4.1. Prevalence and environmental risk factors associated with nasopharyngeal carcinoma (NPC) among the ethnic population of Northeast India

4.1.1. Nasopharyngeal carcinoma in different cancer registries

The global cancer database of international association of research in cancer (IARC), GLOBOCAN in 2012 reported 14.1 million new cancer incidences, and 8.2 million deaths worldwide. Accordingly, in India there were nearly 1 million new incidences and 0.6 million cancer death. In that year, India reported 3947 (0.4%) new cases and 2836 (0.4%) death of NPC. In males there were 2956 (0.6%), and in females 991 (0.2%) new cases of NPC (Table 4.1.1).

Table 4.1.1 NPC incidence, mortality and prevalence: Estimate for India, 2012

	Males (%)	**Females (%)**	**Both (%)**
Incidence of NPC	2956 (0.6)	991 (0.2)	3947 (0.4)
Ages < 65	2305 (78)	811 (82)	3116 (79)
Ages ≥ 65	651 (22)	180 (18)	831 (21)
$CR/10^5$	0.5	0.2	0.3
$ASR/10^5$	0.5	0.2	0.4
Cumulative risk (0-74 years)	0.06	0.002	0.04
Mortality OF NPC	2094 (0.6)	742 (0.2)	2836 (0.4)
$CR/10^5$	0.3	0.1	0.2
$ASR/10^5$	0.4	0.1	0.3
Cumulative risk (0-74 years)	0.05	0.01	0.03
5-years Prevalence (Prop.)	74881 (1.7)	2479 (0.6)	9967 (1.1)
Total cancer incidence	477482	537452	1014934
Total cancer mortality	356730	326100	682830

Proportion by 100,000

CR- Crude incidence rate and AAR- Age standardized rates per 100,000

Adapted from GLOBOCAN 2012 (*http://www.globocan.iarc.fr/*)

NPC incidence were higher in age group above 65 years, with 2305 (78%) and 811 (82%) cases in both males and females, respectively. NPC is rare cancer in Indian sub-continent, the age standardized incidence rates per 100,000 (AARs) was low compare (0.4) to other endemic areas of the world. 5-years prevalence data reported 9967 (0.6%) of NPC in India (Table 4.1.1).

There are 23 Population Based Cancer Registries (PBCR) in India working under the National Cancer Registry Programme of the Indian Council of Medical Research. Of these 9 PBCRs covers the eight Northeasters States of India, including Assam (population: 30.94 million), Arunachal Pradesh (1.255 million), Manipur (2.722 million), Meghalaya (2.651 million), Mizoram (1.016 million), Nagaland (2.275 million), Tripura (3.658 million), and Sikkim (619,000). During 2009-2010, PBCRs in northeast India have reported a high age adjusted incidence rates (AAR) of NPC. The AARs of NPC for males and females registered in north-eastern PBCRs are shown in Table 4.1.2. In males, Nagaland state had AARs of 21/100,000, among the highest reported in the world; followed by Mizoram and Manipur states with a high AAR of 4.9 and 4.5 per 100,000, respectively. While, Sikkim state (AAR= 3/100,000) and Meghalaya state (AAR= 1.8/100,000) reported AAR of NPC. The district wise distribution (population scattered over various districts within a State) of the AARs of NPC in Aizawl District was 6.7/100,000; and the Imphal West district in Manipur State reported AAR of 4/100 000. In female, Nagaland state (AAR= 4.6/100,000) and Mizoram state (AAR= 3.4/100,000) reported the high incidence of NPC. However, Aizawl District had high AAR of 5.2 per 100,000 in females. Several other districts in Mizoram, Meghalaya, Assam and Manipur states recorded high AARs in both males and females, but this cannot be regarded as very significant because only less than 10 cases of cancer were recorded.

Table 4.1.2 Comparison of age adjusted incidence rates (AAR) of PBCRs in northeast India, 2009-2010.

Male			Female		
S. no	**Region**	**AAR**	**S. no**	**Region**	**AAR**
1	Nagaland State	21	1	Aizawl District	5.2
2	Aizawl District	6.7	2	Nagaland State	4.6
3	Mizoram State (MZ)	4.9	3	Mizoram State (MZ)	3.4
4	MR Excl. Imphal West	4.6	4	MZ-Excl. Aizawl	2.4
5	Manipur State (MR)	4.5	5	Sikkim State	1.8
6	Imphal West District	4	6	MR-Excl Imphal West	1.7
7	MZ- Excl. Aizawl	3.9	7	Manipur State	1.6
8	East Khasi Hills	3.3	8	Imphal West District	1.1
9	Sikkim State	3	9	East Khasi Hills	0.9
10	Meghalaya	1.8	10	Meghalaya	0.8
11	Kamrup Urban District	0.9	11	Kamrup Urban District	0.5
12	Cachar District	0.8	12	Dibrugarh District	0.3
13	Tripura State	0.6	13	Cachar District	0.3
14	Dibrugarh District	0.6	14	Tripura State	0.2

AAR- Age standardised rates per 100,000 populations

Adapted from the PBCRs of National Cancer Registry Programme of Indian Council of Medical Research (ICMR), 2009-2010 (*http://www.pbcrindia.org/*)

Based on the PBCR report of 2009-10 our study was focus on three States viz., Nagaland, Manipur and Mizoram having the highest AARs of NPC. Afterwards, we also collected the incidence information of new NPC cases in males and females during 2010-2012, from Regional Institute of Medical Sciences, Imphal (RIMS); Government Civil Hospital, Aizwal (GCH) and Naga Hospital Administration, Kohima (NHAK) which have the PBCRs covering these main study areas. It is clear that during 2010-2012 Nagaland State had the highest incidence cases of NPC both in male (23%) and female (13.7%) followed by Manipur and Mizoram States (Table 4.1.3). However, highest AAR among male was from Nagaland (21.7) followed by Mizoram (4.4) and Manipur (3.6). Similarly, Nagaland (6.1) had the highest AAR among female compare to the other States.

Table 4.1.3 Number of cases, Percentage (%) and Rate of Incidence Cases, 2010-2012

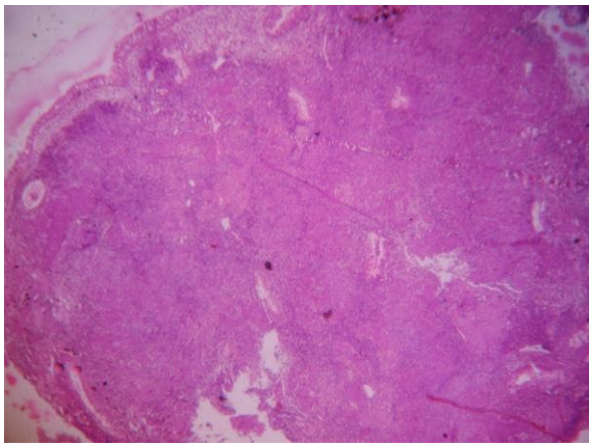
	Manipur (RIMS)	Mizoram (GCH)	Nagaland (NHAK)
No. Cases in males	113	124	150
Percentage (%)	5.9	3.2	23.4
CR	2.7	3.9	13.3
AAR	3.6	4.4	21.7
TR	8.7	10.2	56.1
No. Cases in females	50	65	60
Percentage (%)	2.3	2.2	13.7
CR	1.2	2.3	4.8
AAR	1.5	3.5	6.1
TR	2.7	8.8	18.2

CR- Crude incidence rate, TR- Truncated rate; and AAR- Age standardized rates per 100,000

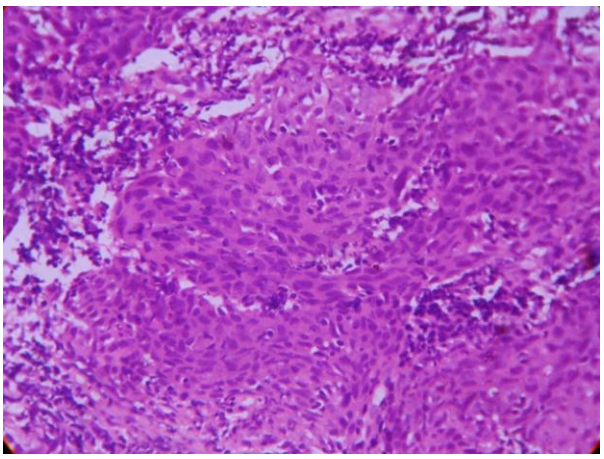
4.1.2. Demographic characteristics of the study population

The demographic characteristics and socio-economic status of the study population is represented in Table 4.1.4. A total of 123 histopathologically confirm NPC cases (Figure 4.1.1), their 100 first degree relatives (FDRs) and 189 controls without family history of cancer participated in the study. The cases and controls samples comprised of three different ethnic groups viz., Manipuri (23.6 & 28.6%), Naga (67.5 & 63%) and Mizo (8.9 & 8.4%) from northeast India. No significance difference were observed with respect to the ethnicity between the cases and controls (P=0.621). Relatively higher proportions of patients were males (59.3%) and large majority were of age group ≥50 years (65.1%) at the time of diagnosis. However, there was no statistically significant difference between the cases and controls in terms of sex (P=0.6985) and age (P=0.9203), suggesting that sex and age matching was effective. Cases tends to have lower education level (70%, P=0.039) and most of them were farmers (55.3%, P=0.0332). Almost, 65% cases were found to have a family history of NPC and reported ear, nose and throat (ENT) related problems (41.5%) compare to controls (23.8%). The difference was significant (P=0.001). Moreover, body mass-index (BMI) of the individual P=0.014), presence of soot inside the house (P<0.0001), and type of fuel used for cooking (P=0.022) also showed significance variation between the cases and controls. We do not observed variation in type of the house; ventilation per room, and kitchen (outside/inside).

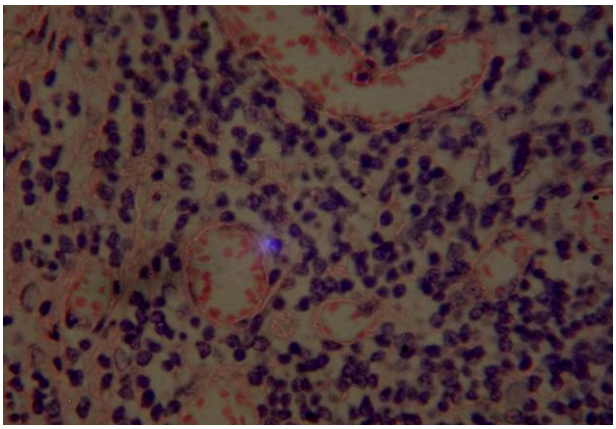
Between the FDRs and controls significance variations were observed in ethnicity (P=0.0024), profession (P=0.0002), and presence of soot in house (P=0.0181). However, no difference was observed with respect to other variables considered in the study population (Table 4.1.4).



a



b



c

Figure 4.1.1 Photomicrograph of nasopharynx showing **a** the respiratory lining and underneath is the lymphoepithelial lesion (H&E stain, 10X), **b** the malignant epithelial nest admixed with lymphocytes (H&E stain, 40X) and **c** section without presence of malignant cells (Images were provided from RIMS, Imphal and NHAK, Kohima)

Table 4.1.4 Demographic characteristics and socioeconomic status of the cases, FDRs and controls

Variables	**Case, N =123 (%)**	**Control, N=189 (%)**	***P* value**	**FDR, N=100 (%)**	***P* value**
Sex:					
Males	73 (59.3)	108 (57.2)	0.6985	69 (69)	0.0656
Females	50 (40.7)	81 (42.8)		31 (31)	
Age Group:					
≥50	80 (65.1)	122 (64.6)	0.9203	59 (59)	0.4237
<50	43 (34.9)	67 (35.4)		41 (41)	
Ethnicity:					
Manipuri	29 (23.6)	54 (28.6)	0.6219	42 (42)	**0.0024**
Naga	83 (67.5)	119 (63)		42 (42)	
Mizo	11 (8.9)	16 (8.4)		16 (16)	
BMI Kg/m^2:					
Underweight (< 18.5)	28 (22.8)	19 (10.1)	**0.0146**	11 (11)	0.4453
Normal Weight (18.5-22.99)	53 (43)	98 (51.9)		45 (45)	
Overweight (23.0-27.49)	31 (25.2)	59 (31.2)		32 (32)	
Obesity (27.5-more)	11 (9)	13 (6.8)		12 (12)	
Profession:					
Famer	68 (55.3)	76 (40.2)	**0.0332**	28 (28)	**0.0002**
Service	32 (26)	66 (35)		23 (23)	
Others	23 (18.7)	47 (24.8)		49 (49)	
Education Level:					
High School or less	86 (70)	109 (57.7)	**0.039**	63 (63)	0.4503
University or More	37 (30)	80 (42.3)		37 (37)	
Type of House:					
*Kacca- Bamboo/Mud/Wood	74 (60.2)	112 (59.3)	0.8625	56 (56)	0.6801
*Pucca- RCC	49 (39.8)	77 (40.7)		44 (44)	
Ventilation per room:					
No Window	12 (9.8)	11 (5.8)	0.4253	7 (7)	0.3135
Single Window	72 (58.5)	114 (60.3)		51 (51)	
≥ 2 Window	39 (31.7)	64 (33.9)		42 (42)	
SOOT in house:					
Present	58 (47.2)	47 (24.9)	**<0.0001**	39 (39)	**0.0181**
Absent	65 (52.8)	142 (75.1)		61 (61)	
Kitchen:					
Separate	102 (82.9)	162 (85.7)	0.5071	79 (79)	0.1963
Not Separate	21 (17.1)	27 (14.3)		21 (21)	
Cooking Fuel Used:					
Gas	53 (43.1)	108 (57.1)	**0.0228**	48 (48)	0.1565
Wood Fire	54 (43.9)	55 (29.1)		30 (30)	
Both	16 (13)	26 (13.8)		22 (22)	
Family History of NPC:					
No	42 (34.1)	-	-	-	-
Yes	81 (65.9)	-		100 (100)	
^^ENT Problem if any:					
Present	51 (41.5)	45 (23.8)	**0.001**	32 (32)	0.1738
Absent	72 (58.5)	144 (76.2)		68 (68)	

*Distribution in frequencies were tested by chi-square test, and $P < 0.05$ is considered statistically significant value ; ^^In patients ENT present before diagnosis

Bold values indicate statistical significance ($P<0.05$)

4.1.3. Environmental risk factors associated with NPC

We investigated the environmental risk factors associated with NPC in the study population (Table 4.1.5). Dietary (smoked meat and fermented fish intake) and lifestyle-related habits (tobacco habits, alcohol intake and herbal medicine use), and viral factor (EBV-infection) were considered.

4.1.3.1. Dietary habits and risk o NPC

In the case-control dataset, we observed that regular consumption of smoked meat and fermented fish were associated with an elevated risk of NPC. The ORs was (OR=2.49, 95% CI: 1.33-4.67; *P*=0.004) and (OR=1.98; 95% CI: 1.09-3.6; P=0.024), respectively (Figure 4.1.2). Similarly, in FDRs when their dietary practices were examined with respect to those of controls, it was observed that there is an increased risk for NPC with regular intake of smoked meat (OR= 1.84; 95 % CI, 1.0-3.39; P=0.048). However, consumption of fermented fish (P>0.05) did not show significant association with NPC.

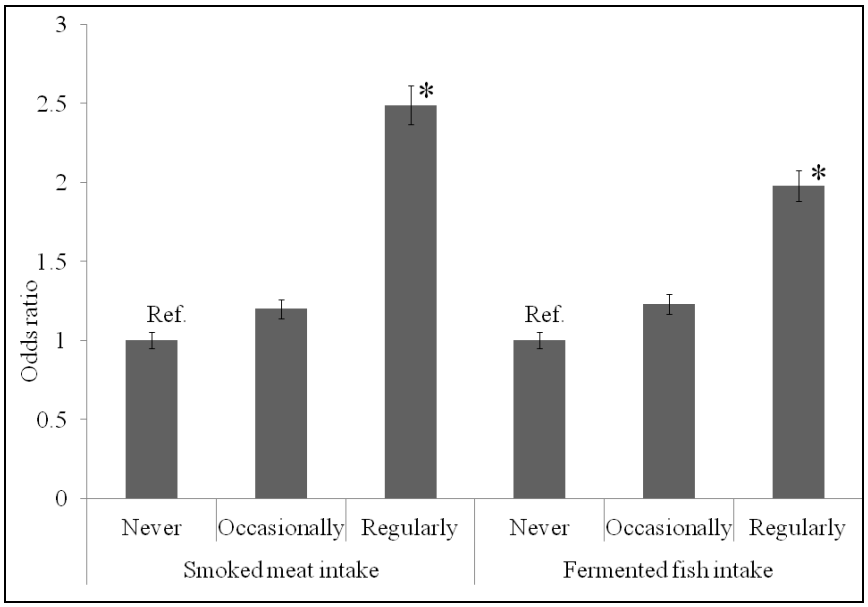


Figure 4.1.2 Bar diagram showing the risk (Odds ratios) of NPC associated with environmental factors. Regular consumption of smoked meat (OR=2.49) and fermented fish (OR=1.98).

4.1.3.2. Tobacco and alcohol habits and herbal medicine use and risk of NPC

Smoking, and tobacco-betel quid chewing showed a dose-dependent risk association with heavy chewers and smoker had higher NPC risk; the ORs were 2.45 (95% CI: 1.24-4.7; P=0.009) and 3.8 (95% CI: 1.95-7.7; P<0.0001), respectively. Light smokers also had a significant increase risk of NPC (OR=2.9; P=0.009). While herbal medicine use had 2.22 fold risk of NPC. However, no significant risk association was observed with alcohol drinking in NPC (Figure 4.1.3). When the association of tobacco habits and alcohol intake and NPC risk were examined in FDRs no significant risk was found (P>0.05). Similarly, herbal medicine use were not associated with NPC risk in FDRs (P>0.05) (Table 4.1.5)

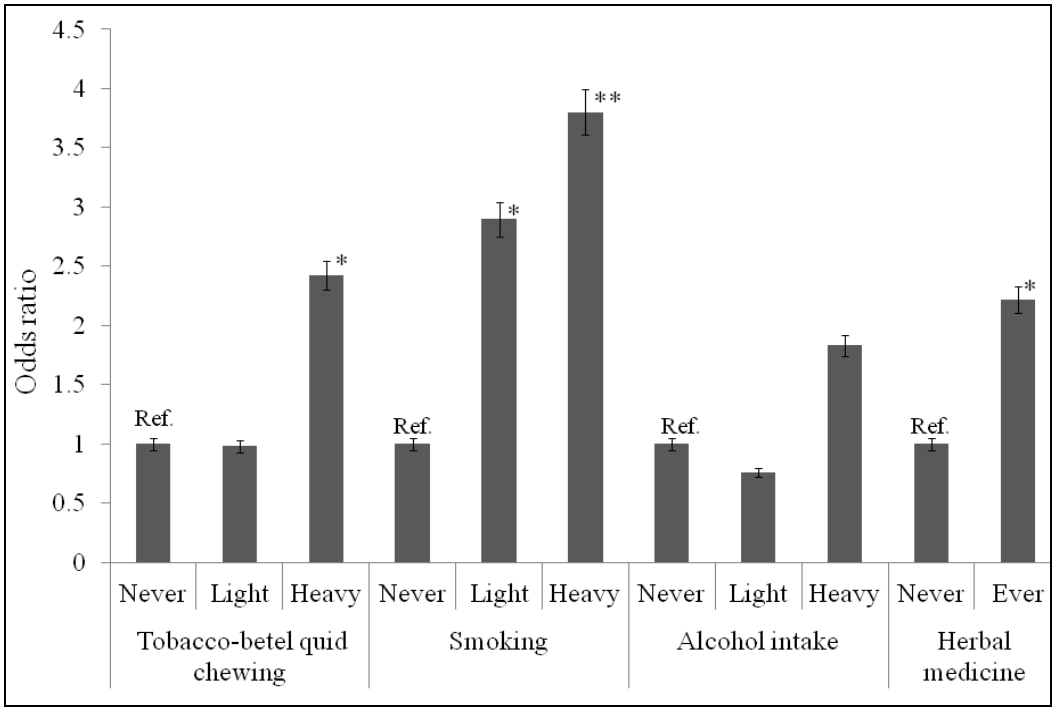


Figure 4.1.3 Bar diagram showing the risk (Odds ratios) of NPC associated with environmental factors. Heavy tobacco-betel quid chewing (OR=2.42) and smoking (OR=3.8) were associated with NPC risk. Alcohol drinking was not associated with NPC risk and herbal medicine use had 2.22 fold risk of NPC risk. (*P<0.05)

*P<0.05; **P<0.0001

Table 4.1.5 Smoked meat, fermented fish, tobacco and alcohol habits and the risk of NPC

Variables	Case, n=123 (%)	Control, n= 189 (%)	^^ORs (95% CI)	*P value	FDRs n= 100 (%)	^^ORs (95% CI)	*P value
Smoked meat intake							
Never	28 (22.8)	78 (41.3)	1.0	Ref.	33 (33)	1.0	Ref.
Occasionally	30 (24.4)	57 (30.2)	1.2 (0.6-2.41)	0.601	23 (23)	0.99 (0.50-1.93)	0.97
Regularly	65 (52.8)	54 (28.5)	2.49 (1.33-4.67)	**0.004**	44 (44)	1.84 (1.0-3.39)	**0.048**
Fermented fish intake							
Never	48 (39)	111 (58.7)	1.0	Ref.	63 (63)	1.0	Ref.
Occasionally	20 (16.3)	28 (14.8)	1.23 (0.57-2.67)	0.592	13 (13)	0.83 (0.37-1.8)	0.627
Regularly	55 (44.7)	50 (26.5)	1.98 (1.09-3.6)	**0.024**	24 (24)	1.05 (0.52-2.0)	0.875
Tobacco-betel quid chewing							
Never	42 (34.2)	84 (44.5)	1.0	Ref.	44 (44)	1.0	Ref.
Light	37 (30.1)	68 (36)	0.98 (0.52-1.84)	0.965	30 (30)	0.86 (0.47-1.58)	0.639
Heavy	44 (35.7)	37 (19.5)	2.42 (1.24-4.7)	**0.009**	26 (26)	1.20 (0.61-2.36)	0.591
Smoking							
Never	63 (51.2)	143 (75.7)	1.0	Ref.	68 (68)	1.0	Ref.
Light	22 (17.9)	21 (11.1)	2.9 (1.3-6.49)	**0.009**	15 (15)	1.39 (0.64-2.99)	0.399
Heavy	38 (30.9)	25 (13.2)	3.8 (1.95-7.7)	**<0.0001**	17 (17)	1.77 (0.83-3.75)	0.137
Alcohol intake							
Never	55 (44.7)	107 (56.6)	1.0	Ref.	46 (46)	1.0	Ref.
Light	33 (26.8)	50 (26.5)	0.76 (0.39-1.48)	0.433	32 (32)	1.33 (0.73-2.43)	0.338
Heavy	35 (28.5)	32 (16.9)	1.83 (0.92-3.65)	0.084	22 (22)	1.44 (0.72-2.89)	0.300
Herbal Medicine							
Never	64 (52)	120 (63.5)	1.0	Ref.	69 (69)	1.0	Ref.
Ever	59 (48)	69 (36.5)	2.22 (1.27-3.87)	**0.005**	31 (31)	0.74 (0.42-1.3)	0.305
EBV-*LMPI*							
Negative	25 (20.3)	69 (36.5)	1.0	Ref.	44 (44)	1.0	Ref.
Positive	98 (79.7)	120 (63.5)	2.87 (1.57-5.23)	**0.001**	56 (56)	0.76 (0.45-1.3)	0.330

Ca/Co: Case/Control Ca *cases*, Co *controls*, FDRs *first degree relatives*

^^ Odds adjusted for sex, age, ethnicity, smoked meat, fermented fish, smoking, tobacco-betel quid, alcohol, herbal medicine and EBV infection as appropriate

***** Fisher's exact test used to calculate *P* value and $P<0.05$ considered as statistically significance; Bold values indicate statistical significance ($P<0.05$)

4.1.3.3. Epstein Barr Virus (EBV) infection and risk of NPC development

We investigate Epstein Barr Virus infection to analyses their risk associated with NPC in our study. EBV infection was detected following PCR amplification of viral *EBV-LMP1* gene and by agarose gel electrophoresis of the PCR products on 2% gel (Figure 4.1.4). It was observed that 79.1% of the cases were *EBV-LMP1* positive; while only 63.5 and 56% of the controls and FDRs were positive for *EBV-LMP1* viral gene. EBV positive subjects were associated with 2.87 fold increased risk of NPC in the cases-control dataset. However, there were no risk association of EBV infection and the risk of NPC among FDRs.

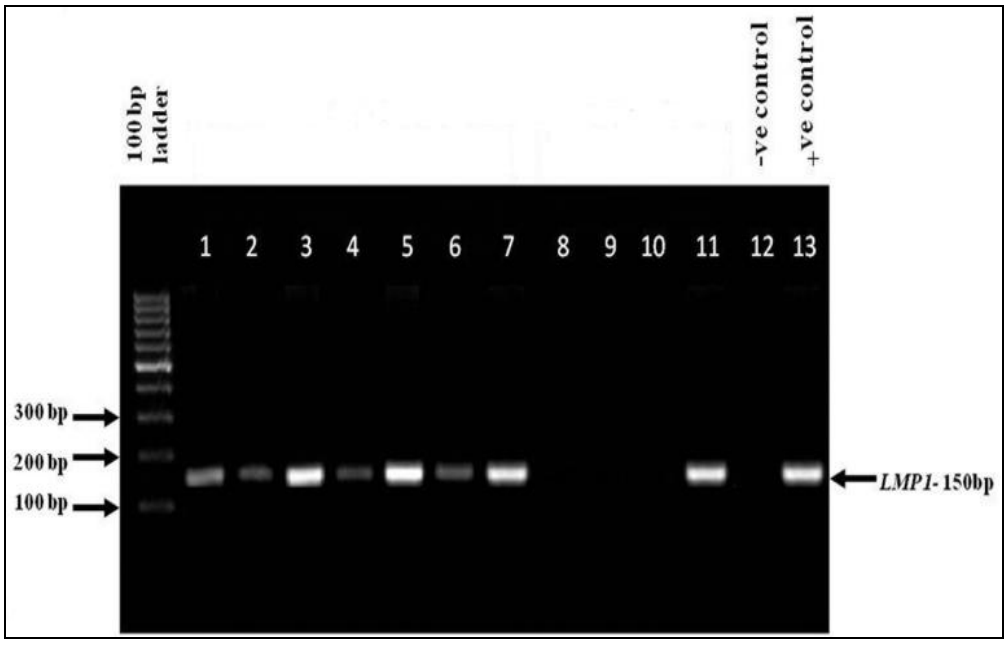


Figure 4.1.4 PCR based detection of *EBV-LMP1* in NPC. Lane 1-7 & 11 represent samples positive for EBV, lane 8-10 represents samples negative for EBV along with –ve control in lane 12

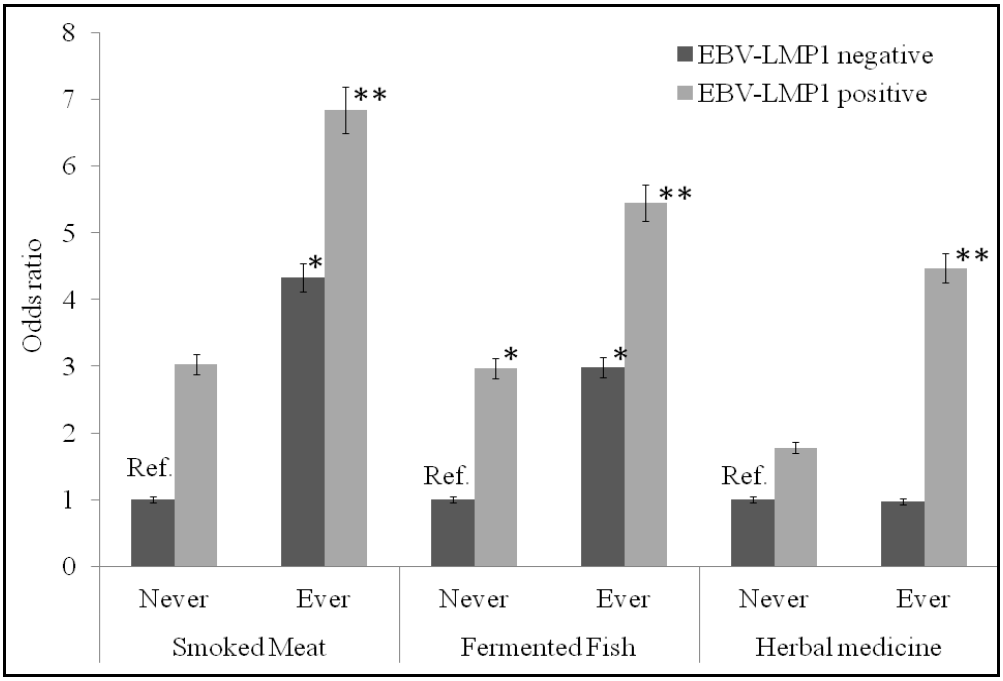
4.1.4. Interaction of Epstein Barr Virus (EBV) and other environmental risk factors in NPC

When the distribution of EBV among the study population was determined (Table 4.1.6); 46.4 % cases and 33.9% controls were found to be positive for EBV among males. Similarly, individual belonging to the age group ≥50 were more infected with EBV both in cases (52%) and controls (40.8%). Ethnicity wise we found that Naga has the highest infected cases (52.9%) and controls (39.6%).

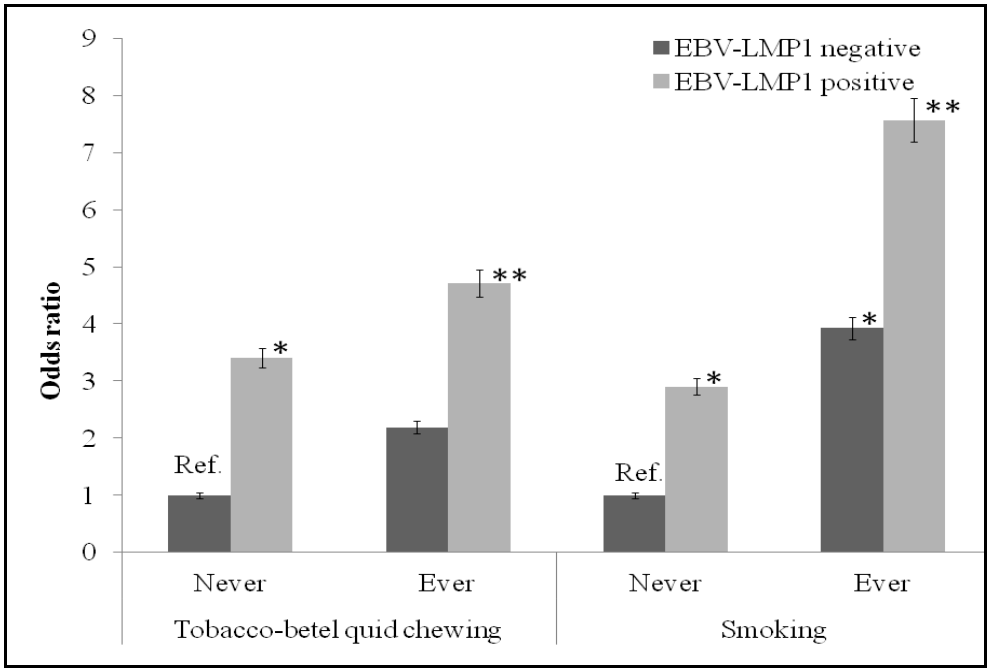
Table 4.1.6 Sex, age and ethnicity wise detection of EBV in the study population

Variables	**Cases (%)**		**Controls (%)**	
	EBV-positive	EBV-negative	EBV- positive	EBV- negative
Sex:				
Males	57 (46.4)	16 (13)	64 (33.9)	44 (23.3)
Females	41 (33.3)	9 (7.3)	51 (27)	30 (15.8)
Age Group:				
≥50	64 (52)	16 (13)	77 (40.8)	45 (23.8)
<50	34 (27.7)	9 (7.3)	38 (20.1)	29 (15.3)
Ethnicity:				
Manipuri	25 (20.3)	4 (3.2)	33 (17.5)	21 (11.1)
Naga	65 (52.9)	18 (14.7)	75 (39.6)	44 (23.4)
Mizo	8 (6.5)	3 (2.4)	7 (3.7)	9 (4.7)

We also analysed whether the relationship between EBV infection and NPC risk were modulated by other environmental risk factors (Figure 4.1.5; Table 4.1.7). In our study we observed that consumption of smoked meat and fermented fish significantly modulate the NPC risk in EBV infected individuals. The risk of NPC were found to increase among EBV positive subjects with intake of smoked meat (OR= 6.84; P<0.0001), fermented fish (OR=5.45; P<0.0001) and herbal medicine use (OR=4.47; P<0.0001). Tobacco habits also interact significantly with EBV. Tobacco-betel quid chewing had 4.71 fold risk (P<0.0001) of NPC in EBV infected subjects. However, highest risk of NPC was observed among smokers (OR=7.57; P<0.0001) positive for EBV.



a



b

Figure 4.1.5 Odds ratio for interaction of EBV and dietary and tobacco habits **a** EBV positive subjects consuming smoked meat (OR=6.84), fermented fish (OR=5.45), and herbal medicine (OR=4.47) have increased risk of NPC; **b** tobacco-betel quid chewing (OR=4.71) and smoking (OR=7.57) also interacts with EBV and show elevated risk.

*P<0.05; **P<0.0001

Table 4.1.7 Risk of NPC by intake of smoked meat, fermented fish, tobacco smoking, herbal medicine use and EBV infection

Variables	EBV-*LMP1* Negative			EBV-*LMP1* Positive		
	Ca /Co	**ORs (95% CI)^^**	***P*- value***	**Ca/Co**	**ORs (95% CI)^^**	***P*- value***
Smoked meat intake						
Never	5/31	1.0	Ref.	23/47	3.03 (1.06-8.71)	0.39
Ever	30/43	4.33 (1.53-12.24)	**0.005**	75/68	6.84 (2.55-18.35)	**<0.0001**
Fermented fish intake						
Never	9/48	1.0	Ref.	39/63	2.97 (1.36-6.51)	**0.007**
Ever	16/26	2.95 (1.18-7.37)	**0.023**	59/52	5.45 (2.52-11.77)	**<0.0001**
Tobacco-betel quid chewing						
Never	7/34	1.0	Ref.	35/50	3.40 (1.37-8.49)	**0.009**
Ever	18/40	2.19 (0.82-5.79)	0.158	63/65	4.71 (1.96-11.29)	**<0.0001**
Smoking						
Never	12/58	1.0	Ref.	51/85	2.90 (1.43-5.88)	0.002
Ever	13/16	3.93 (1.52-10.13)	**0.006**	47/30	7.57 (3.52-16.30)	**<0.0001**
Herbal Medicine						
Never	12/35	1.0	Ref.	52/85	1.78 (0.86-3.72)	0.156
Ever	13/39	0.97 (0.40-2.39)	1.00	46/30	4.47 (2.02-9.89)	**<0.0001**

Ca/Co: Case/Control Ca *cases*, Co *controls*.

* Fisher's exact test used to calculate *P* value and $P<0.05$ considered as statistically significance.

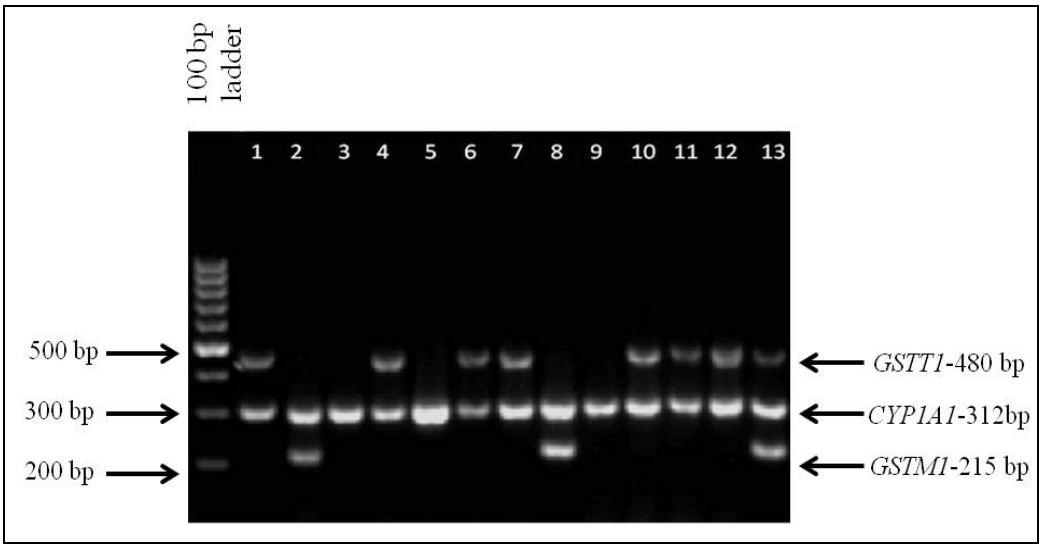
Bold values indicate statistical significance ($P<0.05$)

4.2. Metabolic genes, their interaction with environmental factors in NPC risk

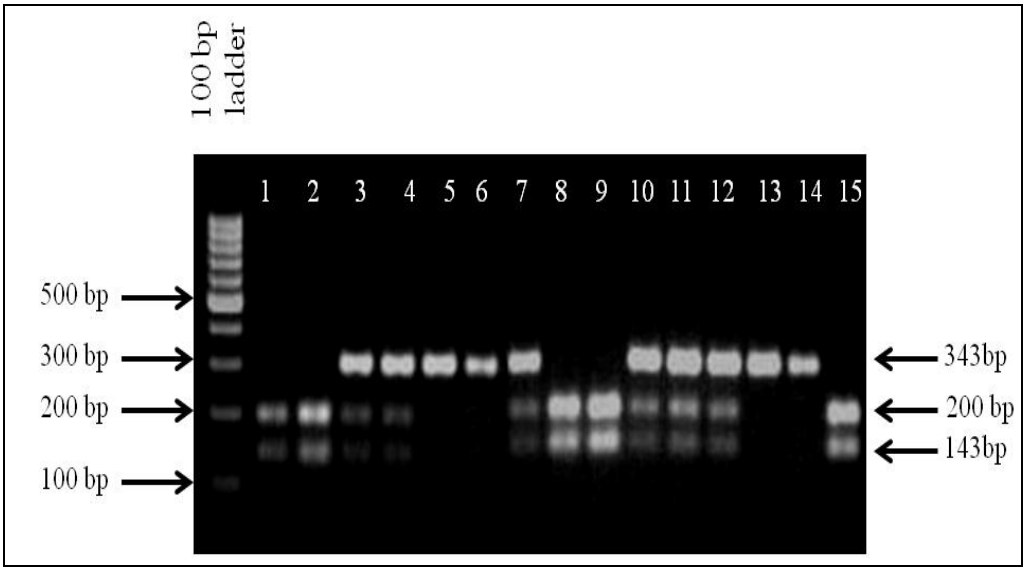
The effect of polymorphisms in major metabolic Phase I (CYPs) *CYP1A1* T3801C, and Phase II (GSTs) *GSTM1* and *GSTT1* gene polymorphisms on NPC, and their differential effect according to diet (smoked meat and fermented fish) and lifestyle (tobacco-betel quid chewing and smoking) were investigated. Further, the degree of risk of NPC among individuals carrying more than one unfavourable genotype was also determined. A total of 402 samples; 123 cases, 90 FDRs and 189 controls were included in the study.

4.2.1. Polymorphisms in metabolic genes (*GSTM1*, *GSTT1* and *CYP1A1*) and the risk associated with NPC

The genotypes of *GSTM1, GSTT1* and *CYP1A1 T3801C,* were determined by observing the band pattern on 1.5% agarose gel (Figure 4.2.1). Furthermore, the PCR product of *CYP1A1* gene was sequenced to confirm the RFLP results (Figure 4.2.2). The frequency distributions were 66.4 and 43.4% for *GSTM1* null genotype, and 45.5 and 36.5%, for *GSTT1* null genotype in cases and controls. The genotypes of *CYP1A1* T3801C viz. wild (TT), heterozygous (TC) and homozygous (CC) variants had frequencies of 40.7, 42.2, and 17.1% in cases while it was 47.1, 36.5, and 16.4% in controls, respectively. Logistic regression method was used to analyses the association between *GSTT1, GSTM1* and *CYP1A1* T3801C genotypes, and NPC risk. It was found that *GSTM1* null genotype was associated with 2.76 fold risk of NPC (95% CI: 1.61-4.71; $P<0.0001$). *GSTT1* and *CYP1A1* T3801C genotypes did not show a significant risk to NPC in the study population (Table 4.2.1; Figure 4.2.3).



a



b

Figure 4.2.1 Polymorphism in *GSTM1*, *GSTT1* and *CYP1A1* metabolic genes **a** Ethidium bromide stained gel *GSTM1* null genotype (lanes 1, 3, 4, 5, 6, 7, 9, 10, 11 and 12); *GSTT1* null genotype (lanes 2, 3, 5, 8 and 9); *GSTM1-GSTT1* wild type genotype (lane 13) and both *GSTM1-GSTT1* NULL genotypes (lanes 3, 5 and 9); **b** Ethidium bromide stained gel *CYP1A1* TT wild genotype (lanes 5, 6, 13 and 14); *CYP1A1* TC heterozygous genotype (lanes 3, 4, 7, 10, 11, and 12); *CYP1A1* CC mutant genotype (lanes 1, 2, 8, 9, and 15).

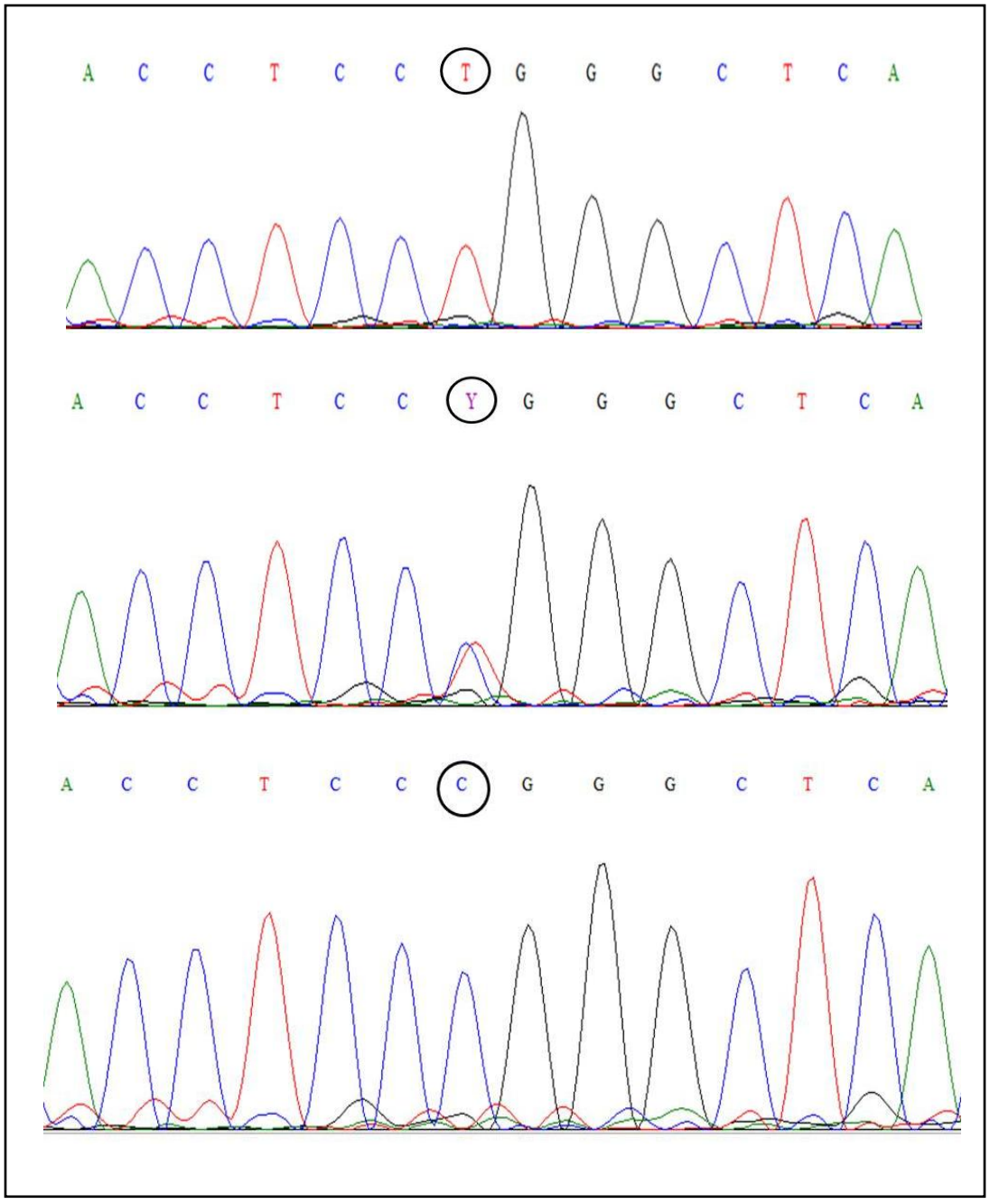


Figure 4.2.2 Sanger sequencing results showing nucleotide changes confirming PCR-RFLP results (mark by black line); Nucleotide change marked as Y indicates the heterozygous genotype, where red peak denotes Thymine (T) while blue peak denotes Cytosine (C).

Table 4.2.1 Distribution of *GSTM1*, *GSTT1* and *CYP1A1* T3801C genotype among the study subjects

Variables	**Case, n=123 (%)**	**Control, n= 189 (%)**	**^^ORs (95% CI)**	***P* value**	**FDR, n= 100 (%)**	**^^ORs (95% CI)**	***P* value**
GSTM1							
Positive	41 (33.3)	107 (56.6)	1.0	Ref.	39 (39)	1.0	Ref.
Negative	82 (66.4)	82 (43.4)	2.76 (1.61-4.71)	**<0.0001**	61 (61)	2.1 (1.22-3.65)	**0.007**
GSTT1							
Positive	67 (54.5)	120 (63.5)	1.0	Ref.	63 (63)		
Negative	56 (45.5)	69 (36.5)	1.36 (0.8-2.31)	0.248	37 (37)	1.25 (0.72-2.18)	0.418
CYP1A1							
TT	50 (40.7)	89 (47.1)	1.0	Ref.	44 (44)	1.0	Ref.
TC	52 (42.2)	69 (36.5)	1.44 (0.79-2.59)	0.225	39 (39)	0.871 (0.47-1.59)	0.665
CC	21 (17.1)	31 (16.4)	1.03 (0.48-2.21)	0.931	17 (17)	1.005 (0.46-2.15)	0.990
TC+CC	73 (59.3)	100 (52.9)	1.3 (0.75-2.23)	0.34	56 (56)	0.911 (0.52-1.58)	0.740
χ^2 (HWE), *P* value	1.34, 0.245	7.1, 0.007	-	-	2.5, 0.112	-	-

Ca/Co: Case/Control Ca *cases*, Co *controls*

^^ Odds adjusted for sex, age, ethnicity, smoked meat, fermented fish, smoking, tobacco-betel quid, alcohol, EBV, herbal medicine and *CYP1A1* T3801C, *GSTM1* and *GSTT1* genotypes as appropriate

***** Fisher's exact test used to calculate *P* value and $P<0.05$ considered as statistically significance

Bold values indicate statistical significance ($P<0.05$)

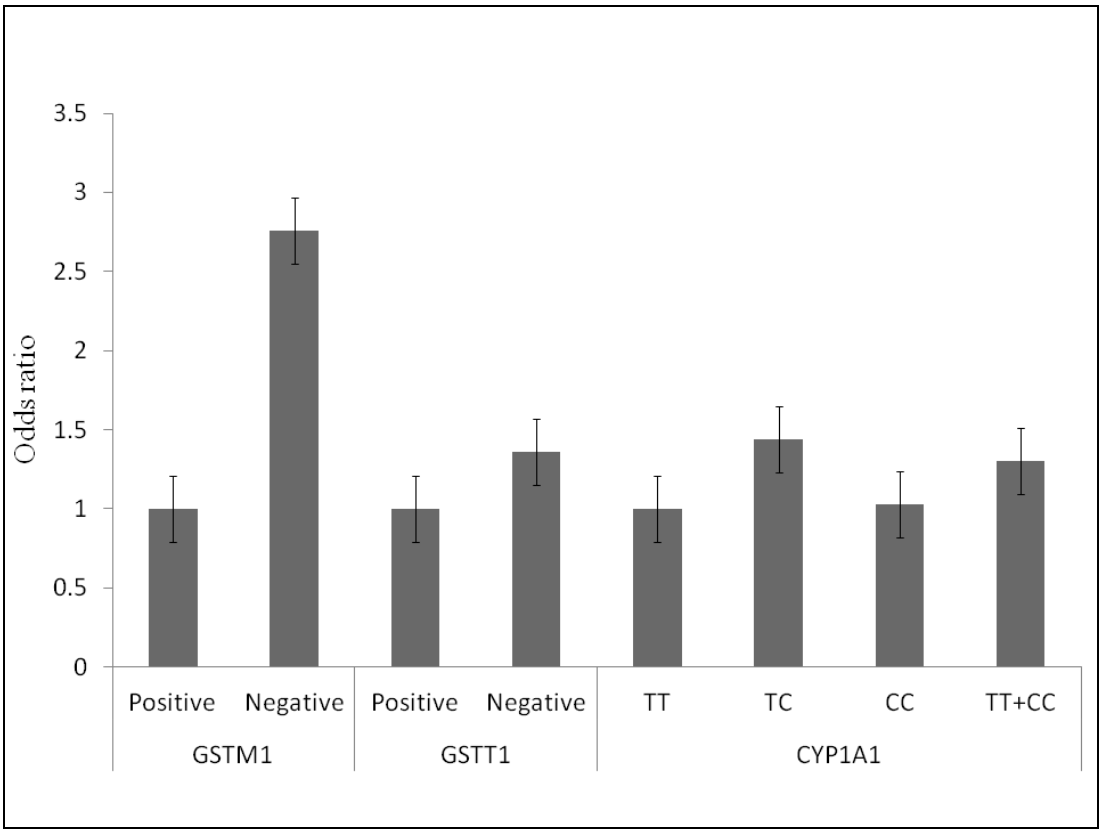


Figure 4.2.3 Bar diagram showing the risk (Odds ratios) of NPC associated with *CYP1A1* T3801C, *GSTT1* and *GSTM1* polymorphism. *GSTM1* null genotypes (OR=2.49) was associated with NPC risk. *CYP1A1* T3801C and *GSTT1* polymorphisms were not associated with NPC risk.

The distribution and association between *GSTT1*, *GSTM1* and *CYP1A1* T3801C polymorphism and NPC risk were also determined in FDRs. *GSTM1* and *GSTT1* null genotypes have frequency distribution of 61% and 37%, respectively. The three genotypes of *CYP1A1* T3801C viz. wild (TT), heterozygous (TC) and homozygous (CC) variants had frequencies of 44, 39, and 17% in FDRs. *GSTM1* null genotype (OR=2.1; P=0.007) was significantly associated with NPC risk in FDRs. However, there were no significant relation between *GSTT1* and *CYP1A1*genotypes and the risk of developing NPC in the FDRs (Table 4.2.1).

The interactions of *GSTT1*, *GSTM1*, *CYP1A1* T3801C genotypes and risk of NPC were also analyzed among the cases (Table 4.2.2). *GSTM1* null individual carrying wild-type GSTT1 genotype had 1.95 fold risk (*P*=0.003) of NPC. Significantly, elevated risk of NPC (OR=3.77, 95% CI: 1.95-7.3; *P*<0.0001) was observed among individuals carrying null genotypes of both *GSTM1* and *GSTT1*. Similarly, *GSTM1* null individual carrying *CYP1A1* T3801C polymorphic variants had 3.2 fold increased risk of NPC (95% CI: 1.65-6.28; *P*=0.001). Both *GSTT1* and *GSTM1* null individual with normal *CYP1A1* gene had 3.77 fold risk (*P*=0.011) of NPC. However, highest risk of NPC (OR=5.71, 95% CI: 2.11-15.45; *P*=0.001) was observed in individuals with defective *GSTM1*, *GSTT1* genotypes and *CYP1A1* T3801C polymorphic variants.

Table 4.2.2 Odds ratios for the interaction of *GSTM1*, *GSTT1* and *CYP1A1* T3801C genotypes in the study subjects

Genotypes	**Cases, n=123(%)**	**Controls, n= 189 (%)**	**ORs (95% CI)**	***P* value**
GSTM1and GSTT1				
M1 (+/+) and T1 (+/+)	26 (21.2)	67 (35.4)	1.0	Ref.
M1 (+/+) and T1 (-/-)	15 (12.2)	40 (21.2)	0.96 (0.45-2.03)	0.928
M1 (-/-) and T1 (+/+)	41(33.3)	53 (28.1)	1.95 (1.06-3.59)	**0.031**
M1 (-/-) and T1 (-/-)	41 (33.3)	29 (15.3)	3.77 (1.95-7.3)	**<0.0001**
***GSTM1*and *CYP1A1* T3801C**				
M1 (+/+) and TT	18 (14.6)	51 (27)	1.0	Ref.
M1 (+/+) and TC or CC	23 (18.7)	56 (29.6)	1.16 (0.57-2.39)	0.716
M1 (-/-) and TT	32 (26)	38 (20.1)	2.39 (1.17-4.85)	**0.021**
M1 (-/-) and TC or CC	50 (40.7)	44 (23.3)	3.22 (1.65-6.28)	**0.001**
***GSTT1*and *CYP1A1* T3801C**				
T1 (+/+) and TT	24 (19.5)	53 (28.1)	1.0	Ref.
T1 (+/+) and TC or CC	43 (34.9)	67 (35.4)	1.46 (0.78-2.7)	0.229
T1 (-/-) and TT	26 (21.2)	36 (19.1)	1.94 (0.97-3.87)	0.057
T1 (-/-) and TC or CC	30 (24.4)	33 (17.4)	1.52 (0.77-3.11)	0.216
***GSTM1, GSTT1* and *CYP1A1* T3801C**				
M1 (+/+), T1 (+/+) and TT	9 (7.3)	30 (15.9)	1.0	Ref.
M1 (+/+), T1 (+/+) and TC or CC	17 (13.8)	37 (19.6)	1.53 (0.59-3.92)	0.374
M1 (-/-), T1 (+/+) and TT	15 (12.2)	23 (12.2)	2.17 (0.8-5.84)	0.124
M1 (-/-), T1 (+/+) and TC or CC	26 (21.2)	30 (15.9)	2.88 (1.61-7.18)	**0.023**
M1 (+/+), T1 (-/-) and TT	9 (7.3)	21 (11.1)	1.42 (0.48-4.2)	0.517
M1 (+/+), T1 (-/-) and TC or CC	6 (4.9)	19 (10)	1.05 (0.32-3.43)	0.932
M1 (-/-), T1 (-/-) and TT	17 (13.8)	15 (7.9)	3.77 (1.36-10.45)	**0.011**
M1 (-/-), T1 (-/-) and TC or CC	24 (19.5)	14 (7.4)	5.71 (2.11-15.45)	**0.001**

Ca/Co: Case/Control Ca *cases*, Co *controls*

* Fisher's exact test used to calculate *P* value and $P<0.05$ considered as statistically significance

Bold values indicate statistical significance ($P<0.05$)

4.2.2. Interaction of metabolic genes and environmental risk factors in NPC progression

The interactions between *GSTM1* genotypes and environmental factors were represented in Table 4.2.3. Occasional and regular smoked meat consumers showed statistically significant interactions among individuals with deletion of *GSTM1*. The ORs of 3.55 (95% CI, 1.50-8.41; *P*=0.005) and 5.56 (95% CI, 2.91-10.62; *P*<0.0001) in *GSTM1* null genotypes carriers where comparatively higher than 0.82 (95% CI, 0.32-2.12; *P*=0.81) and 2.47 (95% CI, 1.06-5.77; *P*=0.053) for those with the gene present. Similarly, occasional and regular fermented fish consumers carrying *GSTM1* null genotype had 6.23 fold (95% CI, 2.47-15.82; *P*<0.0001) and 5.73 fold (95% CI, 2.66-12.34; *P*<0.0001) elevated risk of NPC. Tobacco-betel quit chewers and smokers carrying *GSTM1* null genotypes showed a dose-dependent risk association of NPC. The ORs was 2.81 (95% CI, 1.29-6.12; *P*=0.012) in light and 5.68 (95% CI, 2.46-13.08; *P*<0.0001) in heavy chewers, respectively, for *GSTM1* null individuals compared to 0.88 (95% CI, 0.37-2.13; *P*=0.825) and 2.24 (95% CI, 0.94-5.35; *P*=0.111), respectively, for the wild-type carriers. Light smokers with *GSTM1* null individuals had 7.84 fold (95% CI, 2.80-21.99; *P*<0.0001) increased risk of NPC. However, highest risk of NPC was observed in heavy smokers (OR= 12.67, 95% CI, 4.95-32.39; *P*<0.0001) carrying *GSTM1* null genotypes.

Significant interaction was also observed between *GSTT1* genotypes and regular consumption of smoked meat. The OR was 3.99 (95% CI, 1.84-8.68; *P*=0.001) for individuals with *GSTT1* null genotype compare to 3.46 (95% CI, 1.65-7.23; *P*=0.001) for the wild-type carriers (Table 4.2.4). Similarly, regular fermented fish consumers carrying *GSTT1* null genotype had 3.5 fold (95% CI, 1.73-7.09; *P*=0.001)

risk of NPC compare to 2.27 fold (95% CI, 1.15-4.49; *P*=0.002) among individuals with *GSTT1* gene present. However, there were no interactions between *GSTT1* and occasional smoked meat and fermented fish consumers.

Heavy tobacco-betel quid chewers, carrying *GSTT1* null genotype had 3.51 fold (95% CI, 1.44-9.42) elevated risk of NPC while there was no-risk association in light chewers. Similarly, significant interaction was observed in smokers carrying defective *GSTT1* genotype. The OR was 4.60 (95% CI, 1.29-16.4; *P*=0.025) in light and 4.46 (95% CI, 1.89-10.56; *P*=0.001) in heavy smokers with *GSTT1* null genotype compare 2.89 (95% CI, 1.31-6.37; *P*=0.01) and 4.82 (95% CI, 2.21-10.54; *P*<0.0001) in individuals with wild-type *GSTT1* gene.

Furthermore, we observed a significant interaction of *CYP1A1* polymorphisms with the environmental factors (Table 4.2.5). Regular consumption of smoked meat in individuals with *CYP1A1* T3801C variant (TC + CC) genotypes had 4.12 fold (95% CI, 1.89-8.99; *P*<0.0001) increased risk of NPC whereas wild-type carriers had 4.38 fold (95% CI, 1.85-10.35; *P*=0.001) risk. Similarly, regular fermented fish consumers carrying the *CYP1A1* T3801C variant genotypes had an OR of 3.32 (95% CI, 1.56-7.05; *P*=0.003) versus 2.53 (95% CI, 1.20-5.35; *P*=0.023) in individual with the TT genotype. A significant interaction was noted among heavy tobacco-betel quid chewers and smokers in those individual polymorphic for *CYP1A1* T3801C. The ORs was 2.86 (95% CI, 1.20-6.82; *P*=0.03) and 7.13 (95% CI, 2.88-17.68; *P*<0.0001), respectively, for individuals carrying *CYP1A1* T3801C variant genotypes, which was significantly higher than individuals with *CYP1A1* T3801C TT genotype.

Table 4.2.3 Association between *GSTM1* genotype and NPC, stratified by smoked meat, fermented fish, smoking and tobacco-betel quid habits

Variables	***GSTM1* positive**			***GSTM1* negative**		
	Ca/Co	OR (95% CI)	**P* value	Ca/Co	OR (95% CI)	**P* value
Smoked meat intake						
Never	13/44	1.0	Ref.	15/34	1.49 (0.63-3.52)	0.386
Occasionally	9/37	0.82 (0.32-2.12)	0.810	21/20	3.55 (1.50-8.41)	**0.005**
Regular	19/26	2.47 (1.06-5.77)	0.053	46/28	5.56 (2.91-10.62)	**<0.0001**
Fermented fish intake						
Never	17/59	1.0	Ref.	31/52	2.07 (1.03-4.14)	0.057
Occasionally	2/18	0.39 (0.08-1.77)	0.345	18/19	6.23 (2.47-15.82)	**<0.0001**
Regular	22/30	2.55 (1.19-5.47)	**0.020**	33/20	5.73 (2.66-12.34)	**<0.0001**
Smoking						
Never	18/76	1.0	Ref.	45/67	2.84 (1.5-5.35)	**0.001**
Light	9/14	2.71 (1.03-7.13)	0.054	13/7	7.84 (2.80-21.99)	**<0.0001**
Heavy	14/17	3.48 (1.47-8.24)	**0.008**	24/8	12.67 (4.95-32.39)	**<0.0001**
Tobacco-betel quid chewing						
Never	15/17	1.0	Ref.	27/37	2.29 (1.07-4.88)	**0.039**
Light	11/39	0.88 (0.37-2.13)	0.825	26/29	2.81 (1.29-6.12)	**0.012**
Heavy	15/21	2.24 (0.94-5.35)	0.111	29/16	5.68 (2.46-13.08)	**<0.0001**

Ca/Co: Case/Control Ca *cases*, Co *controls*

* Fisher's exact test used to calculate *P* value and *P*<0.05 considered as statistically significance

Bold values indicate statistical significance (*P*<0.05)

Table 4.2.4 Association between *GSTT1* genotype and NPC, stratified by smoked meat, fermented fish, smoking and tobacco-betel quid habits

Variables	***GSTT1* positive**			***GSTT1* negative**		
	Ca/Co	OR (95% CI)	**P* value	Ca/Co	OR (95% CI)	**P* value
Smoked meat intake						
Never	16/49	1.0	Ref.	12/29	1.27 (0.53-3.02)	0.654
Occasionally	14/40	1.22 (0.55-2.73)	0.682	14/17	2.52 (1.03-6.16)	0.059
Regular	35/31	3.46 (1.65-7.23)	**0.001**	30/23	3.99 (1.84-8.68)	**0.001**
Fermented fish intake						
Never	28/71	1.0	Ref.	20/40	1.27 (0.64-2.52)	0.593
Occasionally	13/20	1.65 (0.73-3.72)	0.279	7/8	2.22 (0.76-6.48)	0.227
Regular	26/29	2.27 (1.15-4.49)	**0.022**	29/21	3.50 (1.73-7.09)	**0.001**
Smoking						
Never	29/89	1.0	Ref.	34/54	1.93 (1.06-3.51)	**0.033**
Light	16/17	2.89 (1.31-6.37)	**0.010**	6/4	4.60 (1.29-16.4)	**0.025**
Heavy	22/14	4.82 (2.21-10.54)	**<0.0001**	16/11	4.46 (1.89-10.56)	**0.001**
Tobacco-betel quid chewing						
Never	24/55	1.0	Ref.	18/29	1.42 (0.62-3.02)	0.435
Light	22/43	1.17 (0.58-2.36)	0.721	15/25	1.38 (0.62-3.03)	0.536
Heavy	21/22	2.19 (1.02-4.67)	0.051	23/15	3.51 (1.58-7.82)	**0.003**

Ca/Co: Case/Control Ca *cases*, Co *controls*

* Fisher's exact test used to calculate *P* value and $P<0.05$ considered as statistically significance

Bold values indicate statistical significance ($P<0.05$)

Table 4.2.5 Association between *CYP1A1* T3801C genotype and NPC, stratified by smoked meat, fermented fish, smoking and tobacco-betel quid habits

Variables	***CYP1A1* TT**			***CYP1A1* TC + CC**		
	Ca/Co	OR (95% CI)	**P* value	Ca/Co	OR (95% CI)	**P* value
Smoked meat intake						
Never	12/42	1.0	Ref.	16/36	1.56 (0.66-3.69)	0.381
Occasionally	13/27	1.69 (0.68-4.19)	0.346	17/30	1.98 (0.83-4.72)	0.131
Regular	25/20	4.38 (1.85-10.35)	**0.001**	40/34	4.12 (1.89-8.99)	**<0.0001**
Fermented fish intake						
Never	19/50	1.0	Ref.	29/61	1.25 (0.63-2.48)	0.602
Occasionally	5/12	1.10 (0.35-3.43)	1.00	15/16	2.47 (1.03-5.88)	0.067
Regular	26/27	2.53 (1.20-5.35)	**0.023**	29/23	3.32 (1.56-7.05)	**0.003**
Smoking						
Never	27/67	1.0	Ref.	36/76	1.18 (0.65-2.13)	0.650
Light	8/5	3.97 (1.24-12.69)	**0.027**	14/16	2.17 (0.94-5.0)	0.078
Heavy	15/17	2.19 (0.97-4.95)	0.082	23/8	7.13 (2.88-17.68)	**<0.0001**
Tobacco-betel quid chewing						
Never	13/31	1.0	Ref.	29/53	1.3 (0.6-2.85)	0.557
Light	17/41	0.99 (0.42-2.32)	1.00	20/27	1.77 (0.75-4.17)	0.275
Heavy	20/17	2.81 (1.14-6.93)	**0.040**	24/20	2.86 (1.20-6.82)	**0.030**

Ca/Co: Case/Control Ca *cases*, Co *controls*

* Fisher's exact test used to calculate *P* value and $P<0.05$ considered as statistically significance

Bold values indicate statistical significance ($P<0.05$)

4.2.3. Multifactorial dimensionality reduction analysis for metabolic gene-environment interactions

Table 4.2.6 summarized the best models in the MDR analysis to detect high-risk interaction of genetic and environmental factors for NPC. The MDR result suggest that smoking was the best predictive one-factor model with CVC 7/10 (TBA=0.5455; $P<0.0001$). When we consider two factors at a time, the combination of fermented fish and smoked meat was the best two-factors model with CVC of 10/10 (TBA=0.6474; $P<0.0001$), and the best three-factor model was the combination of fermented fish, smoked meat and smoking with CVC of 8/10 (TBA=0.624; $P<0.0001$). However the best model predictive among all the models was the four-factors model of *GSTT1* null genotype, fermented fish, smoked meat and smoking with maximum CVC of 10/10 and highest TBA (0.6802) and $P<0.0001$

Table 4.2.6 Summary of multifactorial dimensionality reduction analysis for NPC risk prediction

Model	**TrBA**	**TBA**	**CVC**	***P* value**
TBS	0.6084	0.5455	7/10	<0.0001
FFsh, SMT	0.6478	0.6478	10/10	<0.0001
FFsh, SMT, TBS	0.6734	0.624	8/10	<0.0001
GM1, FFsh, SMT, TBS[a]	**0.7063**	**0.6802**	**10/10**	**<0.0001**
GM1, GST1, FFsh, SMT, TBS	0.7316	0.6634	7/10	<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; [a]Best model prediction for NPC risk with highest TrBA, TBA and maximum CVC.

4.2.4 Interaction entropy graph of metabolic gene-environment interactions

We constructed the interaction entropy graphs for NPC risk to determine whether the observed interactions are synergistic or antagonist (not-synergistic) (Figure 4.2.4 and 4.2.5). Entropy graph shown, tobacco smoking had the highest independent effect (3.19%). Fermented fish and smoked meat also showed independent effect with percentage entropy of 2.69% and 2.00% respectively, and had a synergistic interaction with *GSTM1 null genotype* (1.34%) by removing 0.26% and 0.20% of entropy, respectively. *GSTT1 null genotype* (0.58%) showed a synergistic interaction with *GSTM1 null genotype* by removing 0.28% of entropy.

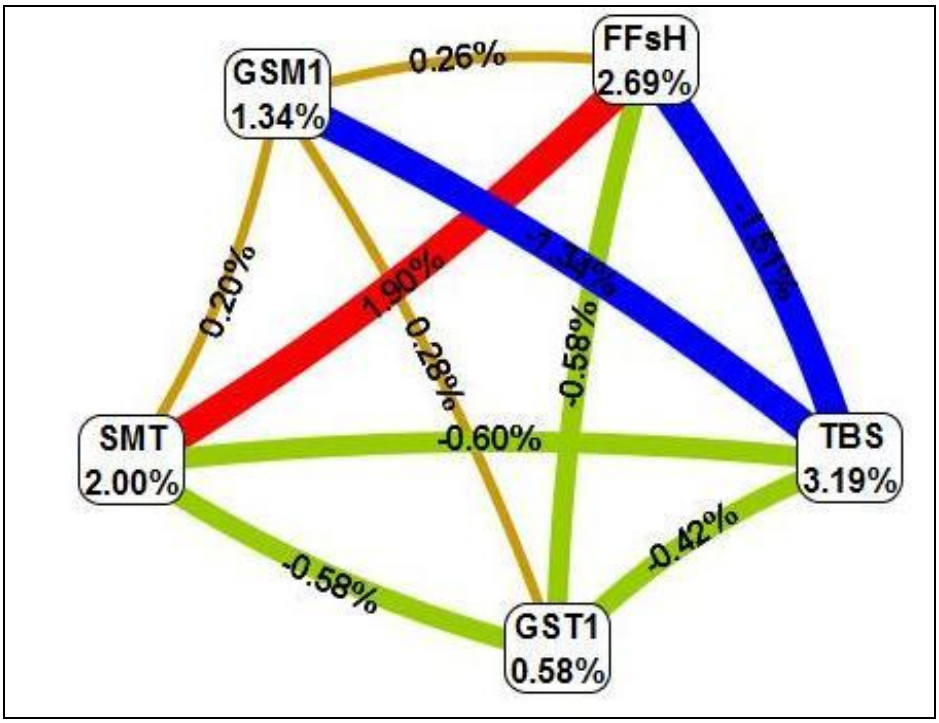


Figure 4.2.4 Interaction entropy graph for gene-environmental interaction in NPC. This graphical model explains the percentage of the entropy in case-control removed by each factor (independent effect) and by pair-wise combination of attributes (interaction effect). Positive percentage of entropy indicating synergistic interaction and negative values of entropy represents redundancy. Red-synergistic (high); Orange- synergistic (low); Gold- Intermediate; Blue- redundancy (high) and Green- redundancy (less).

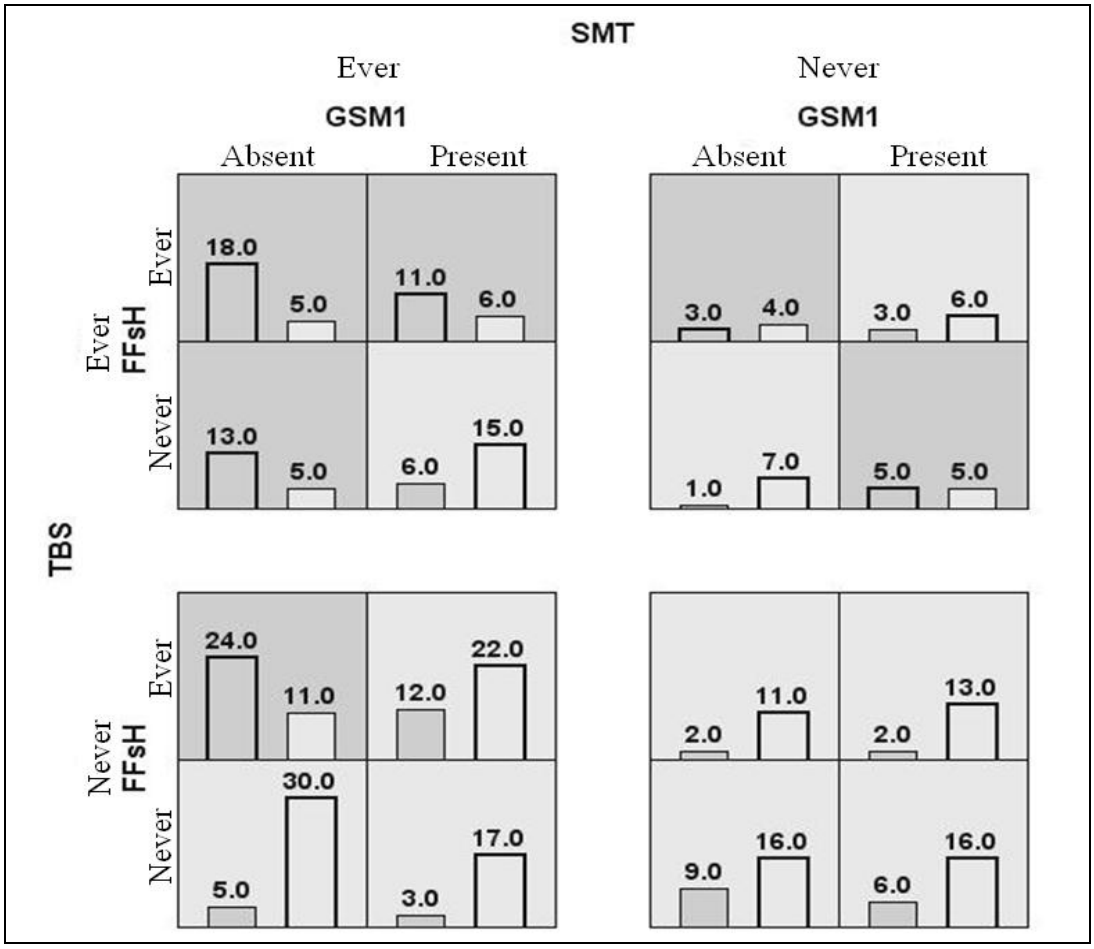


Figure 4.2.5 Summary of the four-factor model (GSTM1, fermented fish, smoked meat and smoking) in MDR analysis. The distribution of high risk (dark shading) and low risk (light shading) combinations associated with NPC risk. The percentage of patients having NPC was represented by left column in each box, whereas right column in each box indicated percentage of controls.

4.2.5. False positive report possibility (FPRP)

We strengthened our results by testing the robustness and consistency of the gene-gene and gene-environment interaction obtained from both LR and MDR using FPRP analysis. The FPRP values for all statistically significant result indicated that the interaction between *GSTT1*, *GSTM1* and *CYP1A1* was noteworthy for low prior probability assumptions upto 0.25 when detecting ORs of 1.5 and 2.0 for an FPRP value of 0.5 (Table 4.2.7). Moreover, the association was also deserving of attention for subject with regular consumption of smoked meat and fermented fish, heavy smoking and tobacco chewing among individual carrying defective *GSTT1, GSTM1* and *CYP1A1* genes (Table 4.2.8). The best predictive models selected by MDR analysis were noteworthy for very low prior probability assumptions (upto 0.1 to 0.001) when detecting ORs of 1.5 and 2.0 for an FPRP value of 0.5 (Table 4.2.9). The relatively greater FPRP values with very low prior probability assumptions (0.001) might be ascribed to the relative small sample size of this study as well as moderate effects of selected SNP. These findings need further validation in investigations with large sample size.

Table 4.2.7 False Positive Reports Probability (FPRP) for odd ratios of the Logistic Regression (LR) analysis in gene-gene interaction

Variables	Odds ratio	OR= 1.5 (Prior Probability)				OR = 2.0 (Prior Probability)			
	**OR (95% CI)	0.25	0.1	0.01	0.001	0.25	0.1	0.01	0.001
GSTT1/GSTM1/CYP1A1 T3801C									
M1 –ve	2.76 (1.61-4.71)	**0.045**	**0.123**	0.606	0.939	**0.005**	**0.015**	**0.141**	0.623
M1(-/-) and T1 (+/+)	1.95 (1.06-3.59)	**0.324**	0.590	0.941	0.994	**0.153**	**0.351**	0.856	0.984
M1 (-/-) and T1 (-/-)	3.77 (1.95-7.3)	**0.073**	**0.192**	0.723	0.963	**0.008**	**0.024**	**0.214**	0.734
M1 (-/-) and TT	2.39 (1.17-4.85)	**0.325**	0.591	0.941	0.994	**0.132**	**0.314**	0.834	0.981
M1 (-/-) and TT or CC	3.22 (1.65-6.28)	**0.126**	**0.302**	0.826	0.980	**0.022**	**0.062**	**0.423**	0.881
M1(-/-), T1 (+/+) and TT or CC	2.88 (1.61-7.18)	**0.463**	0.721	0.996	0.997	**0.243**	**0.491**	0.914	0.991
M1 (-/-), T1 (-/-) and TT	3.77 (1.36-10.45)	**0.457**	0.717	0.965	0.996	**0.224**	**0.464**	0.905	0.990
M1 (-/-), T1 (-/-) and TT or CC	5.71 (2.11-15.45)	**0.299**	0.561	0.934	0.993	**0.085**	**0.218**	0.754	0.969

Prior probability range = 0.25–0.001 to detect OR = 1.5 or 2.0; α level = observed *P* value; Bold value= noteworthy association at 0.5 FPRP; ***P*<0.05

FFsh; Fermented Fish, SMT; Smoked meat, TBS; Tobacco smoking, TBC; Tobacco chewing

Table 4.2.8 False Positive Reports Probability (FPRP) for odd ratios of the Logistic Regression (LR) analysis in gene-environmental interaction

Variables	Odds ratio	OR= 1.5 (Prior Probability)				OR = 2.0 (Prior Probability)			
	**OR (95% CI)	0.25	0.1	0.01	0.001	0.25	0.1	0.01	0.001
GSTM1 (-ve) vs. GSTM1 (+ve)									
Occasional SMT	3.55 (1.50-8.41)	**0.322**	0.588	0.940	0.994	**0.111**	**0.272**	0.804	0.976
Regular SMT	5.56 (2.91-10.62)	**0.017**	**0.048**	**0.357**	0.849	**0.001**	**0.002**	**0.020**	**0.172**
Occasional FFsh	6.23 (2.47-15.82)	**0.207**	**0.439**	0.896	0.989	**0.041**	**0.113**	0.583	0.934
Regular FFsh	5.73 (2.66-12.34)	**0.074**	**0.193**	0.724	0.964	**0.007**	**0.020**	**0.185**	0.696
Light TBS	7.84 (2.80-21.99)	**0.246**	**0.495**	0.915	0.991	**0.055**	**0.148**	0.656	0.951
Heavy TBS	12.67 (4.95-32.39)	**0.076**	**0.197**	0.730	0.965	**0.006**	**0.017**	**0.164**	0.664
Never TBC	2.29 (1.07-4.88)	**0.412**	0.667	0.958	0.996	**0.208**	**0.441**	0.897	0.989
Light TBC	2.81 (1.29-6.12)	**0.328**	0.594	0.942	0.994	**0.124**	**0.229**	0.824	0.979
Heavy TBC	5.68 (2.46-13.08)	**0.133**	**0.315**	0.835	0.981	**0.019**	**0.054**	**0.385**	0.863
GSTT1 (-ve) vs. GSTT1 (+ve)									
Regular SMT	3.99 (1.84-8.68)	**0.176**	**0.390**	0.875	0.986	**0.034**	**0.096**	0.540	0.992
Regular FFsh	3.50 (1.73-7.09)	**0.140**	**0.328**	0.843	0.982	**0.025**	**0.070**	**0.454**	0.893
Light TBS	4.60 (1.29-16.4)	0.571	0.800	0.978	0.998	**0.360**	0.627	0.949	0.995
Heavy TBS	4.46 (1.89-10.56)	**0.234**	**0.479**	0.910	0.990	**0.056**	**0.151**	0.662	0.952
Heavy TBC	3.51 (1.58-7.82)	**0.254**	0.505	0.918	0.991	**0.070**	**0.185**	0.714	0.962
CYP1A1 TC+CC vs. TT									
Regular SMT	4.12 (1.89-8.99)	**0.168**	**0.378**	0.870	0.985	**0.031**	**0.089**	0.517	0.915
Regular FFsh	3.32 (1.56-7.05)	**0.217**	**0.455**	0.902	0.989	**0.054**	**0.147**	0.654	0.950
Heavy TBS	7.13 (2.88-17.68)	**0.149**	**0.345**	0.853	0.983	**0.022**	**0.062**	**0.422**	0.880
Heavy TBC	2.86 (1.20-6.82)	**0.423**	0.688	0.960	0.996	**0.203**	**0.433**	0.893	0.988

Prior probability range = 0.25–0.001 to detect OR = 1.5 or 2.0; α level = observed *P* value; Bold value= noteworthy association at 0.5 FPRP; ***P*<0.05

FFsh; Fermented Fish, SMT; Smoked meat, TBS; Tobacco smoking, TBC; Tobacco chewing

Table 4.2.9 False Positive Reports Probability (FPRP) for odd ratios of Multifactor Dimensionality Reduction (MD) analysis in gene-environmental interaction

Variables	Odds ratio	OR= 1.5 (Prior Probability)				OR = 2.0 (Prior Probability)			
	**OR (95% CI)	0.25	0.1	0.01	0.001	0.25	0.1	0.01	0.001
MDR analysis									
TBS	2.47 (1.51-4.06)	**0.042**	**0.117**	0.593	0.936	**0.005**	**0.016**	**0.150**	0.641
FFsh, SMT	3.69 (2.26-6.02)	**0.003**	**0.010**	**0.098**	0.522	**<0.0001**	**<0.0001**	**0.002**	**0.024**
FFsh, SMT, TBS	4.20 (2.59-6.83)	**0.001**	**0.004**	**0.042**	**0.304**	**<0.0001**	**<0.0001**	**0.001**	**0.005**
GM1, FFsh, SMT, TBS	6.41 (3.8-10.70)	**<0.0001**	**0.001**	**0.008**	0.079	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
GM1, GST1, FFsh, SMT, TBS	7.30 (4.38-12.15)	**<0.0001**	**<0.0001**	**0.004**	**0.034**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**

Prior probability range = 0.25–0.001 to detect OR = 1.5 or 2.0; α level = observed *P* value; Bold value= noteworthy association at 0.5 FPRP; ***P*<0.05

FFsh; Fermented Fish, SMT; Smoked meat, TBS; Tobacco smoking, TBC; Tobacco chewing

4.3. Polymorphism in DNA repair genes, their interaction with environmental factors in NPC

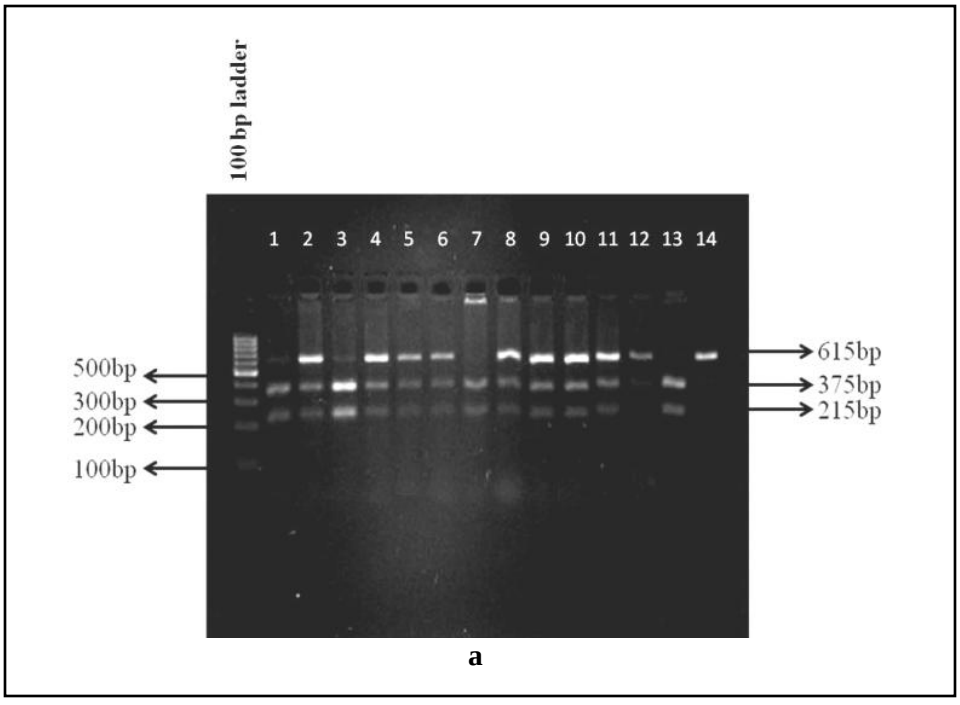
The purpose of this study was to investigate the effect of polymorphisms in *XRCC1* (Arg399Gln) and *XRCC2* (Arg188His) genes in NPC and their role in modulating the relationship between tobacco (smoking and betel-quid chewing) and dietary (smoked meat and fermented fish intake) habits, and NPC risk. Here we analysed 100 cases, 90 FDRs and 120 controls after determining the quality and quantity of DNA in the samples. Furthermore, we also used multifactorial dimensionality reduction (MDR) approach to investigate the high-degree gene-environmental interaction in NPC carcinogenesis in the northeast Indian population. False-positive report probability (FPRP) analysis was also performed to validate all significant findings.

4.3.1. Polymorphisms in DNA repair genes (*XRCC1* and *XRCC2*) and the risk associated with NPC

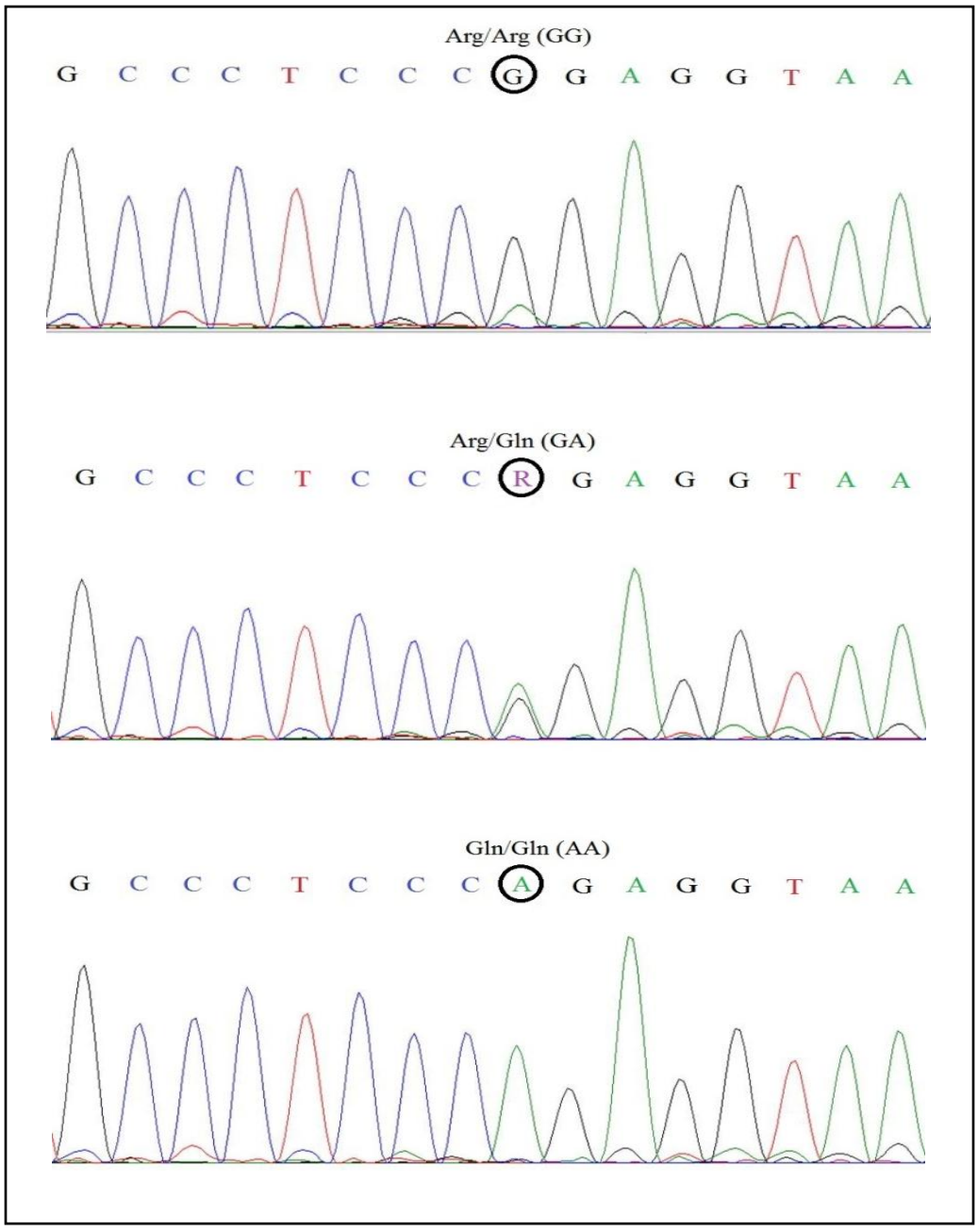
The genotypes of *XRCC1* Arg399Gln and *XRCC2* Arg188His were determined by detecting the PCR-RFLP band pattern on 1.5% agarose gel. Furthermore, the PCR products of *XRCC1* and *XRCC2* genes were sequenced to confirm the RFLP results (Figure 4.3.2 and 4.3.3). The frequency distributions of *XRCC1* Arg399Gln 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 85% and 15% 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 4.3.1). 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* 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 4.3.1).



a



b

Figure 4.3.1 PCR based detection of *XRCC1* Arg399Gln 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 results showing nucleotide variation (in black circle). R indicates heterozygous genotypes, where green peak denotes adenine (A) while black peak denotes guanine (G).

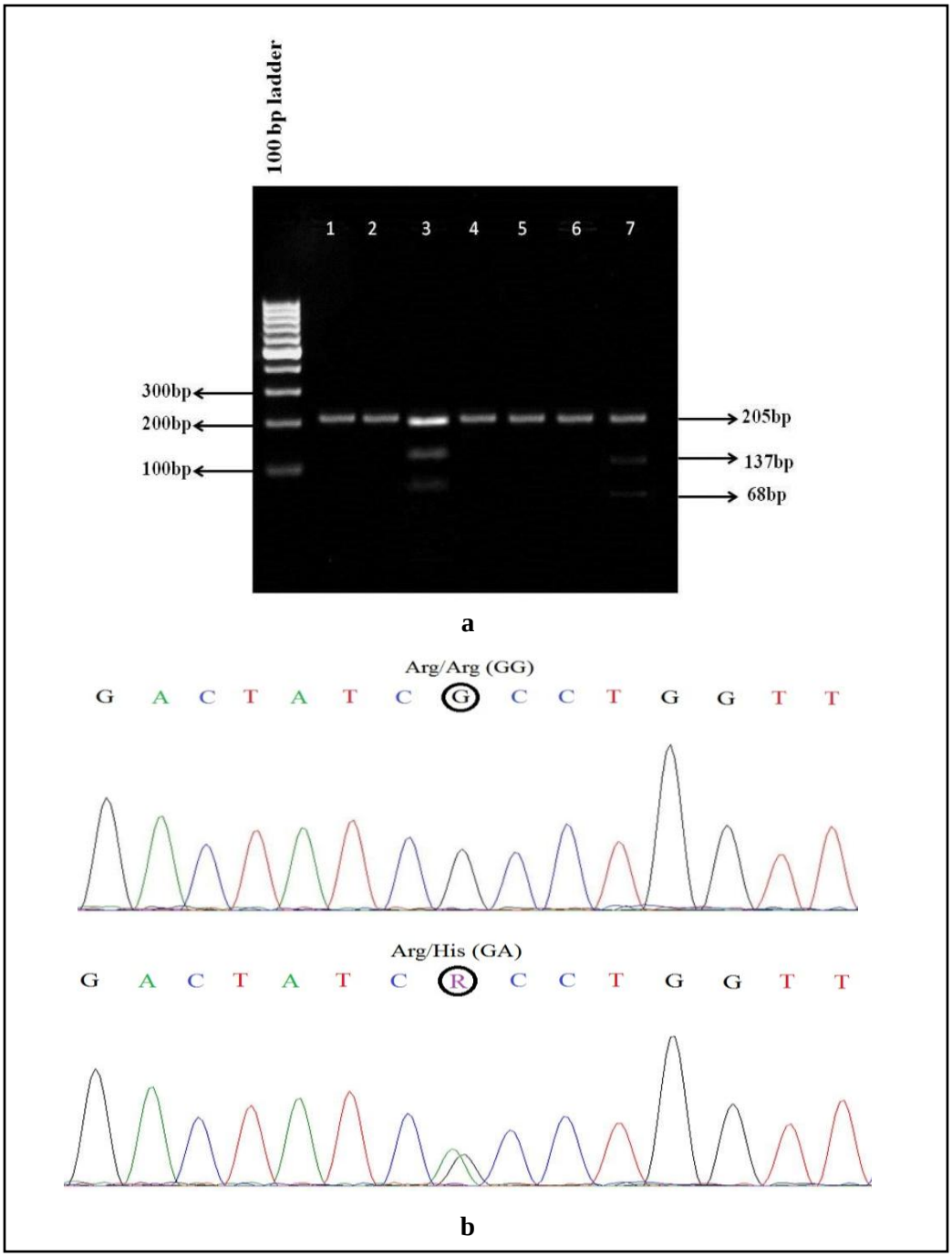


Figure 4.3.2 PCR based detection of *XRCC2* Arg188His gene polymorphism in NPC **a** PCR-RFLP patterns of *XRCC2* Arg188His gene polymorphism separated by agarose gel electrophoresis. Lane 1, 2, 4, 5, and 6 represents wild Arg/Arg (G/G) genotype and lane 3 and 7 represents heterozygous Arg/His (G/A) genotype; **b** DNA sequencing results showing nucleotide variation (in black circle). R indicates heterozygous genotypes, where green peak denotes adenine (A) while black peak denotes guanine (G).

Table 4.3.1 Genotype frequency distribution of XRCC1 Arg399Gln and XRCC2 Arg188His polymorphisms and risk of NPC

Variables	**Case, *n*=100 (%)**	**Controls, *n*=120 (%)**	**OR (95% CI)***	***P* values**	**First-degree relatives, *n*=90 (%)**	**OR (95% CI)***	***P* values**
***XRCCI* Arg399Gln (rs25487)**							
GG (Arg/Arg)	33 (33)	56 (46.7)	1.0	Ref.	39 (43.3)	1.0	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)	0.530
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	1.0	Ref.	116	1.0	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)	85 (85)	107 (89.2)	1.0	Ref.	76 (84.5)	1.0	Ref.
GA(Arg/His)	15 (15)	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	1.0	Ref.	165	1.0	Ref.
A (His) allele frequency	15	13	1.42 (0.67-3.01)^^	0.435	13	1.38 (0.63-3.00)^^	0.540
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 ($P< 0.05$)

The interactions of *XRCC1* Arg399Gln and *XRCC2* Arg188His genotypes and risk of NPC were also analysed among the cases (Table 4.3.2). 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.

Table 4.3.2 Combined genotype analysis of *XRCC1* Arg399Gln and *XRCC2* Arg188His on risk of NPC

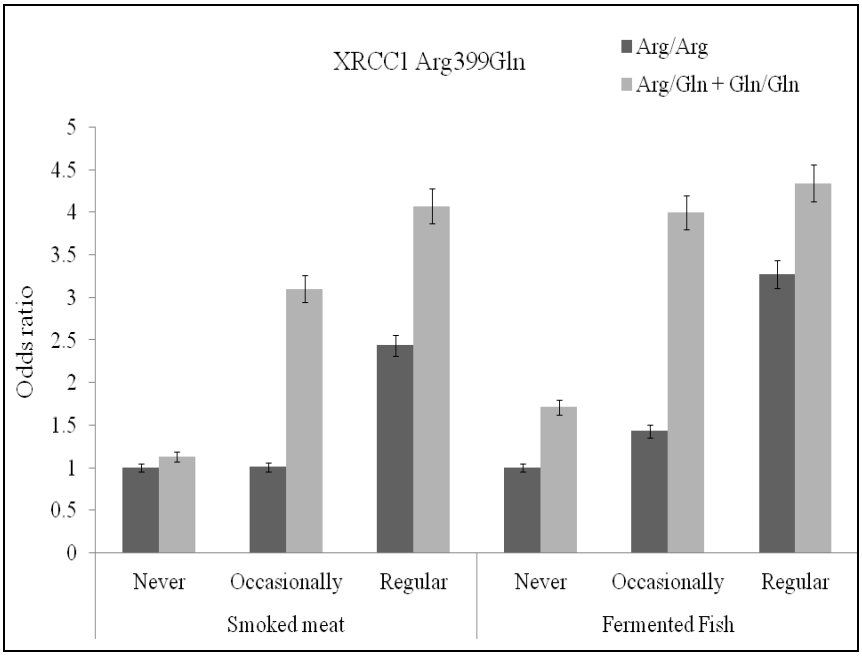
***XRCCI* Arg399Gln (rs25487)**	***XRCC2* Arg188His (rs3218536)**	**Case (*n*=100)**	**Controls (*n*=120)**	**OR (95% CI)**	***P* values**
GG (Arg/Arg)	GG(Arg/Arg)	29	50	1.0	Ref.
GG (Arg/Arg)	GA+AA (Arg/His + His/His)	4	6	1.15 (0.32-4.15)	1.000
GA+AA (Arg/Gln + Gln/Gln)	GG(Arg/Arg)	54	57	1.63 (0.91-2.94)	0.106
GA+AA (Arg/Gln + Gln/Gln)	GA+AA (Arg/His + His/His)	13	7	**3.20 (1.17-8.74)**	**0.041**

Ca/Co: Case/Control Ca *cases*, Co *controls*; *P*< 0.05 is considered statistically significant

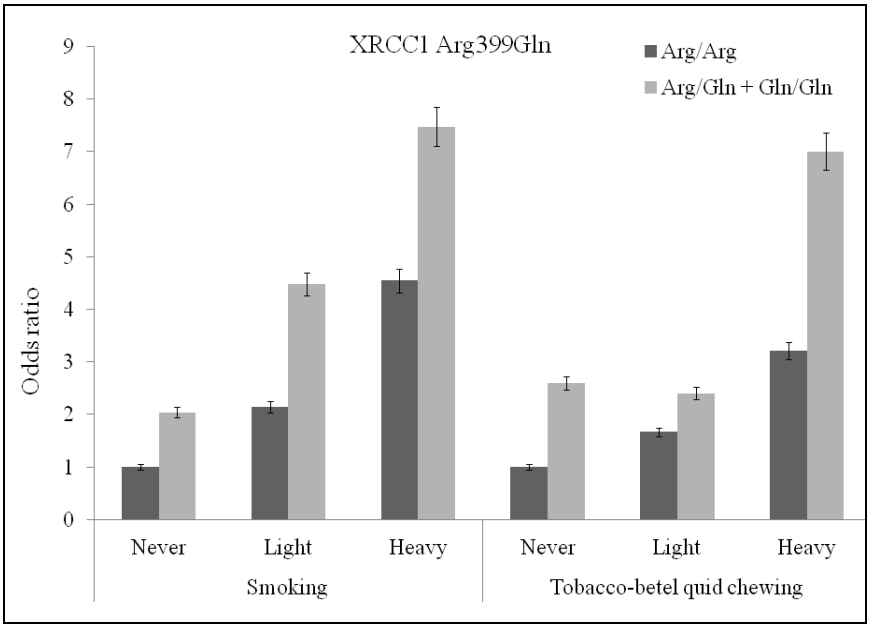
Bold values indicate statistical significance (*P*< 0.05)

4.3.2. Interaction of DNA repair genes and environmental risk factors in NPC progression

The potential interaction between *XRCC1* Arg399Gln genotypes and environment factors in NPC risk were also investigated (Table 4.3.3; Figure 4.3.3). 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.



a



b

Figure 4.3.3 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 of developing NPC, respectively; **b** Heavy smokers and chewers carrying *XRCC1* Gln/Gln genotype had 7.47 and 7 fold increase risk NPC, respectively.

Table 4.3.3 Association of XRCC1 Arg399Gln genotype, stratified by smoked meat, fermented fish, smoking, and tobacco-betel quid habits

Variables	*XRCCI* Arg399Gln GG (Arg/Arg)			*XRCCI* Arg399Gln G/A+AA (Arg/Gln + Gln/Gln)		
	Ca /Co	**OR (95% CI)^^**	***P* value**	**Ca /Co**	**OR (95% CI)^^**	***P* value**
Smoked Meat						
Never	9/22	1.0	Ref.	13/28	1.13 (0.42-3.09)	1.0
Occasionally	7/17	1.01 (0.32-3.18)	1.000	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	1.0	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)	**0.010**
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	1.0	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	1.0	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 Ca *cases*, Co *controls*

^^Fisher's exact test used to calculate *P* value and $P < 0.05$ is considered statistically significant

Bold values indicate statistical significance ($P < 0.05$)

4.3.3. Multifactorial dimensionality reduction analysis for repair gene-environment interaction

MDR analysis was used to determine the best-model interaction of genetic and environmental factors for NPC (Table 4.3.4). 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 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).

Table 4.3.4 Summary of multifactorial dimensionality reduction analysis 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
XRC1, FFsh, SMT, TBS	0.689	0.616	8/10	<0.0001
XRC1, 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

[a]Best model prediction for NPC risk with highest TrBA, TBA and maximum CVC.

4.3.4 Interaction entropy graph of repair gene-environment interactions

Interaction entropy graph was constructed using MDR results to determine synergistic or antagonistic (not-synergistic) interactions between the genetic and environmental factors in NPC risk (Figure 4.3.4 and 4.3.5). Entropy graph shown, tobacco smoking, and fermented fish had 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.

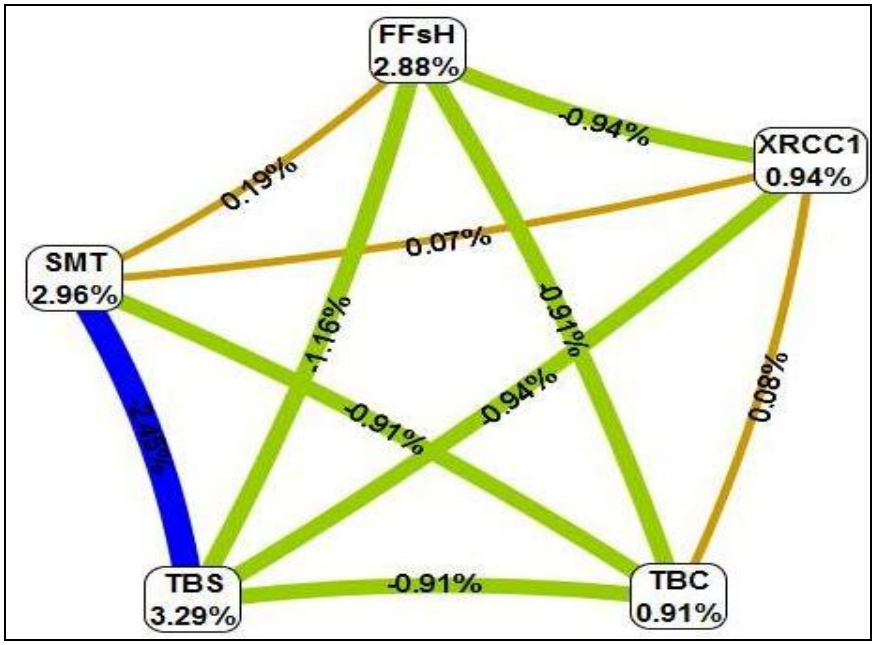


Figure 4.3.4 MDR analysis showing entropy graph for gene-environment interaction and NPC risk. This graphical model explains the percent of the entropy in case-control removed by each factor (independent effect) and by each pair-wise combination of attributes (interaction effect). Positive percentage of entropy indicating synergistic interaction and negative values of entropy represent redundancy. The red color 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.

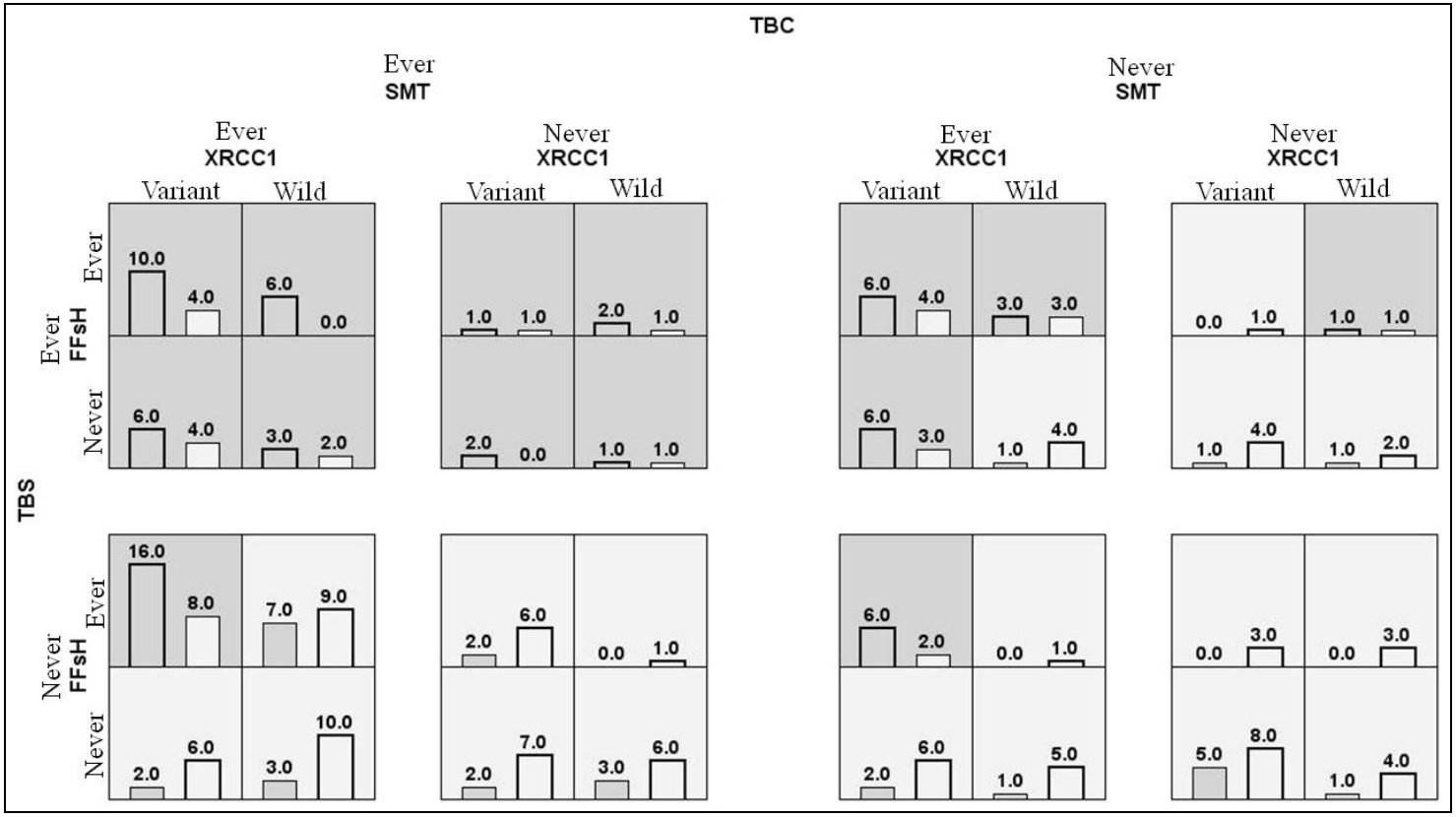


Figure 4.3.5 Summary of the five-factor model (*XRCC1*, fermented fish, smoked meat, smoking and chewing) in MDR analysis; the distribution of high risk (dark shading) and low risk (light shading) combinations associated with NPC risk. The percentage of patients having NPC was represented by left column in each box, whereas right column in each box indicated percentage of controls. **TBS**-tobacco smoking, **SMT**-smoked meat, **FFsh**-fermented fish.

4.3.5 False positive report possibility (FPRP)

We strengthened our data by testing the robustness and consistency of the gene-environment interaction obtained from both LR and MDR using FPRP analysis. The FPRP values for all statistically significant result are summarized in Table 4.3.5

. FPRP analysis indicated that the significant association between *XRCC1* (GA+AA) genotype and NPC risk was noteworthy for low prior probability assumptions (upto 0.1) for detecting OR= 2.0 for an FPRP value of 0.5. Moreover, the association was also deserving of attention for subject with regular consumption of smoked meat and fermented fish, heavy smoking and tobacco chewing among individual carrying defective *XRCC1* Arg399Gln. The best predictive models selected by MDR analysis were noteworthy for very low prior probability assumptions (upto 0.1 to 0.001) when detecting ORs of 1.5 and 2.0 for an FPRP value of 0.5. Relatively greater FPRP values with very low prior probability assumptions (0.001) might be ascribed to the relative small sample size of this study as well as moderate effects of selected SNP. These findings need further validation in investigations with large sample size.

Table 4.3.5 False Positive Reports Probability (FPRP) for odd ratios of the Logistic Regression (LR) and Multifactor Dimensionality Reduction (MDR) analysis

Variables	Odds ratio	OR= 1.5 (Prior Probability)				OR = 2.0 (Prior Probability)			
	OR (95% CI) *P*-value	0.25	0.1	0.01	0.001	0.25	0.1	0.01	0.001
LR analysis									
***XRCC1* Arg399Gln (rs25487)**									
AA vs. GG	2.76 (1.14-6.66) 0.024	**0.450**	0.711	0.964	0.996	**0.232**	**0.476**	0.909	0.990
GA+AA vs. GG	2.03 (1.13-3.62) 0.017	**0.244**	**0.492**	0.914	0.991	**0.093**	**0.236**	0.772	0.972
A allele vs. G allele	1.59 (1.08-2.35) 0.022	**0.135**	**0.318**	0.837	0.981	**0.064**	**0.171**	0.693	0.958
GA+AA vs. GG									
Regular SMT	4.07 (1.60-10.35) 0.004	**0.348**	0.612	0.946	0.994	**0.124**	**0.298**	0.824	0.979
Regular FFsh	4.34 (1.83-10.30) 0.001	**0.247**	**0.496**	0.915	0.991	**0.062**	**0.166**	0.686	0.957
Heavy TBS	7.47 (2.52-22.14) <0.0001	**0.312**	0.577	0.937	0.993	**0.090**	**0.228**	0.765	0.970
Heavy TBC	7.00 (2.33-21.03) 0.001	**0.343**	0.610	0.945	0.994	**0.110**	**0.270**	0.803	0.976
MDR analysis									
TBS	2.43 (1.39-4.23) 0.0016	**0.105**	**0.261**	0.795	0.975	**0.021**	**0.060**	**0.412**	0.876
FFsh, SMT	3.37 (1.91-5.94) <0.0001	**0.030**	**0.086**	0.508	0.912	**0.002**	**0.007**	**0.069**	**0.429**
FFsh, SMT, TBS	4.38 (2.37-8.07) <0.0001	**0.022**	**0.063**	**0.424**	0.881	**0.001**	**0.003**	**0.035**	**0.269**
XRCC1, FFsh, SMT, TBS	4.89 (2.74-7.22) <0.0001	**0.007**	**0.021**	**0.190**	0.703	**<0.001**	**0.001**	**0.006**	**0.056**
XRCC1, FFsh, SMT, TBS, TBC	5.63 (3.14-10.06) <0.0001	**0.004**	**0.012**	**0.119**	**0.576**	**<0.001**	**<0.001**	**0.002**	**0.022**

Prior probability range = 0.25–0.001 to detect OR = 1.5 or 2.0; α level = observed *P* value; Bold value= noteworthy association at 0.5 FPRP

FFsh; Fermented Fish, **SMT**; Smoked meat, **TBS**; Tobacco smoking, **TBC**; Tobacco chewing

4.4. Polymorphism in newly identified susceptible loci (*MECOM, TNFRSF19* and *CDKN2B-AS1*) and risk of NPC in northeast Indian population

Recent, GWAS studies had reported six new variants which were a strongly associated with NPC risk. This study investigate the effect of polymorphisms in selected three genes; *MECOM* (rs6774494), *TNFRSF19* (rs9510787), and *CDKN2B-AS1* (rs1412829) and the risk associated with NPC. Here we analysed 100 cases and 100 controls after determining the quality and quantity of DNA in the samples. We further performed multifactorial dimensionality reduction (MDR) approach to investigate high-degree gene-gene interactions of *GSTT1*, *GSTM1*, *CYP1A1*, *XRCC1*, *MECOM*, and *TNFRSF19* genes in NPC carcinogenesis in the northeast Indian population. False-positive report probability (FPRP) analysis was also performed to validate any significant findings.

4.4.1. Genetic variants of *MECOM, TNFRSF19* and *CDKN2B-AS1* genes and NPC predisposition

The genotypes of *MECOM, TNFRSF19* and *CDKN2B-AS1 genes* were determined by multiplex-PCR amplification (Figure 4.4.1) and Sanger sequencing (Figure 4.4.2, 4.4.3 and 4.4.4) of the PCR products.

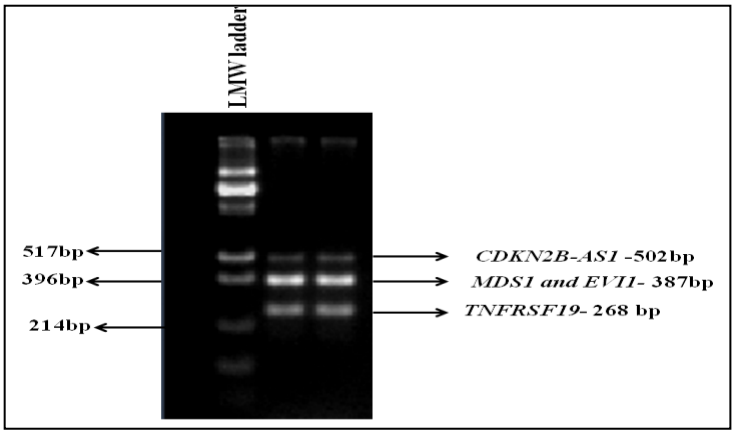


Figure 4.4.1 Multiplex PCR amplification of *MECOM, TNFRSF19* and *CDKN2B-AS1* genes

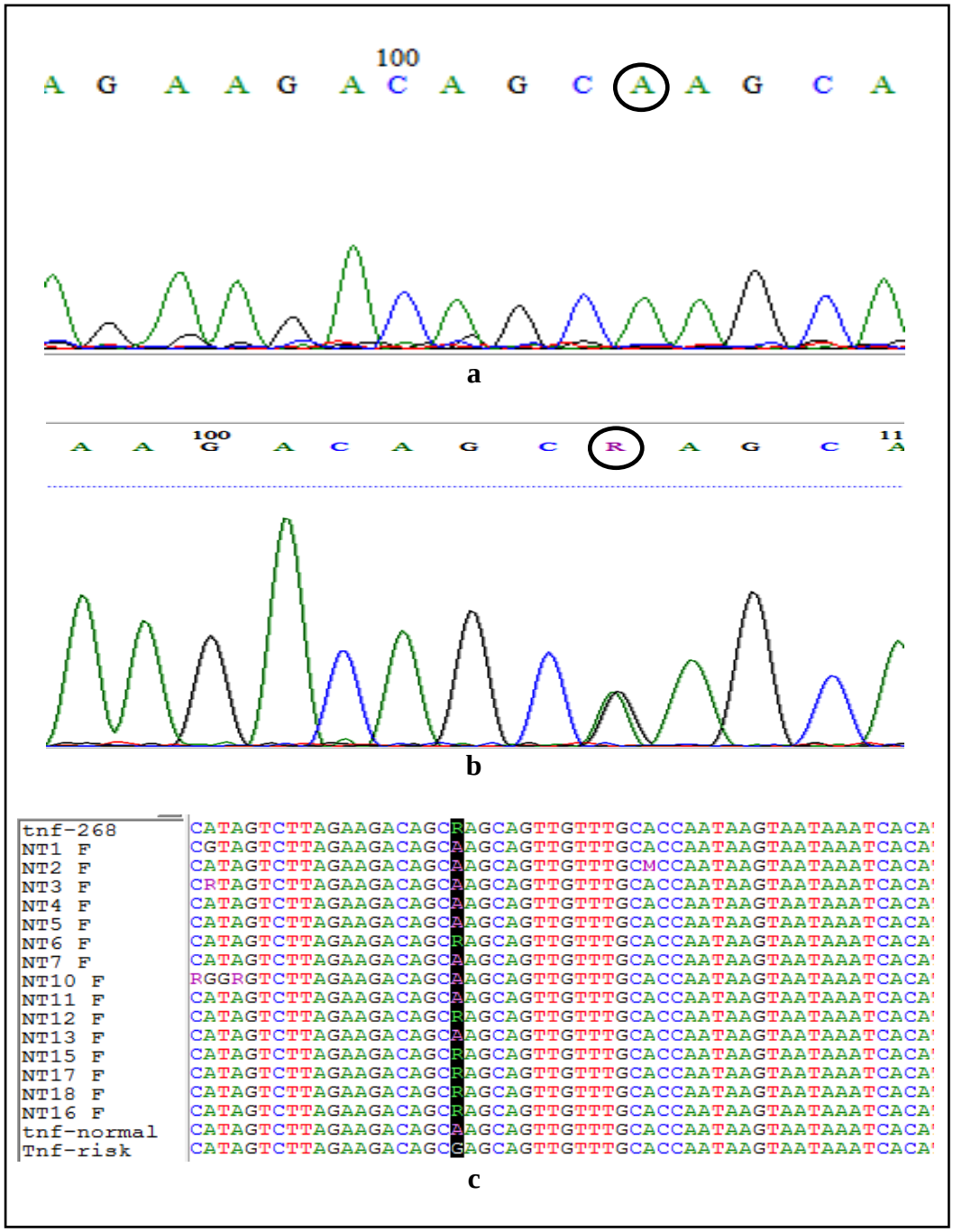


Figure 4.4.2 DNA sequencing result of *TNFRSF19* (rs9510787) gene **a** Chromatogram of the wild type genotype (A allele) **b** heterozygous variant (R allele); and **c** multiple sequence alignment of *TNFRSF19* (rs9510787) gene sequences. Highlighted region (BLACK) shows rs9510787 in *TNFRSF19*. Query sequences (NT *numerical value* F) is aligned with subject sequence (Tnf-268), where A= Reference allele and G= Risk allele, R denotes A/G a heterozygous condition

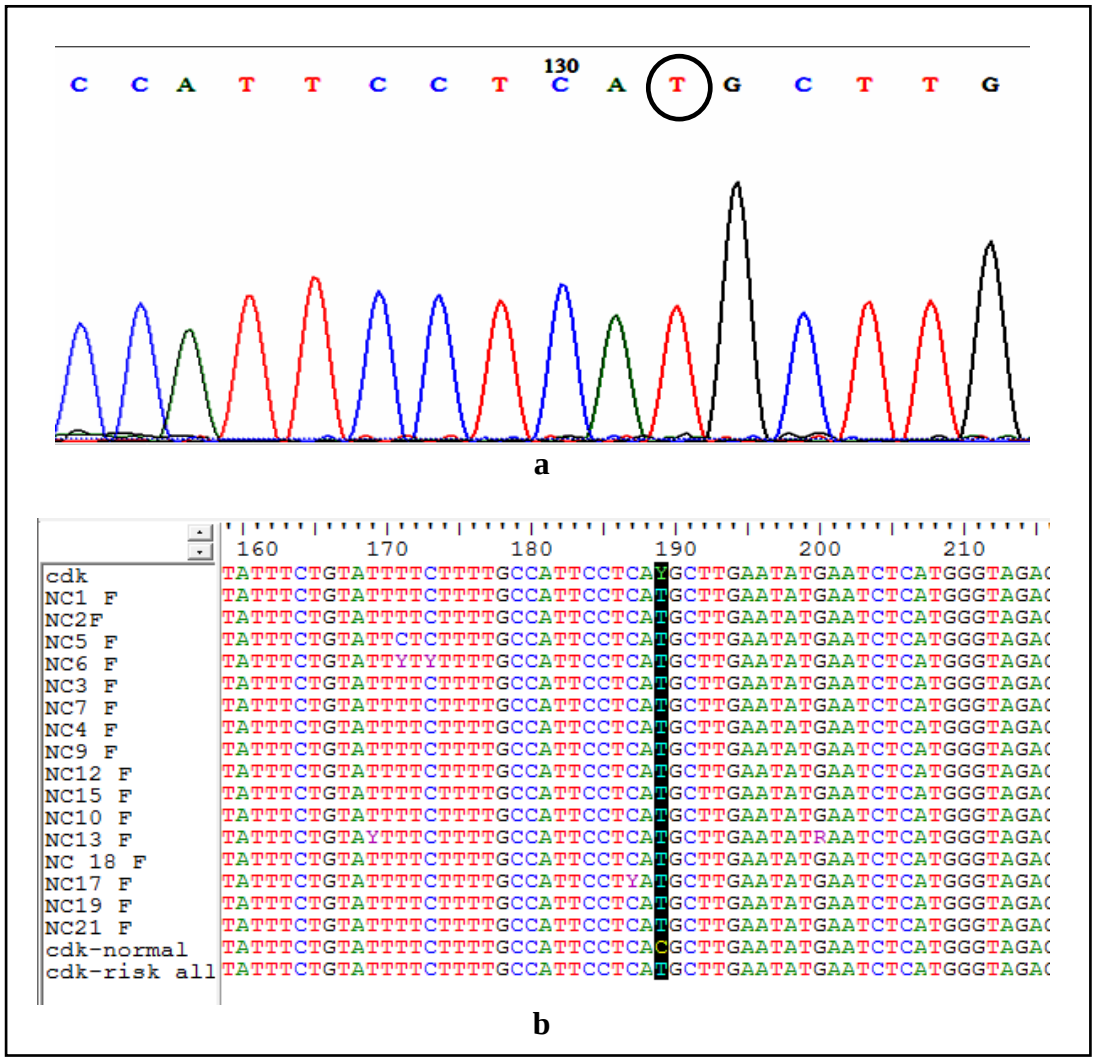


Figure 4.4.3 DNA sequencing result of *CDKN2B-AS1* (rs1412829) gene **a** Chromatogram representing the A allele; and **b** multiple sequence alignment of *CDKN2B-AS1* (rs1412829) sequences showed the presence of single T allele in our study population. Highlighted region (BLACK) shows rs1412829 in *CDKN2B-AS1*. Query sequences (NC *numerical value* F) is aligned with subject sequence (cdk), where C= Reference allele and T= Risk allele, Y denotes C/T a heterozygous condition

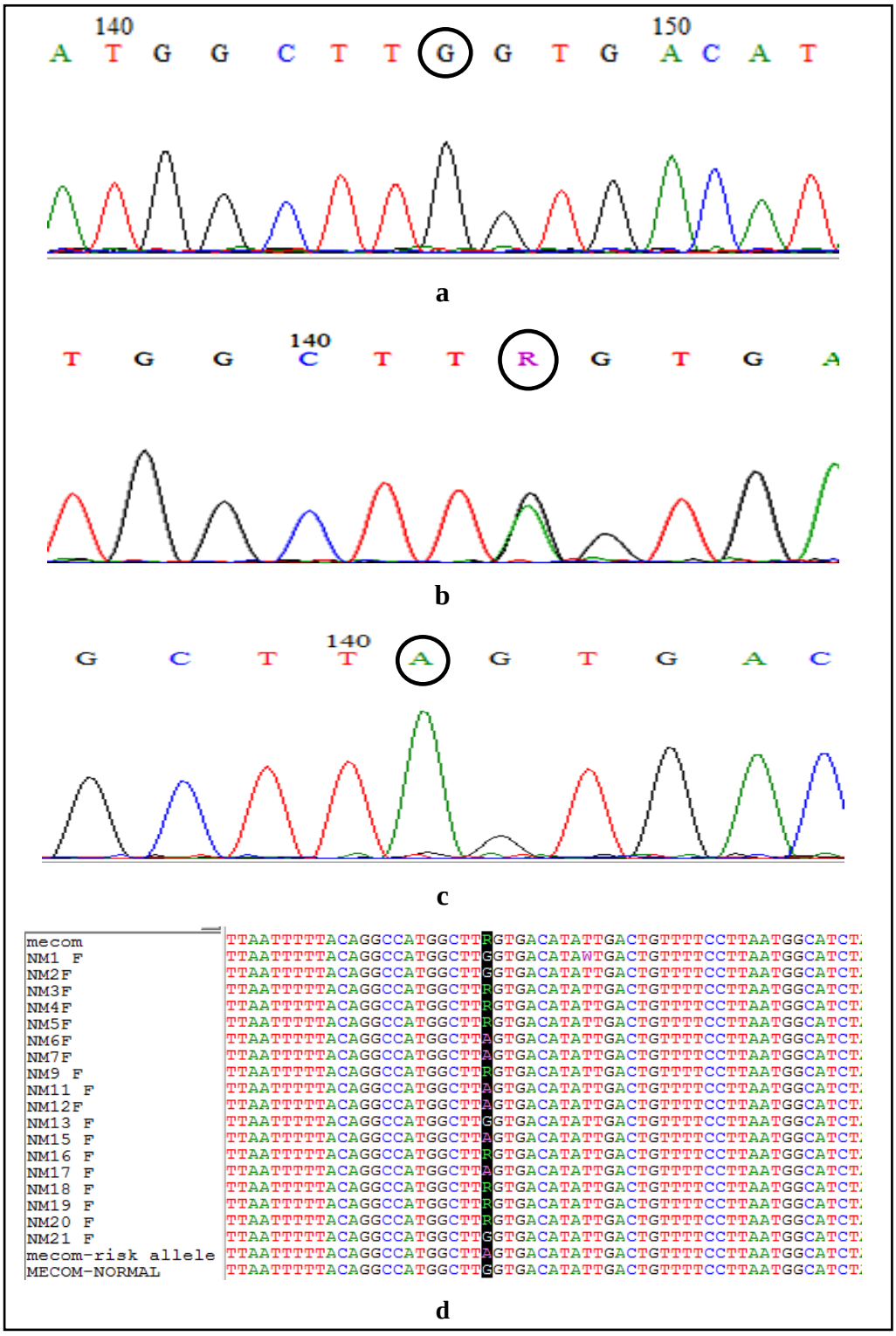


Figure 4.4.4. DNA sequencing result of *MECOM* (rs6774494) gene **a** Chromatogram of the wild type genotype (G allele) **b** heterozygous variant (R allele) **c** homozygous variant (A allele); and **d** multiple sequence alignment of *MECOM* (rs6774494) gene sequences. Highlighted region (BLACK) shows rs6774494 in *MECOM*. Query sequences (NM *numerical value* F) is aligned with subject sequence (mecom), where G= Reference allele and A= Risk allele, R denotes A/G a heterozygous condition.

The percentage of variant genotypes (both heterozygous and homozygous) of the three genes in the case and control samples were determined (Table 4.4.1). Logistic regression analysis revealed that individual carrying the variant genotype of *MECOM* (rs6774494) gene had 2.1 fold risk (95% CI: 1.06-4.14; *P*=0.032) of NPC compared to those with the normal genotype. Similarly, carries of variant genotype of TNFRSF19 (rs9510790) gene had 1.8 fold marginal risk (95% CI: 0.93-3.72; *P*=0.047) of NPC development. However, we did not find any association with *CDNK2B-AS1* (rs1412829) gene in NPC. Genotyping result as determine by Sanger sequencing were further validated by single base extension method using SNaPshot multiplex system (Figure 4.4.5)

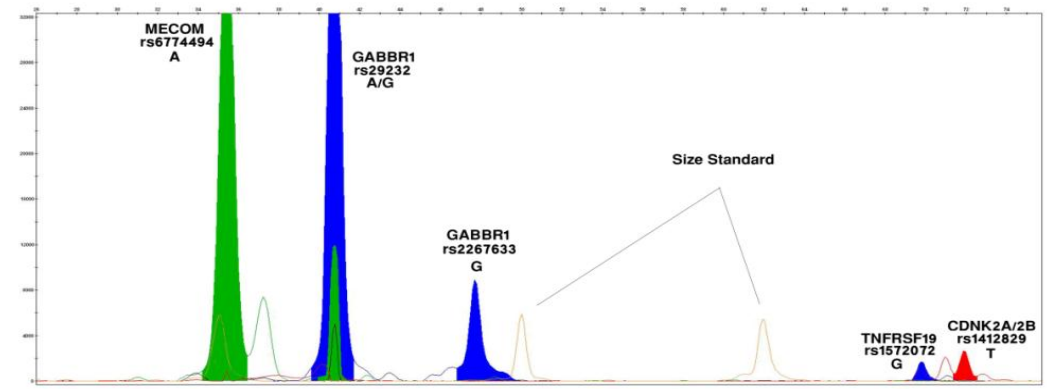
Table 4.4.1 Genetic variants of *MECOM*, *CDNK2B-AS1* and *TNFRSF19* genes and NPC predisposition

Variable	MA	Allele	Variant genotypes		OR (95% CI)[b]	*P* value
			Cases (%)	Controls (%)		
MECOM (rs6774494)	G	A/G	72 (72)	60 (69)	2.1 (1.06-4.14)	**0.032**
CDNK2B-AS1 (rs1412829)	G	T/C	100 (100)	100 (100)	-	-
TNFRSF19 (rs9510790)	G	G/A	39 (39)	30 (30)	1.8 (0.93-3.72)	**0.047**

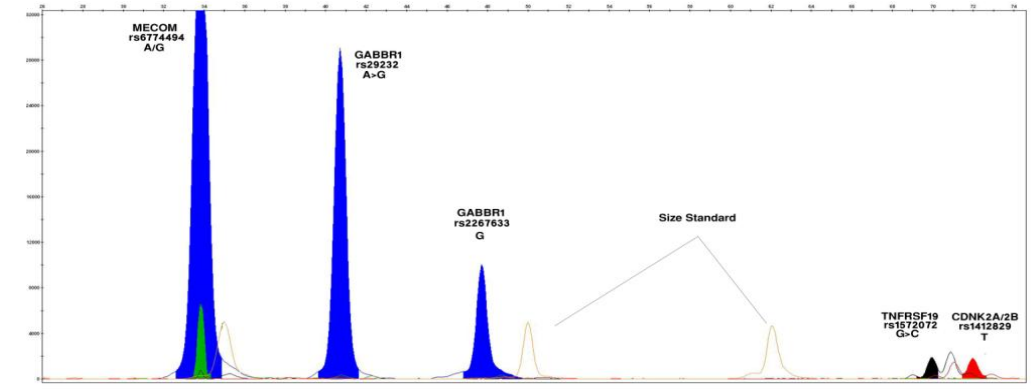
[a] Risk allele/ Reference allele; [b]Adjusted for age, sex, ethnicity, BMI and profession

^^Fisher's exact test used to calculate *P* value and *P*< 0.05 is considered statistically significant

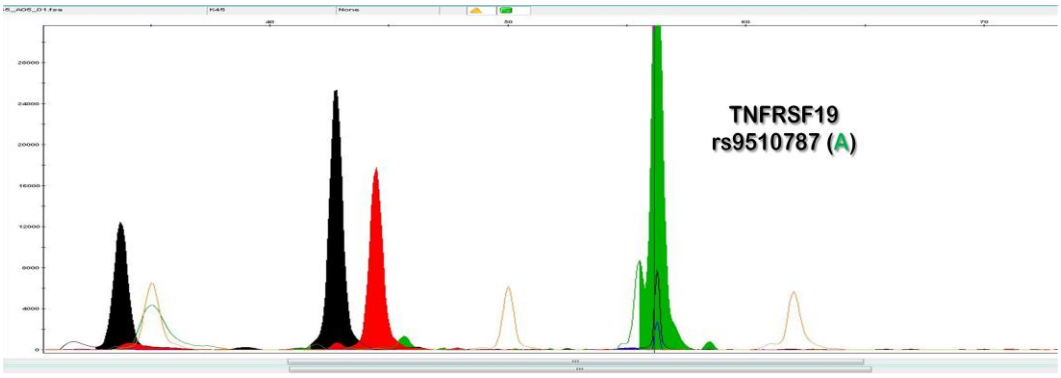
Bold values indicate statistical significance (*P*< 0.05)



a



b



c

Figure 4.4.5 Validation of sequencing results by SNapShot singel base extension method. SNapShot result confirming **a** A allele of *MECOM* and T allele of *CDNK2A/2B* genes**; b** R (A/G) allele of *MECOM* and T allele of *CDNK2A/2B* genes **c** A allele of *TNFRSF19* gene

4.4.2. Multifactorial dimensionality reduction analysis for gene-gene interactions

We performed MDR analysis to invesigate high degree gene-gene interactions of *GSTT1*, *GSTM1*, *CYP1A1*, *XRCC1*, *MECOM*, and *TNFRSF19* genes in NPC carcinogenesis in the northeast Indian population. MDR analyses determine the best-model gene-gene interactions for NPC (Table 4.4.2). The analysis suggest that *GSTM1* null genotype was the best one-factor model with CVC 100/100 (TBA=0.605; *P*=0.0029). While *TNFRSF19* variant and *GSTM1* null genotype was the best two-factors model with CVC of 100/100 (TBA=0.64; *P*<0.0001), the best three-factor model was the combination of *TNFRSF19*, *GSTT1* and *GSTM1* variant genotype with maximum CVC of 84/100 (TBA=0.52; *P*<0.0001). The best four-factors model was the combinations of *MECOM*, *TNFRSF19*, *GSTT1* and *GSTM1* variant genotype with CVC of 86/100 and TBA (0.565) and *P*<0.0001. However, the best model of all the predictive models was the five-factor model combination of *TNFRSF19*, *GSTT1*, *GSTM1*, *CYP1A1*, *and XRCC1* variant genotype having maximum CVC of 100/100 and highest TBA of 0.675 (*P*<0.0001). Interaction entropy graph was also constructed using MDR results to determine synergistic or antagonistic (not-synergistic) interactions between the genetic factors in NPC risk (Figure 4.4.6).

Table 4.4.2 Summary of multifactorial dimensionality reduction analysis for NPC risk prediction

Model	**TrBA**	**TBA**	**CVC**	***P* value**
GSTM1	0.605	0.605	100/010	0.0029
TNFRSF19, GSTM1	0.640	0.640	100/100	<0.0001
TNFRSF19, GSTM1, GSTT1	0.655	0.520	84/100	<0.0001
MECOM, TNFRSF19, GSTM1, GSTT1	0.676	0.565	86/100	<0.0001
TNFRSF19, GSTM1, GSTT1, CYP1A1, XRCC1[a]	**0.720**	**0.675**	**100/100**	**<0.0001**
MECOM, TNFRSF19, GSTM1, GSTT1, CYP1A1, XRCC1	0.760	0.600	100/100	<0.0001

TBA, testing balance accuracy; TrBA, training balance accuracy; CVC, cross-validation consistency; [a]Best model prediction for NPC risk with highest TrBA, TBA and maximum CVC

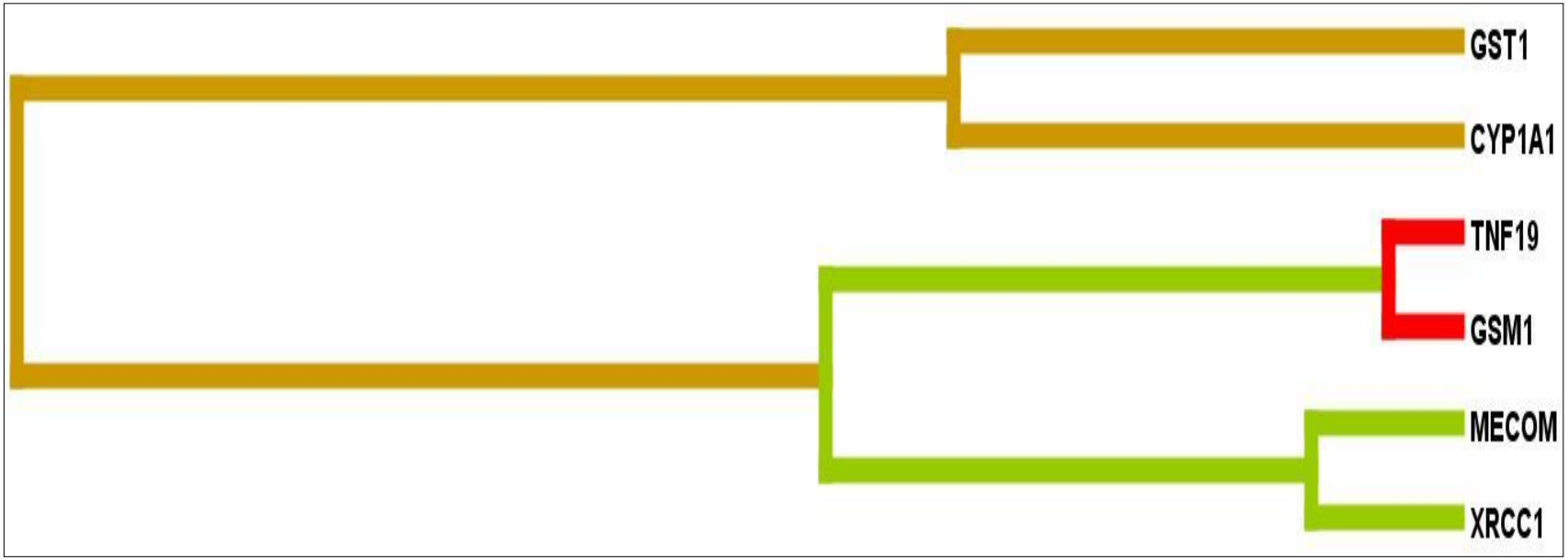


Figure 4.4.6 MDR analysis showing entropy graph for gene-gene interaction and NPC risk in the study. The red color indicating a high degree of synergistic interaction, orange a lesser degree whereas; gold represent midpoint; blue represents the highest level of redundancy followed by green. **GST1**-*GSTT1*; **TNF19**-*TNFRSF19*; and **GSM1**-*GSTM1*

4.4.3 False positive report possibility (FPRP)

We strengthened our data by testing the robustness and consistency of the gene-gene interaction using FPRP analysis. The FPRP values for all statistically significant result are summarized in Table 4.4.3. FPRP analysis indicated that the significant association between *GSTM1* null genotype and NPC risk was noteworthy for low prior probability assumptions (upto 0.1) for detecting OR= 2.0 for an FPRP value of 0.5. Moreover, the association was also deserving of attention for subject with combinations of variant genotypes. The best predictive models selected by MDR analysis were noteworthy for very low prior probability assumptions (upto 0.1 to 0.001) when detecting ORs of 1.5 and 2.0 for an FPRP value of 0.5.

Table 4.4.3 False Positive Reports Probability (FPRP) for odd ratios of the Multifactor Dimensionality Reduction (MDR) analysis

Variables	Odds ratio	OR= 1.5 (Prior Probability)				OR = 2.0 (Prior Probability)			
	OR (95% CI) *P*-value	0.25	0.1	0.01	0.001	0.25	0.1	0.01	0.001
GSTM1	2.35 (1.33-4.14) 0.0029	**0.136**	**0.321**	0.839	0.981	**0.032**	**0.090**	0.521	0.917
TNFRSF19, GSTM1	4.27 (2.17-8.41) <0.0001	**0.060**	**0.161**	0.679	0.955	**0.006**	**0.016**	**0.155**	0.650
TNFRSF19, GSTM1, GSTT1	3.93 (2.13-7.22) <0.0001	**0.032**	**0.090**	0.522	0.917	**0.002**	**0.006**	**0.067**	**0.419**
MECOM, TNFRSF19, GSTM1, GSTT1	4.37 (2.41-7.94) <0.0001	**0.017**	**0.048**	**0.357**	0.848	**0.001**	**0.002**	**0.023**	**0.193**
TNFRSF19, GSTM1, GSTT1, CYP1A1, XRCC1[a]	6.72 (3.61-12.5) <0.0001	**0.005**	**0.015**	**0.143**	0.627	**$<10^{-4}$**	**$<10^{-4}$**	**0.003**	**0.028**
MECOM, TNFRSF19, GSTM1, GSTT1, CYP1A1, XRCC1	11.9 (5.92-24.01) <0.0001	**0.004**	**0.011**	**0.108**	0.549	**$<10^{-4}$**	**$<10^{-4}$**	**0.001**	**0.014**

Prior probability range = 0.25–0.001 to detect OR = 1.5 or 2.0; α level = observed *P* value; Bold value= noteworthy association at 0.5 FPRP

[a]Best model prediction for NPC risk with highest TrBA, TBA and maximum CVC selected by MDR analysis

4.5. **Dysfunction of mitochondria due to environmental carcinogens and the risk of NPC**

The aim of this study was to investigate the effect mitochondrial DNA copy number alterations in NPC development and their relationship between tobacco (smoking and betel-quid chewing) and dietary (smoked meat and fermented fish intake) habits, and NPC risk. The mtDNA copy number per cell is maintained within a constant range to meet the energy requirement of the cell. Variations in the copy number of mitochondria reflects the net results of gene-environmental interactions between unknown hereditary factors and the levels of oxidative stress (an imbalance between ROS production and the antioxidant capacity), caused by a variety of endogenous and exogenous factors, such as dietary and environmental oxidants/ antioxidants and reaction to oxidative damage, all of which are thought to be the risk factors for cancer development. Here we determine the relative mitochondrial DNA (C-tract) content in 100 cases, 88 FDRs and 100 controls after determining the quality and quantity of DNA in the samples.

4.5.1. Relative quantification of mitochondrial DNA copy number (mtDNA) using quantitative real time PCR (qPCR)

Real-time PCR or quantitative PCR (qPCR) is an extension of the capabilities of traditional PCR that amplify and at the same time quantify a specific region of a DNA molecule. It is done so by including in the reaction a fluorescent molecule that reports an increase in the amount of DNA with a proportional increase in fluorescent signal. The measured fluorescence reflects the amount of amplified product in each cycle. In our study mtDNA copy number was measured using relative quantification

method in which the expression of a gene (C-tract) of interest in NPC cases samples was compared to expression of the same gene in controls. The results were expressed as fold change (increase or decrease) in expression of the cases in relation to the controls. A normalizer gene (such as *GAPDH*) is used as a control for experimental variability in this type of quantification.

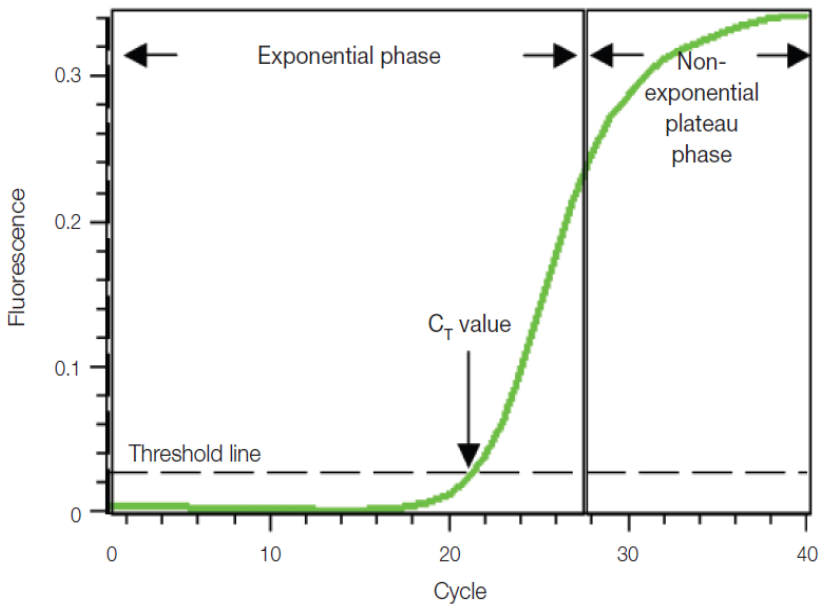
During the Real-time PCR experiment amplification plots were generated (Figure 4.5.1) where the threshold cycle (C_T) was determined i.e. the cycle at which a fluorescent signal is detected after accumulation of enough amplified products. The C_T levels are inversely proportional to the amount of target nucleic acid in the sample (i.e. lower the C_T level greater is the amount of target nucleic acid in the sample). Generally, C_T levels can be used to determine the quality of target nucleic acid as:

1. C_T levels < 29 indicates abundant target nucleic acid in the sample.
2. C_T levels of 30-37 indicates moderate amount of nucleic acid.
3. C_T levels of 38-40 indicates low/minimal amount of target nucleic acid which might represent a disease state or contamination.

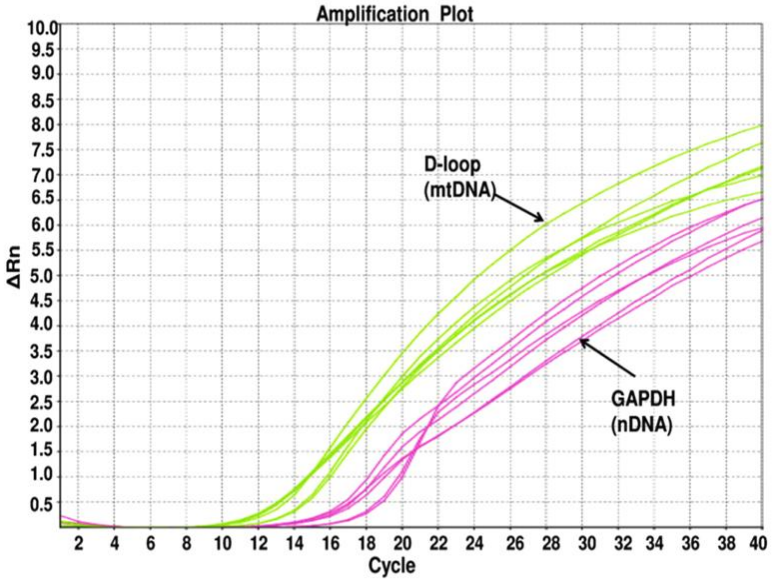
After the C_T values are measured, the $2^{-\Delta\Delta CT}$ (Livak) method was used to determine the expression level of the target gene in the case sample relative to the control sample. This method assumes that both target and reference genes are amplified with efficiencies near 100% and within 5% of each other. First, the C_T of the target gene to that of the reference (ref) gene was normalized, for both the test sample and the calibrator sample:

$$\Delta C_T \text{ (test)} = C_T \text{ (target, test)} - C_T \text{ (ref, test)}$$

$$\Delta C_T \text{ (calibrator)} = C_T \text{ (target, calibrator)} - C_T \text{ (ref, calibrator)}$$



a



b

Figure 4.5.1 Amplification plot **a** showing cycle threshold (C_T) where fluorescence signal is detected; **b** Quantitative PCR of D-loop region and *GAPDH* gene (representative curve). D-loop region and *GAPDH* are ubiquitous genes found in the mitochondrial and nuclear genomes, respectively. Using quantitative PCR in samples from the NPC cases, FDRs and control, the relative mitochondrial content was calculated

Second, the ΔCT of the test sample to the ΔC_T of the calibrator was normalized:

$$\Delta\Delta CT = \Delta C_T\,(test) - \Delta C_T\,(calibrator)$$

Finally, the expression ratio can be calculated as:

$$2^{-\Delta\Delta CT} = \text{Normalized expression ratio}$$

The result obtained was the fold increase (or decrease) of the target gene in the test sample relative to the calibrator sample and was normalized to the expression of a reference gene. Normalizing the expression of the target gene to that of the reference gene compensates for any difference in the amount of sample tissue.

4.5.2. Mitochondrial DNA content and the risk associated with NPC

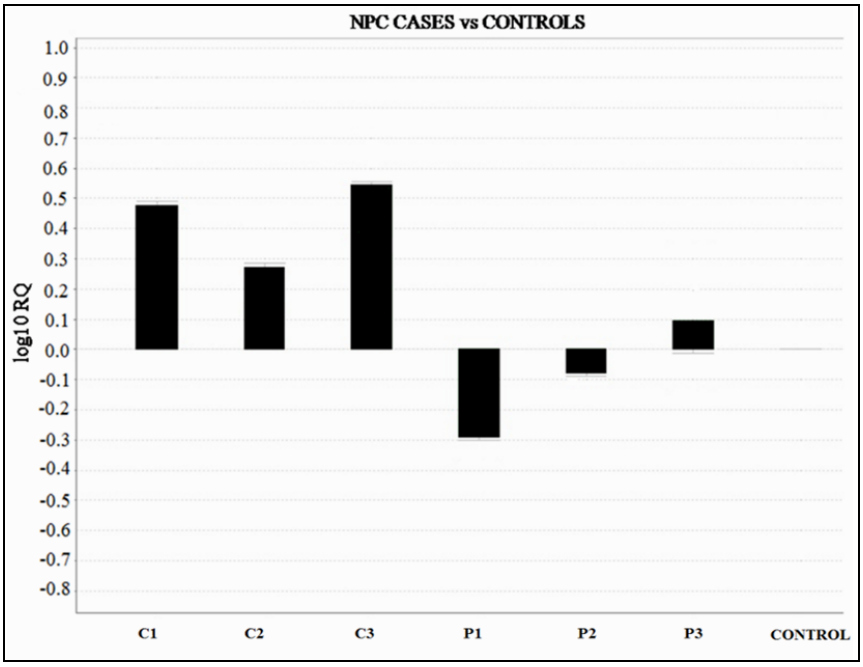
Using quantitative PCR techniques, we determined the relative content of mitochondrial DNA with respect to the *GAPDH* gene in 100 NPC cases, 88 FDRs of the cases and 100 controls without family history of cancer (Figure 4.5.2). The Ct values of GAPDH ranged from 14.60 to 26.89 in blood samples of patients with NPC and from 16.89 to 29.57 in whole blood samples of the control group, respectively. The Ct values of mitochondrial DNA (C-tract) ranged from 15.62 to 32.79 in blood samples of patients with NPC and from 16.1 to 29.78 in whole blood samples of the control group, respectively. In FDRs the Ct values of *GAPDH* ranged from 16.89 to 29.57; while mitochondrial DNA (C-tract) ranged from 16.45 to 28.95. The average Ct values of *GAPDH* amplicon in the normal group were correlated with those of mitochondrial DNA D-loop amplicons (Spearman's rho Test: $P < 0.001$, $r = 0.403$). In contrast, no significant correlation was detected in cancerous group ($P > 0.05$, $r = 0.169$), suggesting

an alteration of the relationship between nuclear DNA and mitochondrial DNA in blood samples of breast cancer patients (Table 4.5.1; Figure 4.5.3). Significant correlation was also detected in FDRs (P < 0.001, r = 0.513). Overall, the relative median of the mitochondrial DNA content is lower in cases (1.98 relative copies) than that in controls (4.11 relative copies) and FDRs (4.69 relative copies).

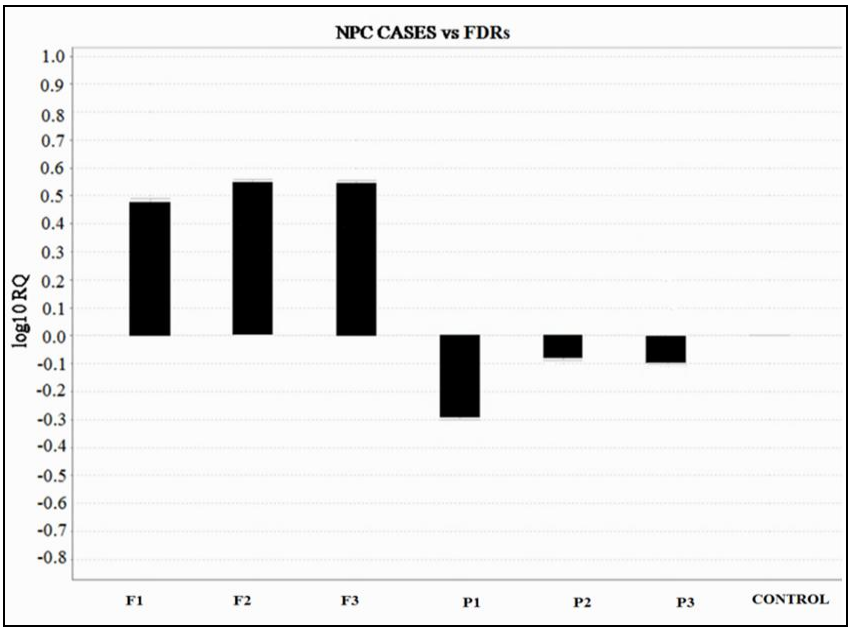
Table 4.5.1 Level of mtDNA content in peripheral blood of the control, FDRs and NPC patients

mtDNA content	Controls N=100	FDRs N=88	NPC N=100
Median	4.11	4.69	1.98
Correlation (Ct_{nDNA} *vs.* Ct_{mtDNA})	r = 0.403 ***P* < 0.001	r = 0.513 ***P* < 0.001	r = 0.169 ***P* = 0.9108

** Spearman's rho test

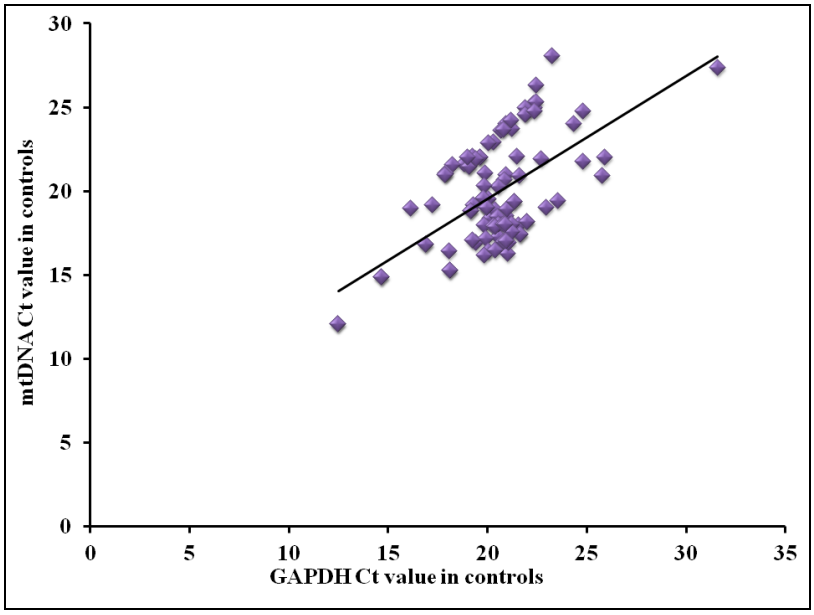


a

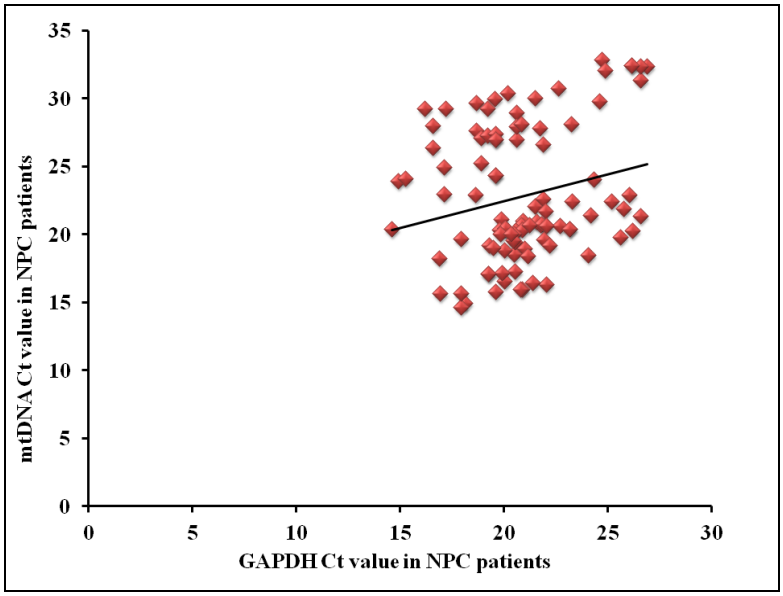


b

Figure 4.5.2 Quantitative PCR for mitochondrial DNA copy number determination **a** Mitochondrial content decreases in NPC cases as compared to that in controls; **b** mitochondrial content decreases in NPC cases in comparison to the FDRs with respect to log10RQ (log fold change).



a



b

Figure 4.5.3 Correlation of nuclear DNA content and mitochondrial DNA content in peripheral blood in normal group and NPC patients **a** Scatter plot illustrates the Ct values of *GAPDH* (X axis) against the Ct values of mitochondrial DNA (C-tract) (Y axis) in blood samples of normal group (n = 100). The correlation rate is highly significant ($P < 0.001$) according the Spearman' rho test; **b** Scatter plot illustrates the Ct values of *GAPDH* (X axis) against the Ct values of mitochondrial DNA (C-tract) (Y axis) in blood samples of NPC

group (n = 100). The correlation rate is not significant (P = 0.9108) according the Spearman' rho test

The non-parametric two-sample Kolmogorov-Smirnov test was used to compare of the mitochondrial DNA content between the normal and the NPC patients group. It was found that the data is normally distributed and the maximum cumulative difference. D=0.3900; P<0.001. It provides a significant difference between the two groups under study.

The relative mitochondrial DNA copy number content was group into four quartile; Quartile 1 (≤ 0.2), Quartile 2 (>0.2 - 2), Quartile 3 (> 2 - 12), and Quartile 4 (< 12) to determine the risk associated with NPC development. NPC cases in the lowest quartile of the mitochondrial DNA copy number experienced a significantly increased risk of 2.90-fold for NPC (95% CI: 1.06‑7.91) compared with those in the highest quartile (Table 4.5.2).We observed that risk of NPC increased with the decrease in mitochondrial DNA copy number (P trend=0.006) compared to FDRs where in the lowest quartile, no risk (OR=0.40; 95% CI: 0.15‑1.03) was observed and P_{trend}=0.05**.**

Table 4.5.2 Odds ratios (ORs) and 95 % CI for relative mitochondrial DNA copy number and risk of NPC

mtDNA copy number quartile	**Cases (N= 100)**	**Controls (N= 100)**	**OR (95% CI)**	**FDR (N= 88)**	**OR (95% CI)**
Quartile 1 (≤ 0.2)	40	29	2.62 (1.08-6.38)	11	0.40 (0.15-1.03)
Quartile 2 (> 0.2 - 2)	25	27	1.76 (0.70-4.44)	28	1.09 (0.47-2.51)
Quartile 3 (> 2 - 12)	25	25	1.90 (0.75-4.82)	31	1.30 (0.56-3.00)
Quartile 4 (> 12)	10	19	1 (ref)	18	1 (ref)
P_{trend}	0.006			0.05	

4.5.3. Mitochondrial DNA copy number and risk of NPC stratified by dietary and tobacco habits

The possible interaction of environmental factors and mitochondrial DNA copy number were investigated (Table 4.5.3). The mitochondrial DNA copy number was categories into 2 groups as High (> 2) and Low (≤ 2). Logistic regression method were use to examined the interaction, stratified by environmental factors. Smoked meat intake with low mitochondrial DNA copy number had 3.37 fold (95% CI: 1.43-7.95; *P*=0.006) increased risk of NPC. However, fermented fish consumers with low mtDNA copy had 5.49 fold (95% CI: 2.41-12.48; *P*<0.001) risk of NPC. Similarly, tobacco habits interact strongly with low mitochondrial DNA copy number. Smokers had 4.55 fold (95% CI: 1.88-11.01; *P*=0.001) and chewers had 3.55 fold (95% CI: 1.51-8.38; *P*=0.004) risk of NPC among individuals with low mitochondrial DNA copy number.

Table 4.5.3 Mitochondrial DNA copy number and risk of NPC stratified by genetic and environmental risk factors

Variables	**Mitochondrial DNA copy number**					
	High >2			**Low ≤ 2**		
	Ca /Co	**OR (95% CI)**	***P* value**	**Ca /Co**	**OR (95% CI)**	***P* value**
Smoked Meat						
Never	10/21	1.0	Ref.	12/23	1.10 (0.40-3.01)	1.0
Ever	25/23	2.28 (0.90-5.78)	0.106	53/33	3.37 (1.43-7.95)	**0.006**
Fermented Fish						
Never	15/35	1.0	Ref.	25/39	1.50 (0.69-3.26)	0.331
Ever	20/9	5.19 (1.95-13.79)	**0.001**	40/17	5.49 (2.41-12.48)	**<0.0001**
Smoking						
Never	17/29	1.0	Ref.	33/44	1.28 (0.61-2.69)	0.572
Ever	18/15	2.05 (0.83-5.02)	0.168	32/12	4.55 (1.88-11.01)	**0.001**
Tobacco betel quid chewing						
Never	10/26	1.0	Ref.	24/26	2.40 (0.97-5.93)	0.075
Ever	25/18	3.61 (1.42-9.21)	**0.012**	41/30	3.55 (1.51-8.38)	**0.004**

Fisher's exact test used to calculate *P* value and *P*<0.05 is considered as statistically significance; Bold values indicate statistical significance (*P* < 0.05)