Department of Biotechnology
Hargobind Khurana School of Life Sciences
Assam University, Silchar
(A Central University constituted under Act XIII of 1989)
Silchar-788011, Assam, India

DECLARATION

I, **Seram Anil Singh**, bearing registration number Ph.D./2177/12 dated 26/09/12, hereby declare that the subject matter of the thesis entitled **"Study of the risk factors associated with Nasopharyngeal Carcinoma in North-East India"** is the record of work done by me and that the contents of this thesis did not form the basis for award of any degree to me or to anybody else to the best of my knowledge. The thesis has not been submitted in any other University / Institute.

Place: Assam University, Silchar **(Seram Anil Singh)**

Date:

ACKNOWLEDGEMENT

I take this opportunity to convey my gratefulness to one and all those who have supported and guided me through the entire tenure of PhD work.

First and foremost, a sincere and heart filled gratitude to my mentor Prof. (Dr) Sankar Kumar Ghosh for being such a wonderful and motivating guide. His stimulating suggestions, experience and encouragement accompanied by the freedom of thought that he granted, has not only helped me to discover my potential but also kept me sparked throughout my Ph.D tenure. I will never forget the generosity that he offered on several occasions. Sir, I thank you for everything and will always be proud to have been mentored by you.

I want to express my warmest thanks to all Molecular Medicine Lab members, Dr. Shilpee Srivastava, Dr. Rosy Mondal, Dr. Fazlur R Talukdar; Dr. Ruhina S Laskar, Bishal Dhar, Shailendra Yadav, Sharbadeb Kundu, Wetetsho Kapfo, Manish Kumar, Saheen Laskar and Raima Das who helped me in innumerable ways during the entire period. A special thanks to my batch mate Javed H Choudhury who has always been there for me. Thanks to Abhijeet Paul, and Sukumar Sonar for all their enumerable help. I would also like to thank all MSc students who help me during their dissertation Sourav Phukan, Niren Khomdram, and Sourav Sen. All these people have kept the lab environment lively during these years and made it a second home.

I would like to convey my heartiest gratitude to Dr. Y. Indibor Singh, Regional Institute of Medical Science (RIMS), Manipur; Dr. V Khamo, Nagaland Hospital Administration Kohima (NAHAK), Nagaland, and Dr. Z Eric, Government Civil Hospital, Aizwal, Mizoram for providing valuable clinical inputs and samples for

this research. Special thanks to Dr. Rajesh Laishram, RIMS for his priceless support while analysing PBCRs data.

I want to express my sincere gratitude to Department of Biotechnology (DBT), Government of India for providing infra-structural facilities (BT/Med/NE-SFC/2009) and (BT/NE/TBP/204(Med)/3/2011) which makes this work a reality.

I am also thankful to Dean Hargobind Khurana School of life Science, Head Department of Biotechnology, and faculty members, technical and official staff of Department of Biotechnology for their kind co-operation.

And the most important people in my life, my parent, my brother and sisters, without their love and support the completion of this work would have not been possible. Finally, Geetanjali, my wife you always have been my greatest strength and source of inspiration.

Date:

Place: Assam University, Silchar (**Seram Anil Singh)**

CONTENTS

Chapter **Topics** **Page**

List of Tables.....i

List of Figures.....iii

1 Introduction.....1-32

1.1 Cancer.....1

1.1.1 Classification of cancer.....2

1.1.2 Global Cancer Scenario.....3

1.2 Nasopharyngeal Cancer.....4

1.2.1 Stages of NPC.....5

1.2.2 Histological subtypes of NPC.....5

1.2.3 Racial and geographical distribution.....6

1.2.3.1 Ethic information of northeast India: A high incidence region of NPC.....7

1.3 Etiology of nasopharyngeal carcinoma.....8

1.3.1 Environmental factors.....8

1.3.1.1 Dietary factors.....8

1.3.1.2 Tobacco use.....9

1.3.1.3 Alcohol consumption.....11

1.3.1.4 Occupational exposure.....12

1.3.1.5 Viral pathogenesis.....12

1.4 Genetic susceptibility of NPC.....14

1.4.1 Metabolic activation and detoxification genes.....15

1.4.1.1 Phase I metabolic gene.....17

Cytochrome P 450, superfamily A, polypeptide 1 (CYP1A1)

1.4.1.2 Phase II metabolic gene.....18

Glutathione S-transferase (*GSTs*)

1.4.2 DNA repair genes.....19

1.4.2.1 Polymorphism in repair genes.....20

XRCC1 (x-ray repair cross-complementation 1)

XRCC2 (x-ray repair cross-complementation 2)

1.4.3 Newly identified susceptible loci.....21

TNFRSF19 (Tumor Necrosis Factor Receptor Superfamily, Member19)

MDS1 and *EVI1* [*MDS1* (myelodysplasia 1) and *EVI1* (ecotropic viral insertion site 1 fusion proteins) complex locus protein]

CDKN2B-AS1 (*CDKN2B* antisense RNA 1)

1.4.4 Mitochondria.....24

1.4.4.1 Mitochondria in cancer.....26

1.4.4.2 Mitochondrial DNA copy number alterations in human cancer.....27

1.5 Rationale of the propose study.....28

1.5.1 Objectives of the study.....32

2 Review of Literature.....33-59

2.1 Global scenario of cancer.....33

2.1.1 India cancer scenario.....35

2.1.2 Nasopharyngeal carcinoma: A global perspective.....36

2.2 Environmental risk factors for NPC.....39

2.2.1 Dietary factors: Preserved foods, herbal medicine, fresh fruits and vegetables.....39

2.2.2 Lifestyle related factors: Tobacco, other smoke, and alcohol.....42

2.2.3 Viral factor: Epstein Barr virus infection.....45

2.3 Genetic factors.....48

2.3.1 Metabolic and DNA repair genes.....48

2.3.2 Newly identified susceptible loci in GWAS.....52

2.3.3 MtDNA copy number variation in human cancers.....55

3 Materials and Methods.....60-81

3.1 Materials.....60

3.1.1 Survey for cancer prevalence.....60

3.1.1.1 Survey of International Cancer Registries.....60

3.1.1.2 Survey of National Cancer Registries.....60

3.1.2 Study design.....60

3.1.2.1 Ethical Clearance.....61

3.1.2.2 Selection of subjects.....61

3.1.2.3 Collection of samples.....62

3.1.3 Chemicals and Reagents.....62

3.1.3.1 Stock solution preparation.....63

3.1.3.2 Buffers preparation.....64

3.2 Methods.....65

3.2.1 Isolation of genomic DNA.....65

3.2.1.1 Isolation from buccal swap.....66

3.2.1.2 Isolation from whole blood.....67

3.2.2 Quality and quantity analysis of the extracted DNA.....68

3.2.2.1 Spectrophotometric Analysis.....68

3.2.2.2 Agarose gel electrophoresis for genomic DNA quality and quantity analysis.....69

3.2.3 Polymerase chain reaction (PCR).....69

3.2.3.1 Primer designing and optimisation.....70

3.2.3.2 Restriction enzyme digestion.....71

3.2.3.3 Multiplex PCR genotyping for *GSTT1* and *GSTM1* metabolic genes.....71

3.2.3.4 PCR-RFLP genotyping of *CYP1A1* T3801C polymorphism.....72

3.2.3.5 PCR amplification of viral *LMP1* gene to detection of EBV infection.....72

3.2.3.6 PCR-RFLP genotyping of *XRCC1* Arg399Gln polymorphism....72

3.2.3.7 PCR-RFLP genotyping of *XRCC2* Arg188His polymorphism...73
3.2.3.8 PCR Amplification of Hotspot Genes Related to NPC.....73
3.2.3.9 Purification of PCR products by Gel purification.....74
3.2.3.10 Sanger Sequencing.....75
3.2.3.11 Relative quantitative of mitochondrial DNA copy number using real-time PCR.....76
3.2.3.12 SNaPShot Multiplex System.....77
3.2.3.13 Genotyping of *MECOM, TNFRSF19* and *CDK2BAS1* genes using SNaPshot assay.....78
3.3 Bioinformatics analysis.....79
3.4 Statistical analysis.....79
3.4.1 Logistic Regression Analysis.....80
3.4.2 Multifactor Dimensionality Reduction (MDR) Analysis.....80
3.4.3 Interaction Entropy Graphs.....81
3.4.4 False Positive Report Probability (FPRP).....81
4 Results.....82-145
4.1 Prevalence and environmental risk factors associated with nasopharyngeal cancer (NPC) among the ethnic population of Northeast India.....82
4.1.1 Nasopharyngeal carcinoma in different cancer registries.....82
4.1.2 Demographic characteristics of the study population.....86
4.1.3 Environmental risk factors associated with NPC.....89
4.1.3.1 Dietary habits and risk o NPC development.....89
4.1.3.2 Tobacco and alcohol habits and herbal medicine use and risk of NPC development.....90
4.1.3.3 Epstein Barr Virus (EBV) infection and risk of NPC development.....92
4.1.4 Interaction of Epstein Barr Virus (EBV) and other environmental risk factors in NPC.....93

4.2. Metabolic genes, their interaction with environmental factors in NPC.....96

4.2.1 Polymorphisms in metabolic genes (*GSTT1, GSTM1* and *CYP1A1*) and the risk associated with NPC.....96

4.2.2 Interaction of metabolic genes and environmental risk factors in NPC progression……..……..….103

4.2.3 Multifactorial dimensionality reduction analysis of metabolic gene-environmental interaction……..……..…..108

4.2.4 Interaction entropy graph of metabolic gene-environment interactions.....109

4.2.5 False positive report possibility (FPRP).....111

4.3 Polymorphism in DNA repair genes, their interaction with environmental risk factors in NPC.....115

4.3.1 Polymorphisms in DNA repair genes (*XRCC1* and *XRCC2*) and the risk of associated with NPC.....115

4.3.2 Interaction of DNA repair genes and environmental risk factors in NPC progression…….……120

4.3.3 Multifactorial dimensionality reduction analysis of repair gene-environmental interaction and NPC ris……..123

4.3.4 Interaction entropy graph of repair gene-environment interactions…124

4.3.5 False positive report possibility (FPRP).....126

4.4 Polymorphism in newly identified susceptible loci (*MECOM, TNFRSF19* and *CDKN2B-AS1*) and risk of NPC in northeast Indian population.....128

4.4.1 Genetic variants of *MECOM, TNFRSF19* and *CDKN2B-AS1* genes and NPC predisposition.....128

4.4.2 Multifactorial dimensionality reduction analysis for gene-gene interactions.....134

4.4.3 False positive report possibility (FPRP).....136

4.5 Dysfunction of mitochondria due to environmental carcinogens and the risk of NPC.....137

4.5.1 Relative quantification of mitochondrial DNA copy number (mtDNA) using quantitative real time PCR (qPCR).....137

4.5.2 Mitochondrial DNA content and the risk associated with NPC.....141

4.5.3 Mitochondrial DNA copy number and risk of NPC stratified by dietary and tobacco habits.....145

5 Discussion.....146

Summary.....159

Bibliography.....163

Appendix

A. List of Abbreviations.....v

B. List of Publications.....viii

C. Research papers presented in scientific conferences.....ix

Published Research Articles

List of Tables

Tables	Page No
Table 1.1 Cancer Incidence, mortality and prevalence in the world	**3**
Table 1.2 Risk allele associated with the three hotspot genes	**22**
Table 2.1 Incidence, mortality and prevalence of NPC in the World	**37**
Table 2.2 NPC (Age-adjusted incidence rate) by sex and world areas	**37**
Table 3.1 List of primers use for amplification of different genes	**70**
Table 3.2 External primer pair for SNaPshot	**78**
Table 3.3 Internal primers for SNaPshot	**78**
Table 4.1.1 NPC incidence, mortality and prevalence: Estimate for India, 2012	**82**
Table 4.1.2 Comparison of age adjusted incidence rates (AAR) of PBCRs in northeast India, 2009-2010	**84**
Table 4.1.3 Number of cases, Percentage (%) and Rate of Incidence Cases, 2010-2012	**85**
Table 4.1.4 Demographic characteristics and socioeconomic status of the cases, FDRs and controls	**88**
Table 4.1.5 Smoked meat, fermented fish, tobacco and alcohol habits and the risk of NPC	**91**
Table 4.1.6 Sex, age and ethnicity wise detection of EBV in the study population	**93**
Table 4.1.7 Risk of NPC by intake of smoked meat, fermented fish, tobacco smoking, herbal medicine use and EBV infection	**95**
Table 4.2.1 Distribution of *GSTM1*, *GSTT1* and *CYP1A1* T3801C genotype among the study subjects	**99**
Table 4.2.2 Odds ratios for the interaction of *GSTM1*, *GSTT1* and *CYP1A1* T3801C genotypes in the study subjects	**102**
Table 4.2.3 Association between *GSTM1* genotype and NPC, stratified by smoked meat, fermented fish, smoking and tobacco-betel quid habits	**105**
Table 4.2.4 Association between *GSTT1* genotype and NPC, stratified by smoked meat, fermented fish, smoking and tobacco-betel quid habits	**106**
Table 4.2.5 Association between *CYP1A1* T3801C genotype and NPC, stratified by smoked meat, fermented fish, smoking and tobacco-betel quid habits	**107**

Table 4.2.6 Summary of multifactorial dimensionality reduction analysis for NPC risk prediction **108**

Table 4.2.7 False Positive Reports Probability (FPRP) for odd ratios of the Logistic Regression (LR) analysis in gene-gene interaction **112**

Table 4.2.8 False Positive Reports Probability (FPRP) for odd ratios of the Logistic Regression (LR) analysis in gene-environmental interaction **113**

Table 4.2.9 False Positive Reports Probability (FPRP) for odd ratios of the Multifactor Dimensionality Reduction (MDR) analysis in gene-environmental interaction **114**

Table 4.3.1 Genotype frequency distribution of *XRCC1* Arg399Gln and XRCC2 Arg188His polymorphisms and risk of NPC **119**

Table 4.3.2 Combined genotype analysis of XRCC1 Arg399Gln and XRCC2 Arg188His on risk of NPC **120**

Table 4.3.3 Association of *XRCC1* Arg399Gln genotype, stratified by smoked meat, fermented fish, smoking, and tobacco-betel quid habits **122**

Table 4.3.4 Summary of multifactorial dimensionality reduction analysis for NPC risk prediction **123**

Table 4.3.5 False Positive Reports Probability (FPRP) for odd ratios of the Logistic Regression (LR) and Multifactor Dimensionality Reduction (MDR) analysis **127**

Table 4.4.1 Genetic variants of *MECOM*, *CDNK2B-AS1* and *TNFRSF19* genes and NPC predisposition **132**

Table 4.4.2 Summary of multifactorial dimensionality reduction analysis for NPC risk prediction **134**

Table 4.3.4 False Positive Reports Probability (FPRP) for odd ratios of the Multifactor Dimensionality Reduction (MDR) analysis **136**

Table 4.5.1 Level of mtDNA content in peripheral blood of the control, FDRs and NPC patients **141**

Table 4.5.2 Odds ratios (ORs) and 95 % CI for relative mitochondrial DNA copy number and risk of NPC **144**

Table 4.5.3 Mitochondrial DNA copy number and risk of NPC stratified by genetic and environmental risk factors **145**

List of Figures

Figures	Page No
Figure 1.1 Main sites of metastases for some common cancer types	**2**
Figure 1.2 Schematic representation of Head and Neck region	**4**
Figure Geographical map of northeast India States showing high NPC incidence areas viz., Manipur, Mizoram and Nagaland	**7**
Figure 1.4 Percentages of different factors contributing to cancer	**8**
Figure 1.5 A model for Epstein Barr virus (EBV) infection and persistence	**14**
Figure 1.6 Schematic representation of xenobiotic detoxification and processes leading to cancer development	**16**
Figure 1.7 Illustration of human mitochondrial genome with its 37 genes and the coding region	**25**
Figure 3.1 Schematic representation of SNaPshot genotyping	**77**
Figure 4.1.1 Photomicrograph of nasopharynx	**87**
Figure 4.1.2 Bar diagram showing the risk (Odds ratios) of NPC associated with environmental factors (smoked meat and fermented fish).	**89**
Figure 4.1.3 Bar diagram showing the risk (Odds ratios) of NPC associated with environmental factors (tobacco-betel quid, smoking, alcohol and herbal medicine)	**90**
Figure 4.1.4 PCR based detection of *EBV-LMP1* in NPC	**92**
Figure 4.1.5 Odds ratio for interaction of EBV and dietary and tobacco habits	**94**
Figure 4.2.1 Polymorphism in *GSTM1*, *GSTT1* and *CYP1A1* metabolic genes	**97**
Figure 4.2.2 Sanger sequencing results showing nucleotide changes confirming PCR-RFLP results	**98**
Figure 4.2.3 Bar diagram showing the risk (Odds ratios) of NPC associated with *CYP1A1* T3801C, *GSTT1* and *GSTM1* polymorphism	**100**
Figure 4.2.4 Interaction entropy graph for gene-environmental interaction in NPC	**109**
Figure 4.2.5 Summary of the four-factor model (*GSTM1*, fermented fish, smoked meat and smoking) in MDR analysis	**110**

Figure 4.3.1 PCR based detection of *XRCC1* Arg399Gln gene polymorphism in NPC	**117**
Figure 4.3.2 PCR based detection of *XRCC2* Arg188His gene polymorphism in NPC	**118**
Figure 4.3.3 Combined effect of dietary and tobacco habits with *XRCC1* Arg399Gln genotypes	**121**
Figure 4.3.4 MDR analysis showing entropy graph for gene-environment interaction and NPC risk	**124**
Figure 4.3.5 Summary of the five-factor model (*XRCC1*, fermented fish, smoked meat, smoking and chewing) in MDR analysis	**125**
Figure 4.4.1 Multiplex PCR amplification of *MECOM*, *TNFRSF19* and *CDKN2B-AS1* genes	**128**
Figure 4.4.2 DNA sequencing result of *TNFRSF19* (rs9510787) gene	**129**
Figure 4.4.3 DNA sequencing result of *CDKN2B-AS1* (rs1412829) gene	**130**
Figure 4.4.4. DNA sequencing result of *MECOM* (rs6774494) gene	**131**
Figure 4.4.5 Validation of sequencing results by SNapShot singel base extension method	**133**
Figure 4.4.6 MDR analysis showing entropy graph for gene-gene interaction and NPC risk in the study	**135**
Figure 4.5.1. Amplification plot showing **a** cycle threshold (CT) where fluorescence signal is detected; **b** Quantitative PCR of D-loop region and GAPDH gene (representative curve**)**	**139**
Figure 4.5.2. Quantitative PCR for mitochondrial DNA copy number determination	**142**
Figure 4.5.3. Correlation of nuclear DNA content and mitochondrial DNA content in peripheral blood in normal group and NPC patients	**143**
Figure 5.1 Schematic representation of the etiology of nasopharyngeal carcinoma (NPC)	**157**

List of Abbreviations

AAR-	Age adjusted incidence rate
BART-	*BamHI* A rightward transcript
Bcl2-	B-cell lymphoma 2
BER-	Base excision repair
BHRF1-	BamHI fragment H rightward open reading frame
CD21-	Cluster of differentiation 21
CDK-	Cyclin dependent kinase
CDKN2B-AS1-	CDKN2B antisense RNA 1
CHRNA3-	Cholinergic receptor nicotinic alpha 3
CIAP-	Calf intestine alkaline phosphatase
COX2-	Cyclooxygenase-2
CR2-	Complement receptor type 2
CSB-	Conserve sequence box
CVC-	Cross validation consistency
CYP1A1-	Cytochrome P450, family 1, member A1
CYP2E1-	Cytochrome P450, family 2, subfamily E, polypeptide 1
DNA-	Deoxyribose nucleic acid
DRC-	DNA repair capacity
EA-	Early antigen
EBER-	Epstein–Barr virus-encoded small RNA
EBNA-	Epstein Barr Nuclear Antigen
EBV-	Epstein-Barr virus
EDTA-	Ethylene diamine tetrachloro acetic acid
ENT-	Ear nose throat
ESCC-	Esophageal squamous cell carcinoma
ETAS-	Extended termination associated sequences
EtBr-	Ethidium bromide
EVI1-	Ecotropic viral insertion site 1
FDRs-	First degree relatives
FFPE-	Formalin fixed paraffin embedded tissues
FMO-	Flavin-containing monooxygenases
FPRP-	False positive report probability
GABBR1-	Gamma amino butyric acid b receptor 1

GAPDH-	Glyceraldehyde 3-phosphate dehydrogenase
GST-	Glutathione-S-transferase
GSTM1-	Glutathione S-transferase mu 1
GSTP1-	Glutathione S-transferase pi 1
GSTT1 -	Glutathione S-transferase theta 1
GWAS-	Genome wide association study
HCC-	Human hepatocellular carcinoma
HLA-	Human leukocyte antigen
HpaII-	*Haemophilus parainfluenzae* II
HphI-	*Haemophilus parahaemolyticus* I
HRR-	Homologous Recombination Repair
HSP-	Heavy strand promoter
JNK-	c-Jun N-terminal kinase
IARC-	International agency for research on cancer
hOGG1-	Human 8-oxoguanine DNA N-glycosylase 1
KSHV-	Kaposi's sarcoma associated herpes virus
LMP1-	Latent membrane protein 1
LR-	Logistic regression
LSP-	Light strand promoter
MDR-	Multifactor dimensionality reduction
MDS 1-	Myelodysplasia 1
MECOM-	Myelodysplasia 1 and ecotropic viral insertion site 1
MspI-	*Moraxella species*
MT-	Methyl transferase
NaCl-	Sodium chloride
NADPH-	Nicotinamide adenine dinucleotide phosphate
NAT2-	N-acetyltransferase 2
NCRP-	National Cancer Registry Programme
NDEA-	Nitrosodiethylamine
NDMA-	Nitrosodimethylamine
NER-	Nucleotide excision repair
NNN-	Nitrosonornicotine
NNK-	4 (Methylnitrosamino)-1,3-pyridyl-1-butanone
NPC-	Nasopharyngeal carcinoma
NPYR-	Nitrosopyrrolidine

NQO-	NADPH-quinone oxidoreductase
NQO1-	NAD (P) H dehydrogenase quinone 1
NSCLC-	Non-small cell lung cancer
OSCC-	Oral squamous cell carcinoma
OXPHOS-	Oxidative phosphorylation
PAH-	Polycyclic aromatic hydrocarbon
POLG-	Polymerase gamma
PRC-	Polycomb repressive complex
rRNA-	Ribosomal RNA
tRNA-	Transfer RNA
RCC-	Renal cell carcinoma
RFLP-	Restriction fragment length polymorphism
ROS-	Reactive oxygen species
SDS-	Sodium dodecyl sulphate
SSC-	Squamous-cell carcinoma
SULT-	Sulfotransferase
TAE-	Tris acetic acid EDTA
TBA-	Testing balance accuracy
TERT-	Telomerase reverse transcriptase
TES-	Tris-HCl EDTA sodium chloride
TGF-	Transforming growth factor
TNFRSF19-	Tumor necrosis factor receptor super family 19
TRADD-	Tumour necrosis factor receptor type 1-associated death domain
TRAF-	TNF receptor-associated factor
TrBA-	Training balance accuracy
TSNA-	Tobacco specific nitrosamines
UGT-	UDP-glucuronosyltransferase
VCA-	Viral capsid antigen
VEGF-	Vascular endothelial growth factor
WHO-	World Health Organization
XRCC1-	X-ray repair cross-complementing protein 1
XRCC2-	X-ray repair cross-complementing protein 2