Nasopharyngeal carcinoma (NPC) is a rare cancer around the world with an incidence rate under 1 per 100,000 populations per annum, but much more common among Cantonese Chinese, natives of South-East Asia, Arabs living in North Africa, Eskimos in Arctic and Native Americans. It is an uncommon malignancy in Indian subcontinent except in the north-eastern part of the country. NPC is caused by the chronic infection of Epstein Barr virus (EBV), salted fish, smoked meat, tobacco, alcohol, and inherited factors. 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. In this study the incidence, mortality and prevalence of NPC in northeast Indian population were analysed. The selected risk factors, mostly dietary, viral, environmental, genetic polymorphisms in metabolic genes (*GSTT1, GSTM1* and *CYP1A1*), DNA repair genes (*XRCC1* and *XRCC2*), highly susceptible loci (*TNFRSF19, CDKN2B-AS1,* and *MECOM*), mitochondrial DNA (mtDNA) copy number variation and their risk were examined, in subjects who are highly prone to NPC in the ethnic groups of northeast India, which include cases, first-degree relatives and controls. Complex disease such as cancer results from interactions of multiple genetic and environmental factors. Studying these factors singularly cannot explain the underlying pathogenic mechanism of the disease. Extensive investigations have been carried in NPC with the genetic and environmental risk factors with less emphasis on their interaction. Here, we determined a high degree gene-gene and gene-environment interactions that modulate an individual's susceptibility to NPC. Databases available in both international and national cancer registries were used to determined NPC prevalence in northeast India. The cases and controls were selected from three ethnic groups (Manipuri, Naga and Mizo) of northeast India with high prevalence of NPC. This case–control family study includes 123 NPC patients, 100

first-degree relatives (FDRs) and 189 controls having no history of cancer. PCR-based detection was done for EBV–latent membrane protein 1 (*LMP1*) gene. Analysis for *GSTM1* and *GSTT1* gene polymorphism was done by multiplex PCR. *CYP1A1* T3801C, *XRCC1* Arg399Gln and *XRCC2* Arg188His polymorphism was determined using PCR-RFLP, and the results were confirmed by sequencing. Genotyping of *TNFRSF19, CDKN2B-AS1,* and *MECOM* genes were done by Sanger sequencing method and further validated by SNaPshot single base extension assay. A comparative ΔCt method was used for the determination of mtDNA content. Logistic regression (LR) and multifactor dimensionality reduction (MDR) approach were applied for statistical analysis. False-positive report probability (FPRP) analysis was also performed to validate all significant findings. The salient findings of the study are as follows:

The global cancer database (GLOBOCAN) of international association of research in cancer (IARC) in 2012 reported 3947 new cases and 2836 death of NPC in India. In that year, there were 2956 and 991 new cases in among males and females, respectively. 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. In females, highest AAR of 5.2 per 100,000 was reported in Aizawl district of Mizoram State.

An increased risk of 1.98–2.87-folds to NPC was observed with regular intake of smoked meat and fermented fish; heavy smoking and tobacco betel quid chewing, herbal medicine use and those who are infected with EBV. The risk of NPC was modulated by lifestyle and dietary habits among EBV infected individuals. Highest risk of NPC was observed in smokers (OR=7.57) with EBV infection.

The *GSTM1* null genotype alone (OR=2.76) was significantly associated with NPC risk. The most remarkable risk was seen among individual carrying *GSTM1* null, *GSTT1* null genotypes and *CYP1A1* T3801C TC+CC genotypes (OR=5.71, P=0.001). Further; analyses demonstrate an enhanced risk of NPC in smoked meat (OR=5.56, P<0.0001) and fermented fish consumers (OR=5.73, P<0.0001) carrying *GSTM1* null genotype. An elevated risk of NPC was noted in smokers (OR=12.67, P<0.0001) and chewers (OR=5.68, P<0.0001) with *GSTM1* null genotype. However, smokers had the highest risk of NPC among individuals carrying *GSTT1* null genotype (OR=4.46, P=0.001) or *CYP1A1* T3801C TC+CC genotype (OR=7.13, P<0.0001). MDR analysis of metabolic gene and environmental interaction revealed the four-factor model combinations of *GSTM1*, fermented, smoked meat and smoking as the best model for NPC risk (CVC=10/10; TBA=0.6802; P<0.0001); while interaction entropy graph revealed synergistic interaction of *GSTM1* with *GSTT1*, fermented fish and smoked meat, respectively.

XRCC1 Gln/Gln genotype showed increased risk (OR=2.76; P<0.024) of NPC. However, individual with both *XRCC1* and *XRCC2* polymorphic variant had 3.2 fold elevated risk (P<0.041). An enhanced risk of NPC was 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 associated with NPC (OR = 7.47; *P*<0.0001). In 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 interaction with *XRCC1*.

Polymorphisms in *MECOM* (rs6774494) and *TNFRSF19* (rs9510790) genes had 2.1 and 1.8 fold risk of NPC, respectively. However, *CDNK2B-AS1* (rs1412829) did not show association with NPC risk. Further, MDR analysis revealed that the five factor model combinations of *TNFRSF19, GSTT1, GSTM1, CYP1A1* and *XRCC1* variant genotypes (CVC=100/100; TBA=0.675) as the best predictive model of NPC risk in our study.

Decreased mtDNA copy number (P trend=0.006) was found to be associated with NPC risk. Interaction study of mtDNA copy number and environmental factors revealed that intake of smoked meat (OR=3.37) and fermented fish (5.49), smoking (OR=4.55) and tobacco betel quid chewing (OR=3.55) had increased risk of NPC among individuals with low mtDNA copy number. A significant difference between GST null genotypes and EBV infection with mtDNA content was also found in the cases (P<0.0001).

To the best of our knowledge, we are the first to conduct a genetic study of NPC in the northeast Indian population, and their combined effect with dietary and tobacco habits that modify susceptibility towards NPC. We showed that EBV infection, polymorphic variants of *GSTM1, XRCC1, MECOM, TNFRESF19* genes and decreased mtDNA content is a strong predisposing risk factor for NPC. 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 NPC risk. The understandings of environment–genetic risk factors and their role in the etiology of NPC are helpful as preventive measures and screening.