Pearson chi-square (χ2 test) test were used to determine statistical differences between the study subjects. Significant variations were observed in terms of sex (P= 0.0250), ethnicity (P= 0.0203) and profession (P< 0.0001) in the study population. However, there were no differences with respect to age, BMI and education level among the subjects. Further, we examine the association between dietary and tobacco habits and the risk of NPC in the study population (Table 1). Regular consumption of smoked meat and fermented fish show elevated risk of NPC. The ORs were 2.45 (95% CI: 1.18-5.10; p= 0.016) and 2.25 (95% CI: 1.10-4.60; p= 0.026), respectively, compare to the reference groups. We also observe as dose dependent risk of NPC with the intensity of tobacco habits. Heavy smokers had 4.14 fold (95% CI: 1.84-9.30; p = 0.001) risk of NPC compare with never smokers. Similarly, heavy chewers had 2.64 fold (95% CI: 1.22-5.72; p= 0.014) risk of NPC, when compare with never chewers. In FDRs, there were no significant association between the environmental factors and NPC risk. However, smoked meat consumption showed 2.54 fold (95% CI: 1.26-5.03; p= 0.009) risk for NPC.

Association between XRCC1 Arg399Gln and XRCC2 Arg188His polymorphisms and NPC risk

The genotypes of *XRCC1* Arg399Gln and *XRCC2* Arg188His were determined by detecting the PCR-RFLP band pattern on 1.5% agarose gel and the result was confirmed by randomly sequencing 10% of the samples (Figure 1). The frequency distributions of *XRCC1* Arg399Gln genotypes viz. GG (Arg/Arg), GA (Arg/Gln) and AA (Gln/Gln) were 33%, 49%, 18% and 46.7%, 43.3%, 10% among cases, and controls, respectively. While, the frequency distributions of the GG (Arg/Arg) and GA (Arg/His) genotypes at *XRCC2* Arg188His among cases were 83% and 17% whereas among controls were 89.2% and 10.8%. Logistic regression method was used to analyse the association between *XRCC1* Arg399Gln and *XRCC2* Arg188His polymorphisms and NPC risk (Table 2). Individuals with the *XRCC1* AA (Gln/Gln) genotype had 2.76 fold risk of NPC compared to those carrying the wild type GG (Arg/Gln) genotype. Combine GA and AA (Arg/Gln + Gln/Gln) genotypes of *XRCC1* also showed an elevated risk (OR=2.03, 95% CI: 1.13-3.62, P=0.017) of NPC. The risk associated with individual allele was also determined. Significant risk of NPC was observed in individual with the A-allele as compared to G- allele (OR=1.59, 95% CI: 1.08-2.35, P=0.022). However, we did not observed significant association between *XRCC2*

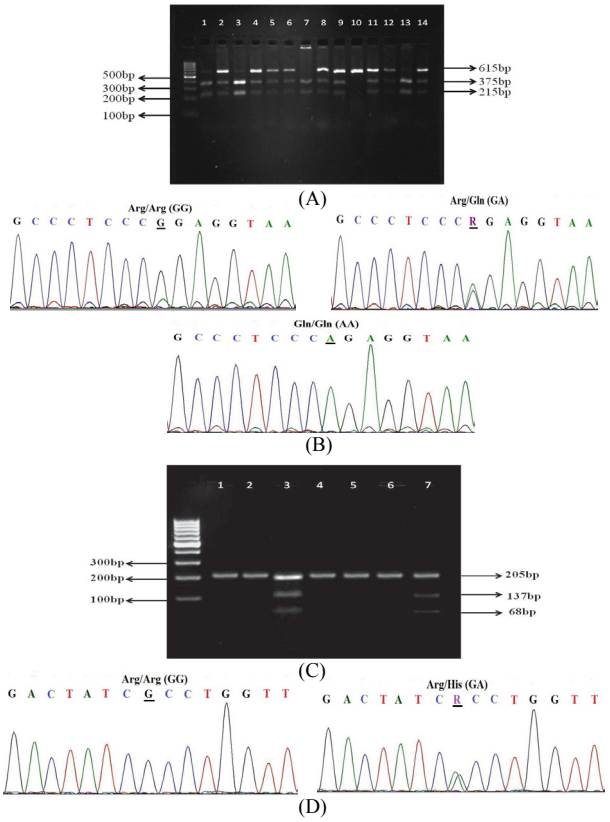


Figure 1. PCR Based Detection of XRCC1 Arg399Gln and XRCC2 Arg188His Gene Polymorphism in NPC. (A) PCR-RFLP patterns of XRCC1 Arg399Gln gene polymorphism separated by agarose gel electrophoresis. Lane 1, 3, 7, and 13 represents wild Arg/Arg (G/G) genotype, lane 2, 4, 5, 6, 8, 9, 11, 12, and 14 represents heterozygous Arg/Gln (G/A) genotype and lane 10 represents mutant Gln/Gln (A/A) genotype. (B) DNA sequencing result of XRCC1 Arg399Gln genotype showing nucleotide variation (in black line). (C) PCR-RFLP patterns of XRCC2 Arg188His gene polymorphism separated by agarose gel electrophoresis. Lane 1, 2, 4, 5, and 6 represents wild Arg/Arg (G/G) genotype and lane 3 and 7 represents heterozygous Arg/His (G/A) genotype. (D) DNA sequencing result of XRCC2 Arg188His genotype showing nucleotide variation (in black line). **R indicates heterozygous genotypes, where green peak denotes adenine (A) while black peak denotes guanine (G)

Table 1. Environmental Risk Factors Associated with NPC

Variables	Case	Controls	*OR (95% CI)	P values	First-degree relatives	*OR (95% CI)	P values
	n=100 (%)	n=120 (%)			n=90 (%)		
Smoked Meat							
Never	22 (22)	50 (41.6)	1	Ref.	26 (28.9)	1	Ref.
Occasionally	26 (26)	32 (26.7)	1.39 (0.61-3.15)	0.426	25 (27.8)	1.74 (0.83-3.64)	0.14
Regular	52 (52)	38 (31.7)	2.45 (1.18-5.10)	0.016	39 (43.3)	2.52 (1.26-5.03)	0.009
Fermented Fish							
Never	40 (40)	75 (62.5)	1	Ref.	63 (68.9)	1	Ref.
Occasionally	19 (19)	18 (15)	1.41 (0.59-3.34)	0.43	12 (13.4)	0.72 (0.31-1.68)	0.452
Regular	41 (41)	27 (22.5)	2.25 (1.10-4.60)	0.026	15 (16.7)	0.56 (0.25-1.19)	0.131
Smoking							
Never	50 (50)	90 (75)	1	Ref.	66 (73.3)	1	Ref.
Light	21 (21)	16 (13.3)	2.87 (1.25-6.59)	0.013	15 (16.7)	1.30 (0.58-2.90)	0.523
Heavy	29 (29)	14 (11.6)	4.14 (1.84-9.30)	0.001	9 (10)	0.81 (0.31-2.09)	0.664
Tobacco-betel quid chewing							
Never	34 (34)	54 (45)	1	Ref.	42 (46.7)	1	Ref.
Light	30 (30)	43 (35.8)	1.04 (0.50-2.21)	0.884	22 (24.5)	0.58 (0.29-1.17)	0.131
Heavy	36 (36)	23 (19.2)	2.64 (1.22-5.72)	0.014	26 (28.8)	1.71 (0.82-3.55)	0.15

Fisher's exact test used to calculate P value and P<0.05 is considered as statistically significance; *Odd ratio (OR) adjusted for age, sex, ethnicity smoked meat and fermented fish intake; smoking and tobacco-betel quid chewing; Bold values indicate statistical significance (P< 0.05)

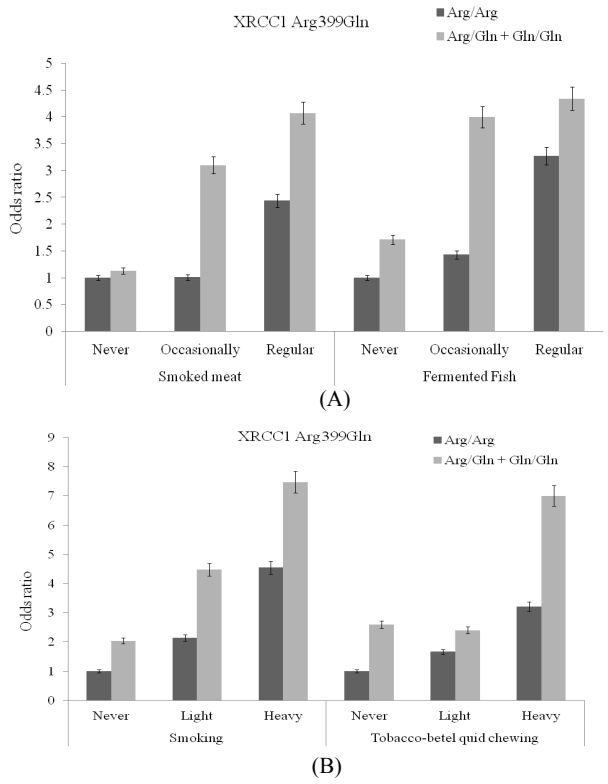


Figure 2. Combined Effect of Dietary and Tobacco Habits with XRCC1 Arg399Gln Genotypes. (A) Regular smoked meat and fermented fish consumers carrying XRCC1 Gln/Gln genotype had 4.07 and 4.43 fold increase risk risk of developing NPC, respectively. (B) Heavy smokers and chewers carrying XRCC1 Gln/Gln genotype had 7.47 and 7 fold increase risk NPC, respectively

Arg188His polymorphism and NPC risk among cases. The distribution and association between *XRCC1* Arg399Gln and *XRCC2* Arg188His polymorphism and NPC risk were also determined among FDRs. The frequencies of *XRCC1* Arg399Gln (43.3%, 42.2% and 14.5%) and *XRCC2* Arg188His (84.5% and 14.5%) genotypes did not show significant variations between the FDRs and the controls. Moreover, there were no significant relation between *XRCC1* and *XRCC2* genotypes and the risk of developing NPC in the FDRs (Table 2).

We further analysed the interaction between *XRCC1* Arg399Gln and *XRCC2* Arg188His genotypes (Table 3). Significantly, elevated risk of NPC (OR=3.2, 95% CI: 1.17-8.74; P=0.041) was observed among individuals carrying defective variants of both *XRCC1* and *XRCC2* genes.

Interaction of XRCC1 Arg399Gln genotypes and environmental factors in NPC

The potential interaction between *XRCC1* Arg399Gln genotypes and environment factors in NPC risk were also investigated (Table 4; Figure 2). Significantly, elevated risk of NPC were observed among regular smoked meat (OR= 4.07, 95% CI, 1.60-10.35; P=0.004) and fermented fish (OR= 4.34, 95% CI, 1.83-10.30; P=0.001) consumers carrying *XRCC1* variant (Arg/Gln + Gln/Gln) genotypes. Among heavy tobacco-betel quid chewers individuals polymorphic for *XRCC1* Arg399Gln had 7 fold (95% CI, 2.33-21.03; P=0.001) risk of NPC compare to non-chewers with the wild type *XRCC1* (GG) genotypes. However, highest risk of NPC was observed among heavy smokers (OR= 7.47, 95% CI, 2.52-22.14; P<0.0001) carrying the defective *XRCC1* gene.

MDR analysis

MDR analysis was used to determine the best-model interaction of genetic and environmental factors for NPC (Table 5). The analysis suggest that smoking was the best one-factor model with CVC 5/10 (TBA=0.547; P<0.0001). While fermented fish and smoked meat was the best two-factors model with CVC of 9/10 (TBA=0.612; P<0.0001), the best three-factor model was the combination of

Table 2. Genotype frequency distribution of XRCC1 Arg399Gln and XRCC2 Arg188His polymorphisms and risk of NPC

Variables	Case	Controls	OR (95% CI)*	P values	First-degree relatives	OR (95% CI)*	P values
	n=100 (%)	n=120 (%)			n=90 (%)		
XRCCI Arg399Gln (rs25487)							
GG (Arg/Arg)	33 (33)	56 (46.7)	1	Ref.	39 (43.3)	1	Ref.
GA (Arg/Gln)	49 (49)	52 (43.3)	1.83 (0.99-3.38)	0.053	38 (42.2)	0.82 (0.45-1.50)	0.53
AA (Gln/Gln)	18 (18)	12 (10)	2.76 (1.14-6.66)	0.024	13 (14.5)	1.13 (0.40-3.18)	0.817
GA+AA (Arg/Gln + Gln/Gln)	67 (67)	64 (53.3)	2.03 (1.13-3.62)	0.017	51 (56.7)	0.82 (0.47-1.42)	0.482
G (Arg) allele frequency	115	164	1	Ref.	116	1	Ref.
A (Gln) allele frequency	85	76	1.59 (1.08-2.35)^^	0.022	64	1.19 (0.79-1.79)^^	0.405
P (HWE) value	0.979	0.988	-	-	0.4555	-	-
XRCC2 Arg188His (rs3218536)							
GG(Arg/Arg)	83 (83)	107 (89.2)	1	Ref.	76 (84.5)	1	Ref.
GA(Arg/His)	17 (17)	13 (10.8)	1.41 (0.61-3.27)	0.423	13 (14.5)	0.68 (0.25-1.79)	0.437
AA(His/His)	-	-	-	-	-	-	-
G (Arg) allele frequency	185	227	1	Ref.	165	1	Ref.
A (His) allele frequency	15	13	1.42 (0.67-3.01)^^	0.435	13	1.38 (0.63-3.00)^^	0.54
P (HWE) value	-	-	-	-	-	-	-

Fisher's exact test used to calculate P value and P<0.05 is considered as statistically significance; *Odd ratio (OR) adjusted for age, sex, ethnicity, smoked meat and fermented fish intake; smoking and tobacco-betel quid chewing; ^^Crude Odd ratio; Bold values indicate statistical significance (P< 0.05)

Table 3. Combined Genotype Analysis of XRCC1 Arg399Gln and XRCC2 Arg188His on Risk of NPC

XRCCI Arg399Gln (rs25487)	XRCC2 Arg188His (rs3218536)	Case	Controls	OR (95% CI)	P values^^
		(n=100)	(n=120)		
GG (Arg/Arg)	GG(Arg/Arg)	29	50	1	Ref.
GG (Arg/Arg)	GA+AA (Arg/His + His/His)	4	6	1.15 (0.32-4.15)	1
GA+AA (Arg/Gln + Gln/Gln)	GG(Arg/Arg)	54	57	1.63 (0.91-2.94)	0.106
GA+AA (Arg/Gln + Gln/Gln)	GA+AA (Arg/His + His/His)	13	7	3.20 (1.17-8.74)	0.041

P< 0.05 is considered statistically significant; Bold values indicate statistical significance (P< 0.05)

Table 4. Association of XRCC1 Arg399Gln Genotype, Stratified by Smoked Meat, Fermented Fish, Smoking, and Tobacco-Betel Quid Habits

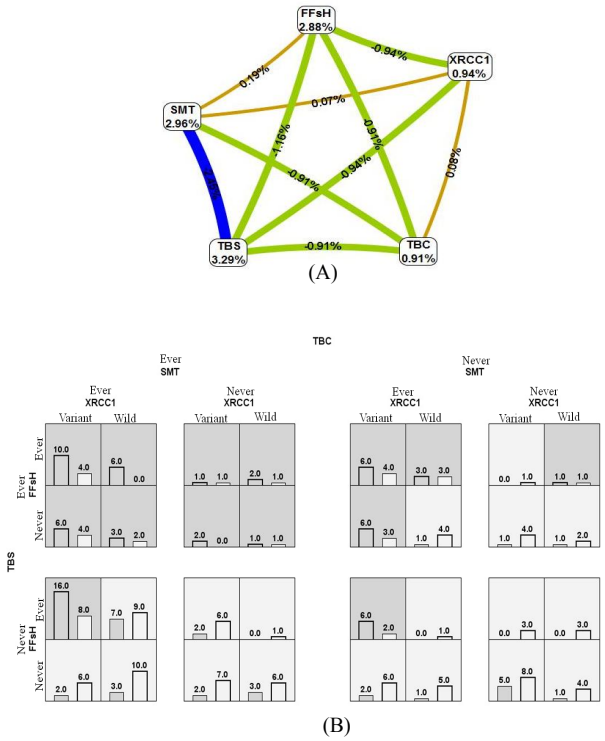
Variables	XRCCI Arg399Gln			XRCCI Arg399Gln		
	GG (Arg/Arg)			GA+AA (Arg/Gln + Gln/Gln)		
	Ca /Co	OR (95% CI)^^	P value	Ca /Co	OR (95% CI)^^	P value
Smoked Meat						
Never	9/22	1	Ref.	13/28	1.13 (0.42-3.09)	1
Occasionally	7/17	1.01 (0.32-3.18)	1	19/15	3.10 (1.12-8.53)	0.045
Regular	17/17	2.44 (0.89-6.71)	0.128	35/21	4.07 (1.60-10.35)	0.004
Fermented Fish						
Never	14/36	1	Ref.	26/39	1.71 (0.78-3.76)	0.236
Occasionally	5/9	1.43 (0.42-4.83)	0.742	14/9	4.00 (1.44-11.11)	0.01
Regular	14/11	3.27 (1.22-8.77)	0.024	27/16	4.34 (1.83-10.30)	0.001
Smoking						
Never	15/42	1	Ref.	35/48	2.04 (0.99-4.23)	0.072
Light	5/6	2.14 (0.59-7.75)	0.294	16/10	4.48 (1.70-11.82)	0.003
Heavy	13/8	4.55 (1.61-12.86)	0.007	16/6	7.47 (2.52-22.14)	<0.0001
Tobacco betel quid chewing						
Never	8/24	1	Ref.	26/30	2.60 (1.01-6.68)	0.068
Light	10/18	1.67 (0.56-4.98)	0.409	20/25	2.40 (0.90-6.39)	0.097
Heavy	15/14	3.21 (1.11-9.32)	0.038	21/9	7.00 (2.33-21.03)	0.001

Ca/Co: Case/Control; ^^Fisher's exact test used to calculate P value and P< 0.05 is considered statistically significant; Bold values indicate statistical significance (P< 0.05)

Table 5. Summary of Multifactorial Dimensionality Reduction Analysis (MDR) for NPC Risk Prediction

Model	TrBA	TBA	CVC	P value
TBS	0.608	0.547	5/10	<0.0001
FFsh, SMT	0.641	0.612	9/10	<0.0001
FFsh, SMT, TBS	0.672	0.629	10/10	<0.0001
XRCC1, FFsh, SMT, TBS	0.689	0.616	8/10	<0.0001
XRCC1, FFsh, SMT, TBS, TBC a	0.707	0.636	10/10	<0.0001

TBA, testing balance accuracy; TrBA, training balance accuracy; CVC, cross-validation consistency; TBS, tobacco smoking; FFsh, fermented fish; SMT, smoked meat, TBC, tobacco chewing; aBest model prediction for NPC risk with highest TrBA, TBA and maximum CVC.



(B)

Figure 3. MDR Analysis of Genetic and Environmental Factors. (A) Interaction entropy graph for gene-environment interaction and NPC risk. This graphical model explains the percent of the entropy in case-control removed by each factor (independent effect) and by each pair-wise combination of attributes (interaction effect). Positive percentage of entropy indicating synergistic interaction and negative values of entropy represent redundancy. The red colour indicating a high degree of synergistic interaction, orange a lesser degree whereas; gold represent midpoint; blue represents the highest level of redundancy followed by green. TBC-tobacco-betel quid chewing, TBS-tobacco smoking, SMT-smoked meat, FFsh-fermented fish. (B) Summary of the five-factor model (XRCC1, fermented fish, smoked meat, smoking and chewing) in MDR analysis. The distribution of high risk (dark shading) and low risk (light shading) combinations associated with NPC risk. The percentage of patients having NPC was represented by left column in each box, whereas right column in each box indicated percentage of controls

fermented fish, smoked meat and smoking with maximum CVC of 10/10 (TBA=0.624; P<0.0001). The best four-factors model was the combinations of *XRCC1* variant genotype, fermented fish, smoked meat and smoking with CVC of 8/10 and highest TBA (0.616) and P<0.0001. However, the best model of all the predictive models was the five-factor model combination of *XRCC1* variant genotype, fermented fish, smoked meat, smoking and chewing having maximum CVC of 10/10 and highest TBA of 0.636 (P<0.0001).

Interaction entropy graphs

Interaction entropy graph was constructed using MDR results to determine synergistic or antagonistic interactions between the genetic and environmental factors in NPC risk (Figure 3). Entropy graph, revealed tobacco smoking, and fermented fish had highest independent effect with percentage entropy of 3.29% and 2.88%, respectively. Smoked meat (2.96%) and tobacco chewing (0.91%) had a synergistic interaction with *XRCC1* variant (0.94%) by removing 0.07% and 0.08% of entropy, respectively.

Discussion

The present study investigates the association between repair genes; *XRCC1* Arg399Gln and *XRCC2* Arg188His polymorphism, and susceptibility to nasopharyngeal cancer in northeast India. The study subjects include NPC patients, their first degree relatives and controls. We use both conventional logistic regression as well as advanced MDR approach to determine high-level interaction between repair genes and environmental factors in NPC risk.

The principle findings suggested that environmental factors (diet and tobacco habits) and *XRCC1* Arg399Gln polymorphism was significant contributor to NPC risk. Previous studies have reported tobacco habits as a major risk factor for cancers (Gupta and Johnson, 2014; Malakar et al., 2014). In a case-control study conducted in northeast India it was seen that tobacco/betel quid chewing and consumption of smoked meat had a profound effect on HNC risk (Mondal et al., 2013; Choudhury et al., 2014; Choudhury and Ghosh, 2014). These environmental factors contain carcinogenic and toxic compound that form DNA adducts causing a wide diversity of DNA damage; gene mutations and genomic instability, and ultimately leading to tumorigenesis. Primary repair system in our body which includes the *XRCC1* and *XRCC2* genes eliminates such damage keeping the genomic integrity of the cells. *XRCC1* gene of the BER pathway encodes a protein having three functional domains that interact with different enzymes to initiate DNA repair of different stages and types (Dianova et al., 2004). *XRCC2* encoded protein play a key role in DNA repair pathway and is required

for RAD51 focus formation. It is an essential part of the HRR pathway and a functional candidate to be involved in tumor progression (Jiao et al., 2008). Polymorphism in *XRCC1* Arg399Gln and *XRCC2* Arg188His may affect the DNA repair capacity of an individual, which can further cause genomic instability and finally leads to cancer. The polymorphisms in these genes have been well documented with cancers (Flores-Obando et al., 2010; Sterpone et al., 2010; Guo et al., 2013; Choudhury et al., 2014). The observed frequencies of 32% for the *XRCC1* 399Gln allele and 10% for the Gln/Gln genotype, are very similar to previous studies conducted on northeast Indian and Chinese populations (Li et al., 2013; Choudhury et al., 2014), but comparatively lower than Turkish population (Kocabas and Karahalil, 2006). Our finding suggests that carriers of AA genotype of *XRCC1* gene had significantly increased risk (2.76 fold) of NPC compare to GG genotype. Association of NPC risk was also observed with the 399Gln (A) allele, compare with the Arg (G) allele (OR=1.59). In consistent to our result a recent study conducted in China showed NPC risk in individual with defective *XRCC1* gene (Li et al., 2013). Furthermore, meta-analysis suggests high associations of *XRCC1* Arg399Gln polymorphism and NPC risk (Huang et al., 2011). However, studies conducted in Chinese (Cao et al., 2006), Taiwanese (Cho et al., 2003), Malaysian (Visuvanathan et al., 2014) and North African (Laantri et al., 2011) population did not reported such association. Unfortunately, there was no report that linked *XRCC2* Arg188His polymorphisms and NPC. In our study also, we did not find any association of *XRCC2* variant genotype and NPC risk.

Single gene polymorphism cannot explain complex disease such as NPC. Therefore, we explore the risk conferred by the combine interaction of *XRCC1* and *XRCC2* polymorphism and NPC risk. Similar to previous reports (Choudhury et al., 2014), we also observed a 3.2 fold elevated risk ($p<0.041$) of NPC in individual carrying the polymorphic variant of both the *XRCC1* and *XRCC2* genes.

Environmental factors alone cannot explain the complex mechanism involved in NPC carcinogenesis. Therefore, the interaction of repair gene and dietary and tobacco habits were also investigated in the study population. Several studies have reported the association of tobacco and DNA repair genes in cancer (Saikia et al., 2014; Uppal et al., 2014). In contrary to studies conducted in north India and China (Cao et al., 2006; Kumar et al., 2012; Li et al., 2013), tobacco habits were found to increase the risk of HNC in individual carrying *XRCC1* Arg399Gln and *XRCC2* Arg188His polymorphic variants (Choudhury et al., 2014) in northeast India. Tobacco and betel-quid constitute large number of polycyclic aromatic hydrocarbon (PAH), alkaloids and other phenolic compounds which are considered as prime risk factors of cancer (Nagaraj et al., 2006; Sabitha et al., 2010; Talukdar et al., 2013). Similarly, smoked meat contains nitrosodimethylamine (NDMA), nitrosodiethylamine (NDEA) and nitrosopyrrolidine (NPYR) which are known mutagen and has proven to be risk factors for NPC (Sarkar et al., 1989; Ghosh et al., 2014b). In our study we used conventional LR method and advanced MDR approach to investigate high-level interaction between repair gene polymorphism and environmental factors in NPC. Tobacco smoking was a strongly associated with NPC risk as shown in LR analysis (Table 1); and the best predictive single-factor model in the MDR analysis (Table 5). In our study, both conventional LR method and advanced MDR approach reveal a high-level interaction between *XRCC1* Arg399Gln polymorphism and environmental factors in NPC risk (Table 4 & 5). Moreover, MDR analysis identify the best model for NPC risk as the interaction between *XRCC1* Arg399Gln genotype, fermented fish, smoked meat, smoking and chewing having 100% CVC and highest TBA. Study in bladder cancer showed four factor model of gene-environmental interaction as the best model to determine the cancer risk (Huang et al., 2007). Similarly, MDR analysis among the northeast Indian population showed four factor model had the best ability to predict HNC risk with TBA of 0.6737 maximum CVC (10/10; $p<0.0001$) (Choudhury and Ghosh, 2014). However, in our study we found the five-factor model as the best model for NPC risk. Interaction entropy graphs were drawn to determine whether the gene-environmental interactions shown by MDR analysis have synergistic or antagonistic effects. Tobacco smoking, smoked meat and fermented fish had the highest individual effects. Though *XRCC1* and tobacco betel-quid chewing had low effects, both showed synergistic interactions in NPC risk. Similarly, interaction was also observed between *XRCC1* and smoked meat. Our investigations indicated that besides environmental factors, individual genetic make-up also plays vital role in NPC carcinogenesis. Our study might have certain drawback for predicting high-order interactions due to the relatively small sample size; however, MDR approach overcomes the limitation for low sample sizes by using cross validation and permutation testing strategy.

We showed that *XRCC1* Arg399Gln polymorphic variant is a strong predisposing risk factor for NPC in the northeast Indian population. Using both conventional LR as well as advanced MDR approach our study further confirmed that the interaction of genetic and environmental factors further increased the risk of NPC. Moreover, implementation of advanced statistical tools in interaction study enables us to identify high-risk marker for cancer risk. Further studies with larger samples size and also in other populations are required before the clinical implications can be considered.

Acknowledgements

Our humble acknowledgement goes to the Department of Biotechnology (DBT), Govt. of India for providing infra-structural facilities (BT/Med/NE-SFC/2009) and (BT/NE/TBP/204(Med)/3/2011) for conducting research on Cancer, and Naga Hospital Administration, Kohima; RIMS, Imphal; Civil Hospital, Aizwal for the biological samples.

References

Bei JX, Jia WH, Zeng YX (2012). Familial and large-scale

case-control studies identify genes associated with nasopharyngeal carcinoma. *Semin Cancer Biol*, **22**, 96-106.
Cao Y, Miao XP, Huang MY, et al (2006). Polymorphisms of *XRCC1* genes and risk of nasopharyngeal carcinoma in the Cantonese population. *BMC Cancer*, **6**, 167.
Chang F, Hu T, Wang G (2006). [Relationship between CYP1A1 and GSTM1 genetic polymorphisms and lung cancer susceptibility in population of Inner Mongolia]. *Zhongguo Fei Ai Za Zhi*, **9**, 413-7.
Cho EY, Hildesheim A, Chen CJ, et al (2003). Nasopharyngeal carcinoma and genetic polymorphisms of DNA repair enzymes *XRCC1* and hOGG1. *Cancer Epidemiol Biomarkers Prev*, **12**, 1100-4.
Choudhury JH, Choudhury B, Kundu S, et al (2014). Combined effect of tobacco and DNA repair genes polymorphisms of *XRCC1* and *XRCC2* influence high risk of head and neck squamous cell carcinoma in northeast Indian population. *Med Oncol*, **31**, 67.
Choudhury JH, Ghosh SK (2014). Gene-environment interaction and susceptibility in head and neck cancer patients and in their first-degree relatives: a study of Northeast Indian population. *J Oral Pathol Med*.
Dianova, II, Sleeth KM, Allinson SL, et al (2004). *XRCC1*-DNA polymerase beta interaction is required for efficient base excision repair. *Nucleic Acids Res*, **32**, 2550-5.
Flores-Obando RE, Gollin SM, Ragin CC (2010). Polymorphisms in DNA damage response genes and head and neck cancer risk. *Biomarkers*, **15**, 379-99.
Forat-Yazdi M, Gholi-Nataj M, Neamatzadeh H, et al (2015). Association of *XRCC1* Arg399Gln Polymorphism with Colorectal Cancer Risk: A HuGE Meta Analysis of 35 Studies. *Asian Pac J Cancer Prev*, **16**, 3285-91.
Ghosh SK, Mondal R (2012). Quick diagnosis of female genital tuberculosis using multiplex fast polymerase chain reaction in Southern Assam, India. *Int J Gynaecol Obstet*, **118**, 72-3.
Ghosh SK, Singh AS, Mondal R, et al (2014a). Dysfunction of mitochondria due to environmental carcinogens in nasopharyngeal carcinoma in the ethnic group of Northeast Indian population. *Tumour Biol*.
Ghosh SK, Singh AS, Mondal R, et al (2014b). Dysfunction of mitochondria due to environmental carcinogens in nasopharyngeal carcinoma in the ethnic group of Northeast Indian population. *Tumour Biol*, **35**, 6715-24.
Guo S, Li X, Gao M, et al (2013). The relationship between *XRCC1* and XRCC3 gene polymorphisms and lung cancer risk in northeastern Chinese. *PLoS One*, **8**, 56213.
Gupta B, Johnson NW (2014). Systematic review and meta-analysis of association of smokeless tobacco and of betel quid without tobacco with incidence of oral cancer in South Asia and the Pacific. *PLoS One*, **9**, 113385.
Hahn LW, Ritchie MD, Moore JH (2003). Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics*, **19**, 376-82.
Huang GL, Guo HQ, Yu CY, et al (2011). *XRCC1* polymorphisms and risk of nasopharyngeal carcinoma: a meta-analysis. *Asian Pac J Cancer Prev*, **12**, 2329-33.
Huang M, Dinney CP, Lin X, et al (2007). High-order interactions among genetic variants in DNA base excision repair pathway genes and smoking in bladder cancer susceptibility. *Cancer Epidemiol Biomarkers Prev*, **16**, 84-91.
Jia WH, Luo XY, Feng BJ, et al (2010). Traditional Cantonese diet and nasopharyngeal carcinoma risk: a large-scale case-control study in Guangdong, *China*. *BMC Cancer*, **10**, 446.
Jiao L, Hassan MM, Bondy ML, et al (2008). *XRCC2* and XRCC3 gene polymorphism and risk of pancreatic cancer. *Am J Gastroenterol*, **103**, 360-7.
Johnson RD, Liu N, Jasin M (1999). Mammalian *XRCC2* promotes the repair of DNA double-strand breaks by homologous recombination. *Nature*, **401**, 397-9.
Kataki AC, Simons MJ, Das AK, et al (2011). Nasopharyngeal carcinoma in the Northeastern states of India. *Chin J Cancer*, **30**, 106-13.
Kocabas NA, Karahalil B (2006). *XRCC1* Arg399Gln genetic polymorphism in a Turkish population. *Int J Toxicol*, **25**, 419-22.
Kumar A, Pant MC, Singh HS, et al (2012). Associated risk of *XRCC1* and XPD cross talk and life style factors in progression of head and neck cancer in north Indian population. *Mutat Res*, **729**, 24-34.
Laantri N, Jalbout M, Khyatti M, et al (2011). *XRCC1* and hOGG1 genes and risk of nasopharyngeal carcinoma in North African countries. *Mol Carcinog*, **50**, 732-7.
Lamerdin JE, Montgomery MA, Stilwagen SA, et al (1995). Genomic sequence comparison of the human and mouse *XRCC1* DNA repair gene regions. *Genomics*, **25**, 547-54.
Lavender NA, Benford ML, VanCleave TT, et al (2009). Examination of polymorphic glutathione S-transferase (GST) genes, tobacco smoking and prostate cancer risk among men of African descent: a case-control study. *BMC Cancer*, **9**, 397.
Li Q, Wang JM, Peng Y, et al (2013). Association of DNA base-excision repair *XRCC1*, OGG1 and APE1 gene polymorphisms with nasopharyngeal carcinoma susceptibility in a Chinese population. *Asian Pac J Cancer Prev*, **14**, 5145-51.
Malakar M, Devi KR, Phukan RK, et al (2014). CYP2E1 genetic polymorphism with dietary, tobacco, alcohol habits, H. pylori infection status and susceptibility to stomach cancer in Mizoram, India. *Asian Pac J Cancer Prev*, **15**, 8815-22.
Metsola K, Kataja V, Sillanpaa P, et al (2005). *XRCC1* and XPD genetic polymorphisms, smoking and breast cancer risk in a Finnish case-control study. *Breast Cancer Res*, **7**, 987-97.
Mondal R, Ghosh SK, Choudhury JH, et al (2013). Mitochondrial DNA copy number and risk of oral cancer: a report from Northeast India. *PLoS One*, **8**, 57771.
Nagaraj NS, Beckers S, Mensah JK, et al (2006). Cigarette smoke condensate induces cytochromes P450 and aldo-keto reductases in oral cancer cells. *Toxicol Lett*, **165**, 182-94.
Ruan HL, Xu FH, Liu WS, et al (2010). Alcohol and tea consumption in relation to the risk of nasopharyngeal carcinoma in Guangdong, China. *Front Med China*, **4**, 448-56.
Sabitha K, Reddy MV, Jamil K (2010). Smoking related risk involved in individuals carrying genetic variants of CYP1A1 gene in head and neck cancer. *Cancer Epidemiol*, **34**, 587-92.
Saikia BJ, Phukan RK, Sharma SK, et al (2014). Interaction of *XRCC1* and XPD gene polymorphisms with lifestyle and environmental factors regarding susceptibility to lung cancer in a high incidence population in North East India. *Asian Pac J Cancer Prev*, **15**, 1993-9.
Sarkar S, Nagabhushan M, Soman CS, et al (1989). Mutagenicity and carcinogenicity of smoked meat from Nagaland, a region of India prone to a high incidence of nasopharyngeal cancer. *Carcinogenesis*, **10**, 733-6.
Shi Q, Wang LE, Bondy ML, et al (2004). Reduced DNA repair of benzo[a]pyrene diol epoxide-induced adducts and common XPD polymorphisms in breast cancer patients. *Carcinogenesis*, **25**, 1695-700.
Singh SA, Choudhury JH, Kapfo W, et al (2015). Influence of the CYP1A1 T3801C Polymorphism on Tobacco and Alcohol-Associated Head and Neck Cancer Susceptibility in Northeast India. *Asian Pac J Cancer Prev*, **16**, 6953-61.
Sterpone S, Mastellone V, Padua L, et al (2010). Single-

nucleotide polymorphisms in BER and HRR genes, *XRCC1* haplotypes and breast cancer risk in Caucasian women. *J Cancer Res Clin Oncol*, **136**, 631-6.

Talukdar FR, Ghosh SK, Laskar RS, et al (2013). Epigenetic, genetic and environmental interactions in esophageal squamous cell carcinoma from northeast India. *PLoS One*, **8**, 60996.

Tsao SW, Yip YL, Tsang CM, et al (2014). Etiological factors of nasopharyngeal carcinoma. *Oral Oncol*, **50**, 330-8.

Uppal V, Mehndiratta M, Mohapatra D, et al (2014). XRCC-1 Gene Polymorphism (Arg399Gln) and Susceptibility to Development of Lung Cancer in Cohort of North Indian Population: A Pilot Study. *J Clin Diagn Res*, **8**, CC17-20.

Visuvanathan S, Chong PP, Yap YY, et al (2014). Distribution and haplotype associations of XPD Lys751Gln, *XRCC1* Arg280His and *XRCC1* Arg399Gln polymorphisms with nasopharyngeal carcinoma in the Malaysian population. *Asian Pac J Cancer Prev*, **15**, 2747-51.

Xue WQ, Qin HD, Ruan HL, et al (2013). Quantitative association of tobacco smoking with the risk of nasopharyngeal carcinoma: a comprehensive meta-analysis of studies conducted between 1979 and 2011. *Am J Epidemiol*, **178**, 325-38.

Yu HP, Zhang XY, Wang XL, et al (2004). DNA repair gene *XRCC1* polymorphisms, smoking, and esophageal cancer risk. *Cancer Detect Prev*, **28**, 194-9.

Yu X, Liu J, Zhu H, et al (2014). Synergistic association of DNA repair relevant gene polymorphisms with the risk of coronary artery disease in northeastern Han Chinese. *Thromb Res*, 133, 229-34.