Department of Biotechnology
School of Life Sciences
ASSAM UNIVERSITY, SILCHAR
(A Central University Constituted under Act XIII of 1989)
Silchar-788011, Assam, India

DECLARATION

I, **Javed Hussain Choudhury**, bearing Ph.D. registration number PhD/2175/12 dated 26/09/2012, hereby declare that the subject matter of the thesis entitled **"Studies on Human Papillomavirus (HPV) Prevalence, Genetic Polymorphisms and Epigenetics in Head and Neck Cancer"** is the record of work done by me and 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

Date: Javed Hussain Choudhury

Dedicated to

People Suffering from Cancer
and
Their Families

ACKNOWLEDGEMENTS

First and foremost, I express heartfelt and deepest sense of gratitude to my supervisor Prof. (Dr) Sankar Kumar Ghosh for his constant guidance, suggestions, constructive criticisms, scholarly contribution and consistent support at every step of my Ph.D. programme. I feel grateful to express my appreciation for the trust he put in me, and the liberty that he granted, which helped me to discover my potential and to work fluently throughout my doctoral programme. Thank you Sir, for giving me the opportunity to work with your research team.

I am very much thankful to Honourable Vice Chancellor, Dean School of Life Sciences, Head of the Department of Biotechnology and Board of Research Studies, Assam University for their support and co-operation. I would like to convey my heartiest gratitude to all the faculty members and non-teaching staffs of the Department of Biotechnology, Assam University for providing moral and technical support at various phases of this research tenure.

I express my sincere and heart filled gratitude to all my lab members Boni A Laskar, Shilpee Srivastava, Rosy Mondal, Joyraj Bhattachrjee, Jagadish Hansa, Pradosh Mahadani, Shantanu Kundu, Fazlur R Talukdar, Ruhina S Laskar, Biswadeep Choudhury, Bishal Dhar, Shailendra Yadav, Sharbadeb Kundu, Wetetsho Kapfo, Manish Kumar, Saheen Laskar and Raima Das for their helping hands, gestures, lunches and providing adequate atmosphere for research. How can I ever forget my Ph.D. colleague Seram Anil Singh and roommates Sambuddha Das and Sayantan Nath, they have been always there for me and made my Ph.D. life easier. I would like to thanks to Abhijeet Paul and Sukumar Sonar for all their active co-operation and help during the experiment. I would also like to thank all M.Sc project students who help me during their project work.

I express my deepest gratitude to my family members for their love, moral support, good wishes, blessing and inspiration over the years. I am proudly grateful to my beloved parents, Amzad H Choudhury and Jahanara B Choudhury, who are the strength and source of inspiration for me, who sacrificed a lot for making me whatever I am today.

I am thankful to the Department of Biotechnology (DBT), Government of India for providing fellowship and infrastructure facilities (BT/MED/NE-SFC/2009) during my PhD tenure.

I am extremely grateful to Cachar Cancer Hospital and Research Centre (CCHRC) and Silchar Medical College and Hospital (SMC), Assam; Agartala Government Medical College, Tripura; Naga Hospital Administration, Nagaland; RIMS, Manipur; and Civil Hospital, Mizoram for providing samples for research.

Finally yet importantly, I am very much thankful to everyone who directly or indirectly helped me in completion of my research work.

Above all, I thank Almighty God, who blessed me with peace of mind and courage to carry out this work.

Javed Hussain Choudhury

CONTENTS

	Page No.
List of Tables	i-ii
List of Figures	iii-iv
Abbreviations	v-vii
CHAPTER-1: INTRODUCTION	**1-27**
1.1 The Head and Neck Cancer: Overview	1
1.2 Epidemiology of Head and Neck Cancer	2
1.2.1 Global Incidence and Mortality	2
1.2.2 Incidence in India	5
1.3 Aaetiology of Head and Neck Cancer	6
1.3.1 Tobacco Smoking	6
1.3.2 Smokeless Tobacco and Betel Quid Chewing	7
1.3.3 Alcohol Consumption	8
1.3.4 Diet	8
1.3.5 Human Papillomavirus (HPV) Infection	9
1.4 Clinical Perceptive of Head and Neck Cancer	10
1.4.1 Sign and Symptoms	10
1.4.2 Staging	11
1.4.3 Treatment	12
1.5 Genetic Alternations	13
1.5.1 Metabolic Genes Polymorphisms	14
1.5.2 DNA Repair Genes Polymorphisms	16
1.6 Epigenetic Alterations	18
1.6.1 Promoter Methylation of Tumour Suppression Genes	22
1.6.2 CpG Island Methylator Phenotype (CIMP)	23
1.7 Rationale of the Propose Study	24
1.8 Objectives of the Study	27
CHAPTER-2: REVIEW OF LITERATURE	
2.1 Head and Neck Cancer Incidence	**28-45**
2.1.1 The Global Scenario	28
2.1.2 The Indian scenario	30
2.2 Aetiology of Head and Neck Cancer	31
2.2.1 Environmental Risk Factors	31
2.2.2 Human Papillomavirus (HPV)	34
2.3 Genetic Alteration in Head and Neck Cancer	36
2.3.1 Genetic Polymorphisms of Metabolic Genes	36

2.3.2 Genetic Polymorphisms of DNA Repair Genes 39
2.4 Epigenetic Alterations and Head and Neck Cancer 42
2.4.1 Promoter Methylation of Tumour-Related Genes 42
2.4.2 CpGIsland Methylator Phenotype (CIMP) 44

CHAPTER-3: MATERIALS AND METHOS 46-66

3.1 Materials 46
3.1.1 Study Design 46
3.1.2 Ethical Clearance 47
3.1.3 Chemicals and Reagents 47
3.1.3.1 Stock Solution 47
3.1.3.2 Buffers 49
3.1.2.3 Primers 50
3.2. Methods 52
3.2.1 Isolation and Purification of Genomic DNA from Tissue and Blood 52
3.2.1.1 Isolation of Genomic DNA from Tissue 52
3.2.1.2 Isolation of Genomic DNA from Blood 53
3.2.2 Purity Check and Yield of the Extracted DNA 53
3.2.2.1 Spectrophotometric Analysis 54
3.2.2.2 Agarose Gel Electrophoresis for Quantification and Quality Analysis 54
3.2.3 Genotyping of *GSTM1* and *GSTT1* by Multiplex Polymerase Chain Reaction 55
3.2.4 Genotyping of *CYP1A1*, *XRCC1* and *XRCC2* 56
3.2.4.1 Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP) 56
3.2.3.2 Sanger Sequencing 58
3.2.5 Bioinformatic Analysis 60
3.2.6 Methylation Analysis 60
3.2.6.1 Bisulfite Treatment of Genomic DNA 60
3.2.6.2 PCR Amplification of the Modified Genomic DNA by Methylation Specific PCR (MSP) 63
3.2.7 Statistical Analysis 64
3.2.7.1 Logistic Regression (LR) Analysis 65
3.2.7.2 Multifactor Dimensionality Reduction (MDR) 65
3.2.7.3 Interaction Entropy Graphs 66
3.2.7.4 False Positive Report Possibility (FPRP) 66
3.2.7.5 Survival Analysis 66

CHAPTER-4: RESULTS **67-113**

4.1 Prevalence of HPV and Environmental Risk Factors Associated with Head and Neck Cancer 68
4.1.1 Demographic Characteristics of Study Subjects 68
4.1.2 The Association between Environmental Risk Factors and HNC 70
4.1.3 Prevalence of HPV in Head and Neck Cancer 72
4.1.3.1 PCR assay for detection of HPV 72
4.1.3.2 Association of HPV and survival of HNC patients 74
4.2 Genetic Polymorphisms of Carcinogen Metabolizing Genes in Head and Neck Cancer 75
4.2.1 Distribution of Genotypes of *GSTM1* and *GSTT1* in Cases, First-degree relatives and Controls 75
4.2.2 Association between Genetic Polymorphisms of *GSTM1* and *GSTT1* gene and HNC 76
4.2.3 Genetic Polymorphisms in *CYP1A1* Gene and HNC Risk 79
4.2.4 Gene-Gene and Gene-environment Interactions 81
4.2.4.1 Gene-Gene Interaction of *CYP1A1* T3801C, *GSTM1* and *GSTT1* Genes in the Risk of HNC 81
4.2.4.2 Combined effects of Smoking, betel quid and tobacco chewing and *GST* polymorphisms on HNC risk 82
4.2.4.3 Interaction of Smoking and Combined Genotypes of *CYP1A1* T3801C, *GSTM1* and *GSTT1* 84
4.2.4.4 Interaction of tobacco-betel quid and combined genotypes of *CYP1A1* T3801C, *GSTM1* and *GSTT1* 85
4.2.4.5 Multifactor Dimensionality Reduction (MDR) 86
4.2.4.6 Interaction Entropy Graphs 86
4.2.4.7 False Positive Report Possibility (FPRP) 86
4.3 Genetic Polymorphism of DNA Repair Genes (*XRCC1* and *XRCC2*) and risk of HNSCC 90
4.3.1 Distribution of the Genotype Frequencies of *XRCC1* Arg399Gln and *XRCC2* Arg188His 93
4.3.2 Association between *XRCC1* Arg399Gln Polymorphisms and HNSCC 94
4.3.3 Association between *XRCC2* Arg188His polymorphisms and HNSCC 94
4.3.4 Gene-gene interactions of *XRCC1* and *XRCC2* 94
4.3.5 Interaction of tobacco with *XRCC1* and *XRCC2* gene polymorphism 95
4.4 Epigenetic Alterations in Head and Neck Cancer 99
4.4.1 Frequencies of Promoter Hypermethylation in Tumour and Normal Tissues 100
4.4.2 CpG Island Methylator Phenotype in HNSCC Tumour Tissues 103

4.5 Correlation between HPV, Environmental and Genetic Factors with Promoter Methylation Profile of HNSCC patients 105
4.5.1 Promoter Methylation Status in HPV-positive and HPV-negative HNSCC 105
4.5.2 Correlation between Environmental and Genetic Factors and CpG Island Methylator Phenotype (CIMP) 107
4.5.3 Identified Tumour Clusters and Correlation with Environmental, Genetic and CIMP characteristics 111
4.5.4 Correlation between Survival data and promoter methylation profiles 115

CHAPTER-5: DISCUSSION **117-129**

5.1 Prevalence of HPV in Head and Neck Cancer 117
5.2 Genetic Polymorphisms of Metabolic and DNA Repair Genes and Risk of Head and Neck Cancer 119
5.3 Promoter Methylation Profiles of Head and Neck Cancer 125

SUMMARY **130-131**

BIBLIOGRAPHY **132-162**

APPENDIX **i-iii**

1. Annexure i
2. List of Publications ii-iii
   Published Research Articles

LIST OF TABLES

Table No.	Title	Page No.
Table 2.1	Estimated numbers of cancer new cases in males and females and world ASR per 100,000	29
Table 3.1	List of primers used for amplification of different genes	50
Table 3.2	List of primers used for methylation specific PCR (MSP)	51
Table 4.1.1	Demographic characteristics of patients, first-degree relatives and controls	69
Table 4.1.2	The comparison between patients, first-degree relatives and controls in terms of tobacco habits	71
Table 4.1.3	Adjusted and crude odds ratio of environmental risk factors for assessment of HNSCC risk	71
Table 4.1.4	HPV prevalence in HNSCC patients and healthy controls	73
Table 4.1.5	Characteristics of HPV-positive and HPV-negative samples	73
Table 4.1.6	Summary of survival data with respect to HPV	74
Table 4.2.1	Distribution of *GSTM1* and *GSTT1* genotypes among patients, first-degree relatives (FDRs) and controls	78
Table 4.2.2	Adjusted and crude odds ratio of risk factors (*GSTM1* and *GSTT1* gene) for assessment of HNC risk	78
Table 4.2.3	Distribution of *CYP1A1* (T3801C) genotype among the study population	79
Table 4.2.4	Combined genotypes of *CYP1A1* T3801C, *GSTM1* and *GSTT1* genes and HNC risk	81
Table 4.2.5	Combine effects of Smoking, betel quid and tobacco chewing and *GSTs* genes polymorphisms	83
Table 4.2.6	Interaction of smoking and combined genotypes of *CYP1A1* T3801C, *GSTM1* and *GSTT1*	84
Table 4.2.7	Interaction of tobacco-betel quid chewing and combined	85

genotypes of *CYP1A1* T3801C, *GSTM1* and *GSTT1*

Table 4.2.8 Summary of MDR analysis for HNC risk prediction 88

Table 4.2.9 False Positive Reports Probability (FPRP) for odd ratios of the Logistic Regression (LR) and Multifactor Dimensionality Reduction (MDR) analysis 89

Table 4.3.1 Genotype frequency distribution of *XRCC1* Arg399Gln and *XRCC2*Arg188His polymorphisms and risk of HNSCC 93

Table 4.3.2 Combined genotype analysis of *XRCC1* Arg399Gln and *XRCC2* Arg188His on risk of HNSCC 95

Table 4.3.3 Interaction of tobacco and *XRCC1* Arg399Gln polymorphisms 96

Table 4.3.4 Interaction of tobacco and *XRCC2* Arg188His polymorphisms 97

Table 4.4.1 Clinicopathological data of HNSCC tumour samples 99

Table 4.4.2 Frequency of methylation of tumour suppressor genes in tumour and normal tissues 101

Table 4.4.3 Frequency of methylation of tumour-specific loci in tumour and normal tissues 102

Table 4.5.1 Frequency of DNA methylation of genes/loci and CIMP status analyzed in HPV-positive and HPV-negative HNSCC 106

Table 4.5.2 Correlation between environmental factors and CIMP-status 108

Table 4.5.3 Correlation between genetic factors and CIMP-status 109

Table 4.5.4 Environmental characteristics of the two identified clusters 113

Table 4.5.5 Genetic, HPV and CIMP characteristics of the two identified clusters 114

Table 4.5.6 Summary of survival data of HNSCC patients in clusters and CIMP group 116

LIST OF FIGURES

Figure No.	Title	Page No.
Figure 1.1:	Diagrammatic representations of head and neck areas.	2
Figure 1.2:	Age-standardized incidence rates (per 100,000 men and women) of head and neck cancer in the world.	4
Figure 1.3:	Base excision repair (BER) pathway	17
Figure 1.4:	Schematic representation of DNA methylation.	19
Figure 1.5:	DNA methylation promotes events in tumourigenesis.	21
Figure 1.6:	Diagrammatic representation of proposed study from Northeast India	26
Figure 4.1.1:	Age and gender distribution among patients, first-degree relatives and controls	70
Figure 4.1.2:	Agarose gel showing HPV presence or absence.	72
Figure 4.1.3:	Kaplan-Meier survival plots for HPV	74
Figure 4.2.1:	Polymorphism of *GSTM1* and *GSTT1* metabolic genes.	76
Figure 4.2.2:	Ethidium bromide stained gel showing *CYP1A1* T3801C polymorphism.	77
Figure 4.2.3:	Frequency of genotype distribution of *GSTM1* and *GSTT1* in patients, first-degree relatives and controls	77
Figure 4.2.4:	Frequency of genotype distribution of *CYP1A1* T3801C in patients and controls	80
Figure 4.2.5:	MDR analysis for high-order interaction	87
Figure 4.2.6:	Summary of the four-factor model (*CYP1A1*, *GSTM1*, tobacco-betel quid chewing, and smoking) in MDR analysis	88
Figure 4.3.1:	Ethidium bromide stained gel showing *XRCC1* Arg399Gln DNA repair gene polymorphism.	90

Figure 4.3.2: DNA sequencing results of *XRCC1* (Arg399Gln) gene 91

Figure 4.3.3: Polymorphism of *XRCC2* Arg188His DNA repair genes. 92

Figure 4.3.4: Combined effect of tobacco habits with *XRCC1* Arg399Gln and *XRCC2* Arg188His genotypes 98

Figure 4.4.1: Agarose Gel Electrophoresis of MSP products 100

Figure 4.4.2: Frequencies of promoter hypermethylation in tumour and normal tissues 102

Figure 4.4.3: Methylation index (MI) in three CIMP-groups 103

Figure 4.4.4: Promoter methylation status of 10 genes in the HNSCC tumour tissues 104

Figure 4.5.1: Frequency of promoter methylation of 10 tumour-related genes/loci in HPV (+) and HPV (-) HNSCC 107

Figure 4.5.2: Methylation index (MI) stratified by genetic and habit related risk factors in HNSCC 110

Figure 4.5.3: Correlation between environmental, genetic factors and CIMP-status. 111

Figure 4.5.4: Hierarchical clustering and heatmap was constructed based on promoter methylation data of HNSCC in 7 tumour-related genes/loci. 112

Figure 4.5.5: Kaplan-Meier survival plot for CIMP-status 115

Figure 4.5.6: Kaplan-Meier survival plots for indentified Clusters 116

Figure 5.1: A mechanism of DNA damage response and cancer progression. 124

Figure 5.2: Proposed model for head and neck cancer development as evident from our present study 127

ABBREVIATIONS

AAR – Age Adjusted Incidence Rate

ASR – Age-Standardized Incidence Rates

Arg – Arginine

BER – Base Excision Repair

BLAST – Basic Local Alignment Search tool

BRCA1 – Breast Cancer – 1

BQC – Betel Quid Chewing

CDK – Cyclin Dependent Kinase

CDK2A – Cyclin Dependent Kinase Inhibitors – 2A

CIMP – CpG Island Methylator Phenotype

CVC – Cross Validation Consistency

CYP – Cytochrome P450

CYP1A1– Cytochrome P450, Family 1, Member A1

CYP2E1– Cytochrome P450, Family 2, Subfamily E, Polypeptide 1

DAPK – Death Associated Protein Kinase

DNA – Deoxyribose Nucleic Acid

DNMT – DNA Methyltransferase

dNTP – Deoxynucleotide Triphosphate

ECAD – E-Cadherinis

EDTA – Ethylene Diamine Tetrachloro Acetic Acid

EGFR – Epidermal Growth Factor Receptor

ESCC – Esophageal Squamous Cell Carcinoma

FDRs – First-degree Relatives

FFPE – Formalin Fixed Paraffin Embedded Tissues

FMO – Flavin-Containing Monooxygenases

FPRP – False Positive Report Probability

GAPDH – Glyceraldehyde 3-phosphate dehydrogenase

Gln – Glutamine

GST – Glutathione S-transferase

GSTM1– Glutathione S-transferase mu 1

GSTP1– Glutathione S-transferase pi 1

GSTT1 – Glutathione S-transferase theta 1

His – Histidine

HpaII – Haemophilus parainfluenzae II

HphI – Haemophilus parahaemolyticus I

HRR – Homologous Recombination Repair

HNC – Head and Neck Cancer

HNSCC – Head and Neck Squamous Carcinoma

HBCR – Hospital Based Cancer Registry

HPV – Human Papillomavirus

IARC – International Agency for Research on Cancer

IRB – Institutional Ethical Committee

LR – Logistic Regression

MDR – Multifactor Dimensionality Reduction

MI – Methylation Index

MINT1 – Methylated in Tumour – 1

MINT2 – Methylated in Tumour – 2

MINT31 – Methylated in Tumour – 31

MLH1 – MutL Homolog – 1

MSP – Methylation Specific PCR

MspI – *Moraxella species* I

NCRP – National Cancer Registry Programme

NER – Nucleotide Excision Repair

NNN – Nitrosonornicotine

NNK – 4 (Methylnitrosamino)-1,3-Pyridyl-1-Butanone

NPC – Nasopharyngeal Carcinoma

OPSCC – Oropharyngeal Squamous Cell Carcinoma

OR – Odds Ratio

OSCC – Oral Squamous Cell Carcinoma

PAH – Polycyclic Aromatic Hydrocarbon

RASSF1 – Ras Association Protein Family

RB – Retinoblastoma

RFLP – Restriction Fragment Length Polymorphism

ROS – Reactive Oxygen Species

SDS – Sodium Dodecyl Sulphate

SSC – Squamous Cell Carcinoma

TAE – Tris Acetic Acid EDTA

TBA – Testing Balance Accuracy

TE – Tris-HCl EDTA

TES –Tris-HCl EDTA Sodium Chloride

TGF –Transforming Growth Factor

TNM – Tumour Lymph node Metastasis

TrBA– Training Balance Accuracy

Trp –Tryptophan

WHO – World Health Organization

XRCC1– X-Ray Repair Cross-Complementing Protein 1

XRCC2 – X-Ray Repair Cross-Complementing Protein 2