CHAPTER 1

INTRODUCTION
1.1 Cancer

Cancer is a group of diseases due to unregulated cell growth and division caused by complex genetic, epigenetic and environmental interactions. Gene-environment interactions are thought to be mediated by epigenetic modifications of the genome, whereas, the epigenetic changes of the genome often arise in response to environmental factors. The gene-environment-epigenetic interactions are dynamic processes that lead to cascades of cellular events to facilitate the adaptation of an individual cell to its environment (Liu et al., 2008).

**Figure 1.1** Etiological factors of cancer: environmental, genetic and epigenetic interaction.

An inherent property of cancer cells is to spread to other parts of the body by lymphatic system or bloodstream (Zaman, 2013). However, all tumours do not spread throughout the body and are not cancerous. Only malignant tumours are capable of spreading to other parts in its advanced stages whereas benign tumours do not invade as well as spread to other parts of the body. Invading to nearby organs and spreading to other parts of the body is called metastasis, which is the most lethal property of cancerous cells. The newly spread tumours are called metastatic tumours, while the original is called the primary tumour. Usually all cancers can metastasize (spread from its primary site to other organs) and this characteristic is responsible for most
cancer deaths. The common phases in metastasis are local invasion, intravasation into the blood or lymph, circulation through the body, extravasation into the new tissue, proliferation, and angiogenesis (Hanahan and Weinberg, 2011).

![Figure 1.2](image.png)

**Figure 1.2** Staging of esophageal cancer according to the American Joint Committee on Cancer tumour-node-metastasis (TNM) classification (source: [http://www.clevelandclinicmeded.com/esophageal-cancer/](http://www.clevelandclinicmeded.com/esophageal-cancer/))

Defection of cancer is done by many ways like the presence of certain signs and symptoms, screening tests, or medical imaging. Initially cancers are recognized either by the appearance of signs or symptoms or through screening. The suspected cancer cases are investigated with medical tests which are usually blood tests, X-rays, CT scans and endoscopy. As, neither of these are conclusive diagnosis, examination of a tissue samples is required by a pathologist for confirmation of the disease. The histopathological diagnosis report of the tissue specifies the type of cell that is proliferating, its histological grade, histological abnormalities, and other features of the tumour. The histopathological analysis is useful to choose the best treatment strategy and to evaluate prognosis of the patients. Further information on cellular and chromosomal abnormalities can be gathered by performing
immunohistochemistry and cytogenetic analysis of the cancer tissue. These investigations may provide information about the probable cellular and molecular changes that has happened in the cancer cells, may consequently also point towards the future behaviour of the cancer (prognosis) and formulation of better treatment regime (Rathore et al., 2013)

Chemotherapy (anticancer drugs), radiation therapy and surgery are the basic treatment regime for cancers. The probabilities of cancer survival primarily depend upon the type and location of the cancer and the extent of disease at the start of treatment (Yim et al., 2008). All ages of people can be affected by cancer but few cancers are prevalent in certain ages and the overall risk of developing the disease usually increases with age (Wen et al., 2006). 11 million new cancer cases were estimated and 7 million cancer deaths were reported worldwide with almost 25 million people were living with the disease, in 2002 (Kamangar et al., 2006). The incidence rates are gradually increasing as more people live to an old age due to the development in medical science and immense lifestyle changes occur in the developing world (Kamangar et al., 2006).

The cancerous state of a cell involves multiple factors with complex interaction which are yet to be clearly understood. There are diverse factors responsible for the genesis of cancer including genetic, epigenetic and environmental (Talukdar et al., 2013). Genetic and epigenetic makeup defines the susceptibility of the cell towards cancer and its prognosis after treatment, whereas environmental factors play a crucial role in genesis of the disease. Tobacco consumption, certain infections, environmental pollutants, exposure to radiation, lack of physical activity, obesity, and dietary habits are among the known and established environmental and lifestyle related risk factors of cancers (Travis et al., 2013). These factors can individually damage the normal cellular process by directly affecting the genes or may combine with existing genetic defects of the cell to contribute in the development of cancer. Around 5–10% of cancers can be traced directly to inherited genetic defects. The environmental factors play an important role in the cancer development and progression and can be prevented by changing lifestyle and habits by not smoking, eating more vegetables, fruits and whole grains, eating less meat and refined carbohydrates, maintaining a healthy weight, exercising, minimizing sunlight exposure, and being vaccinated.
against some infectious diseases. The probable factors responsible for carcinogenesis in the Asian subcontinent especially north-eastern region of India are elaborately discussed further.

1.1.1 Oral Squamous Cell Carcinoma

Oral squamous cell carcinoma (OSCC) is an anatomically heterogeneous group of neoplasms arising from the mucosal surface of the oral cavity, oropharynx, hypopharynx and other sites within the upper aerodigestive tract. Countries with high tobacco and betel quid consumption such as Southeast Asian countries like India constitute nearly 25% of oral cancer patients (Gupta, 1999; Sinha et al., 2011). In fact, according to Gupta, oral cancer may now be assumed as a “new epidemic”, as incidence rates and prevalence are continuously increasing, possibly due to the common practice of consuming tobacco and areca nut products (Gupta, 1999). OSCC is usually the first or second most common site for malignant cancer in the Indian subcontinent (Elango et al., 2009). Age-adjusted rate of oral cancer is 20 per 100,000 population and accounts for over 30% of all cancers in India which is very high (Sankaranarayanan et al., 2013). The variation in incidence and pattern of the disease is probably due to the combined effect of ageing, as well as regional differences in the prevalence of disease-specific risk factors (Manoharan et al., 2010). OSCC diagnosis and treatment is a major concern for India as it is diagnosed at later stages resulting in low treatment outcomes. Moreover, rural areas in middle- and low-income countries have inadequate access to health services and limited trained clinicians. Therefore, this delay in diagnosis results to detection of OSCC at its advanced stages (Kumar et al., 2001). Early detection of the disease offers the best chance for better survival and potentiality to improve treatment outcomes and make healthcare affordable. It has become really important to enhance our knowledge of the disease aetiology and regional distribution of risk factors. Efforts towards early detection, prevention and treatment will only be possible, if the genetic, epigenetic and environmental aetiology of the disease can be deciphered. In this study, we focus on the epigenetic aberrations associated with major environmental risk factors in the high incident population of northeast India.
1.1.2 Esophageal Squamous Cell Carcinoma

Worldwide, esophageal squamous cell carcinoma (ESCC) varies greatly in its incidence rate and pattern, with around 100 folds difference among the areas of highest and lowest incidences within a distance of few hundred kilometres (Lambert and Hainaut, 2007). Although, the overall incidence of ESCC is low in western populations, Asian countries have an exceptionally high incidence rate. Among the Asian countries northern china, north-eastern Iran and India consists of the population which are at highest risk of ESCC and referred as the ‘Asian Belt’ of esophageal cancer. The incidence in these regions is often greater than 100 cases per 100,000 inhabitants per year (Lambert and Hainaut, 2007).

In recent years India has emerged as one of the high incidence regions. Incidence data from population-based cancer registries in India showed high prevalence of ESCC to occur in Northeast India (AAR ~12 per 100000) . The reason for this difference in incidence rate and pattern is not precisely clear. Previous studies on ESCC in Indian population specify the role of tobacco and betel quid (BQ) consumption in its carcinogenesis. Betel quid chewing with or without tobacco has been shown to be independently associated with the development of ESCC with clear dose-related responses that indicate a causal affect. Moreover, tobacco smoking and alcohol consumption are also strongly associated with ESCC (Coleman et al., 2012). Similar reports on the role of tobacco, betel quid and alcohol consumption in the incidence of ESCC were also reported from other parts Southeast Asia.

1.2 Aetiological Risk Factors of Cancer

Cancer is complex disease and multiple factors interact among themselves for the genesis of the disease. Usually genetic, epigenetic and environmental elements are the major risk factors for cancer (Talukdar et al., 2013). Therefore, in the current study we investigated the association of all these factors in the development of the disease.

1.2.1 Environmental Risk Factors

Exposure to various substances present in the environment may interfere with normal cellular metabolism and accounts for majority of the cancer cause worldwide. These environmental factors include different lifestyle habits like
tobacco consumption, betel quid chewing, excessive alcohol consumption, poor diet, lack of exercise, excessive sunlight exposure and sexual behaviour that increases exposure to certain viruses (Kaur et al., 2005; Koo and Ho, 1996).

1.2.1.1 Tobacco

There are more than 70 species of tobacco plants belonging from the genus Nicotiana within the Solanaceae (nightshade) family. There are many products manufactured from dried tobacco leaves which include cigars, cigarettes, snuff, pipe tobacco, chewing tobacco and flavoured shisha tobacco.

More than one billion people use tobacco in some form which is up to 1/3 of the adult population. Tobacco smoking rates have declined in developed countries, but continue to rise in developing countries. Moreover, powerful addictive properties of tobacco help to develop tolerance and dependence among its consumers. According to the World Health Organization (WHO), tobacco is the only greatest cause of preventable death globally. According to a WHO (2008), it was estimated that tobacco is responsible for 6 million deaths per year (Yach, 2014). The prolonged tobacco consumption may lead to certain diseases affecting lungs, liver and heart with smoking being a major risk factor for chronic obstructive pulmonary diseases, heart attacks, strokes, and cancer (particularly lung cancer, cancers of the larynx and mouth, and pancreatic cancer).

Tobacco is generally consumed in many forms and with number of different methods. Some common examples are following, but not limited to such forms and usage.

- **Bidi** is made of tobacco wrapped in a tendu leaf, and secured with coloured thread at one end, usually found common in India.

- **Chewing tobacco** is one of the oldest methods of consuming tobacco leaves. It is consumed orally mainly in two forms: through sweetened strands, or in a shredded form. While consuming the long sweetened strands, the tobacco is lightly chewed and compacted into a ball. When consuming the shredded tobacco, small amounts are placed between the gum and the teeth, at the bottom lip where it is gently compacted often called as *dipping tobacco*. Both of these methods stimulate saliva glands leading to the development of the spittoon.
• **Cigars** are tightly rolled bundles of dried and fermented tobacco, which is ignited so its smoke may be drawn into the smoker's mouth.

• **Cigarettes** are manufactured from cured and finely cut tobacco leaves and reconstituted tobacco, often combined with other additives, then rolled into a paper cylinder and consumed through inhalation of smoke by its user.

• **Creamy snuffs** are tobacco paste, consisting of tobacco, clove oil, glycerine, spearmint, menthol, and camphor often sold in a toothpaste tube. It is locally known as "mishri" in some parts of Maharashtra and marketed mainly to women in India.

• **Dipping tobaccos** are a form of smokeless tobacco. Dip referred to "chew", and is commonly confused with chewing tobacco, which encompasses a wider range of products. A dip is placed between the lower or upper lip and gums.

• **Gutka** is a prepared by crushing betel nut, tobacco, and sweet or savoury flavourings. It is manufactured in India and even exported to a few other countries. It is sold across India in small individual-size packets as a mild stimulant.

• **Hookah** is a single or multi-stemmed (often glass-based) water pipe for smoking. Hookahs were first used in India and Persia; it has also gained immense popularity, especially in the Middle East. A hookah operates by water filtration and indirect heat. It may be used for smoking herbal fruits, a mixture of tobacco, flavouring and honey or glycerine.

• **Kretek**s are cigarettes made with a complex blend of tobacco, cloves and flavouring "sauce". It was first introduced in the 1880s in Kudus, Java, to deliver the medicinal eugenol of cloves to the lungs.

• **Roll-Your-Own**, often called 'rollies' or 'roll-ups', are relatively popular in some European countries. It is prepared by the user, from loose tobacco, cigarette papers and filters all bought separately. They are usually cheaper to make.

• **Pipe smoking** usually comprises of a small chamber (the bowl) for the combustion of the tobacco to be smoked and a thin stem (Baasandorj *et al.*) that ends in a mouthpiece (the bit). Shredded pieces of tobacco are placed into the chamber and ignited to generate the smoke.

• **Snuff** is a crushed smokeless form of tobacco, inhaled or "snuffed" through the nose.
• **Snus** is a steam-cured moist powder tobacco product consumed by placing it in the mouth against the gums for an extended period of time. It is not fermented, and does not induce salivation.

• **Topical tobacco paste** is prepared by meshing an amount equivalent to the contents of a cigarette in a cup with about a 0.5 to 1 teaspoon of water to make a paste that is then applied to the affected areas as a treatment for wasp, hornet, fire ant, scorpion, and bee stings.

• **Tobacco water** or dust is used in domestic gardening as a traditional organic insecticide. It is produced by boiling strong tobacco in water, or by steeping the tobacco in water for a longer period. After cooling, the mixture can be applied as a spray, or ‘painted’ on to the leaves of garden plants, where it kills insects. Tobacco is however banned from use as pesticide in certified organic production.

**Figure 1.3.** Different forms of smoked and smokeless tobacco

**1.2.1.2 Alcohol**

Alcohol usually refers to ethanol or ethyl alcohol, which is a chemical substance found in beer, wine, and liquor. It is also constituent of some medicines, mouthwashes, household products, and essential oils (scented liquids taken from plants). Alcohol is produced by the fermentation of sugars and
starches by yeast. A standard alcoholic drink in the United States contains 14.0 grams (0.6 ounces) of pure alcohol, according to the National Institute on Alcohol Abuse and Alcoholism. Extensive research studies affirmed a strong scientific unanimity of an association between alcohol drinking and several types of cancer (Bergmann et al., 2013). The research evidence indicates the increased risk of developing an alcohol-associated cancer with more alcohol a person drinks regularly over time and is a known human carcinogen. A variety of carcinogenic contaminants may also be present in alcoholic beverages that are introduced during fermentation and production such as nitrosamines, asbestos fibers, phenols, and hydrocarbons. Based on data from 2009, an estimated 3.5 present of all cancer deaths in the United States (about 19,500 deaths) were alcohol related.

Several studies have detected the association between alcohol consumption and the risk of cancers like cancers of the pancreas, ovary, prostate, stomach, uterus, and bladder (Toskes, 2009). Moreover, Epidemiologic research shows that people who use both alcohol and tobacco have increased odds of developing cancers of the oral cavity (Petti et al., 2013), pharynx (throat), larynx, and esophagus than people who use either alcohol or tobacco alone.

The identified possible ways that alcohol consumption may increase the risk of cancer are as follows:

- Ethanol present in alcoholic drinks is broken down (metabolized) into acetaldehyde (a toxic chemical and a probable human carcinogen), which can damage both DNA and proteins
- Contribution in the generation of reactive oxygen species (chemically reactive molecules that contain highly reactive oxygen), capable of damaging DNA, proteins, and lipids (fats) by oxidation.
- It also hamper the body’s ability to break down and absorb a range of nutrients that may be associated with cancer risk, including vitamin A; nutrients in the vitamin B complex, such as folate; vitamin C; vitamin D; vitamin E; and carotenoids.
1.2.1.3 Viral aetiology

The first known tumorigenic virus in birds was discovered by Peyton Rous in 1911. Later on in 1964, Michael Anthony Epstein and Yvonne Barr first discovered a virus identified to be involved in human cancer and named it Epstein–Barr virus (Lefebvre et al.). It is well established that viruses infect different kinds of human cells and cause a range of diseases, from the common cold to AIDS and cancer. The disease causing mechanism of viruses varies depending on the type of virus and infected cell.

Certain viruses are capable of persisting over time, despite active immune system of the host cell without causing any apparent changes to the cell. This characteristic of latency is the general property of herpes virus group, among which herpes simplex virus can induce cell proliferation without causing malignancy (Whitley et al., 2007), or several members of the human papilloma virus (HPV) group can contribute in the genesis of cancer (Yugawa and Kiyono, 2009). The involvement of many viruses is high in certain cancer types, like human hepatitis B virus (HBV) and human hepatitis C virus (HCV) are associated with 80% of hepatocellular carcinomas (HCCs), HPV is positive in 49.5% of cervical carcinomas (zur Hausen, 2009a) and EBV is associated with 30% of Hodgkin’s lymphomas and nasopharyngeal cancer. The proteins expressed from these oncogenic viruses (cancer causing viruses) are capable of disrupting different crucial cellular pathways, such as cell-cycle checkpoint activation and apoptosis (McLaughlin-Drubin and Munger, 2009). However, viruses alone cannot be responsible for carcinogenesis and additional factors are vital for the development of cancer. These factors include evasion from host immune system and cellular aberrations like mutations are also very essential for the initiation of tumour process (McLaughlin-Drubin and Munger, 2008).

Latent infection producing viruses usually evade recognition by the immune system to avoid its elimination by the host cells. Many of these evasion strategies have been identified with a common mechanism of camouflaging the virus in the host cell for restricting the expression of viral genes and proteins that are vital for viral persistency, and avoiding the expression of genes associated with host immune response. Changes in the DNA methylation pattern of viral genome is proposed to be one of the masking mechanism by which many viruses
are able to evade host immune response (Fernandez et al., 2009). It is also proposed to be one of the mechanisms for genome-defence system to prevent translocations, chromosomal instability and gene disruption caused by the reactivation of the oncogenes or transposable DNA sequences (Rollins et al., 2006; Yoder et al., 1997). Moreover, eukaryotic cells have developed several defence mechanisms against the uptake, integration and continued expression of foreign DNA (such as viruses) where gene-specific methylation plays an important role (Doerfler, 1991). According to recent findings, viruses can influence methylation of the host genome by interacting with the epigenetic machinery of the host cell resulting in the silencing of vital host cellular genes. Viruses are also capable of altering the activity of proteins related with the specific histone marks, chromatin remodelling complexes and miRNA processing (Javier, 2008). It is well established fact that viruses are capable of destabilizing the host genome by insertion of mutations and chromosomal rearrangements inclining the virus infected cells to develop cancer. Moreover, along with influencing genetic changes, viral infection is also associated with aberrant methylation and repression of crucial genes involved in cancer prevention.

1.2.1.3.1 Role of Human Papillomavirus (HPV) in cancer

The relationship between papillomavirus infections and cervical cancer was first established in the 1970s (Gissmann et al., 1977). Since then, many studies have established that Human Papillomavirus (HPV) can contribute in the carcinogenesis of many cancers including head and neck, esophageal and colorectal cancers in addition to cervical cancer (zur Hausen, 2009b). HPVs are currently assumed to be the most common cause of cancers induced by an infectious agent (Whiteside et al., 2008; zur Hausen, 2009b). HPVs are a group of viruses of the Papillomaviridae family. HPV genome contains double-stranded circular DNA of nearly 8kb and like other viruses depends on the host DNA machinery to replicate and proliferate. Typically, the HPV genome is divided into three regions: an upstream regulatory region or long control region (LCR); also called as the early region, consisting of six open reading frames (ORFs) known as E1, E2, E4, E5, E6 and E7; and the ‘late region’, with two ORFs coding for viral structural proteins L1 and L2 (Zheng and Baker, 2006). The
virus mainly infects the squamous epithelia of many species including humans and 4100 papillomavirus genotypes have been identified in humans. They primarily differ in sequence among themselves by at least 10% within the most conserved region (L1) (de Villiers et al., 2004; zur Hausen, 2006). Although HPVs are capable of producing a range of benign proliferations, it is assumed that there are potential five high risk cancer-causing strains of HPV. Among these, HPV16 and HPV18 are the most prevalent types and infection with any two of the high-risk mucosal HPVs causes squamous intraepithelial lesions which can progress to invasive squamous cell carcinoma. These two stains are almost associated with >95% of cervical and >70% of anal cancers (zur Hausen, 2006).

The most widely studied viral genes are E2, E6 and E7 because of their associations with oncogenesis. The E2 gene product is involved in viral DNA replication and the regulation of early transcription, whereas, E6 and E7 are considered as genuine viral oncogenes and their expression induces cell immortalization and transformation. The HPV genome often integrates into the host cell chromosome during the carcinogenic transformation of the host cell.

**Figure 1.4** HPV infections in the squamous epithelial cells. [Source: (Moody and Laimins, 2010)]

In this approach, the virus can persist in the host cell influence oncogenesis simultaneously it also completes its life cycle. This integration of viral genome often disrupts the E1–E2 region, resulting in a loss of expression of the E2 viral protein, which is a transcriptional repressor of E6 and E7. The decreased
expression of E2 protein results to an increase in expression of both oncoproteins E6 and E7. E6 and E7 inactivate p53 and pRb, the two crucial tumour-suppressor genes of cell cycle (Zheng and Baker, 2006) and directly destabilizing the regulation of the cell cycle of the host cell. Other studies have also revealed that E6 and E7 bind to other proteins involved in the processes vital cellular characteristics such as adhesion, apoptosis, cell cycle, DNA repair, metabolism, signal transduction, gene expression and other functions (Horikawa and Barrett, 2003; Howie et al., 2009; McLaughlin-Drubin and Munger, 2009; Whiteside et al., 2008). However, HPV is not only capable of introducing genetic changes but also assumed to be a potent factor for epigenetic reprogramming by aberrant DNA methylation, histone modifications and miRNA expression, finally contributing in the oncogenesis. Extensive screening of DNA methylation in premalignant and invasive cervical carcinoma samples associated with HPV is proposed to be a potential biomarker for early detection (Bergeron et al., 2014; Mir et al., 2014; Whiteside et al., 2008). Moreover, HPV is also associated with other gastrointestinal tract cancers like oral, laryngeal, esophageal and anal cancers but the role and exact mechanism of carcinogenesis is still remained unresolved.

1.2.1.4 Additional Factors

Along with the factors discussed above, there are multiple risk factors associated with these which interact and contribute greatly in the cancer development. These factors include dietary habits, physical activity, suitable hygiene etc. (Sinha et al., 2011). Eating habit are also very crucial as consumption of fibrous fruits and vegetables have showed to reduce the risk of cancer development. Moreover, consuming macro and micronutrients in the diet also enhance cellular metabolism as many of these nutrients acts as cofactors for most of the metabolic enzymes. Sufficient dietary supplements of these nutrients are essential for proper functioning of such enzymes.

1.2.2 Genetic aetiology

Cancer mainly arises due to the unregulated cell and tissue growth. The genes which regulate cell growth and differentiation are very crucial and must be altered for the transformation of the normal to cancerous cell. These crucial cancer related genes are divided into two broad categories namely oncogenes and tumour suppressor genes (TSGs). Oncogenes promote cell growth
and reproduction whereas TSGs inhibit cell cycle, division and survival. Usually in normal cells the expression of oncogenes and TSGs are very finely regulated. Over-expression of oncogenes and under-expression of TSGs generally deregulate cell cycle and finally contribute in the transformation into cancerous cells. These genetic chances may occur by different mechanisms, like gain or loss of an entire chromosome can occur through errors in mitosis or by mutations which are changes in the nucleotide sequence of genomic DNA. These abnormalities may occur in the promoter region of a gene and affect its transcription, or may occur in the coding region of the gene resulting in the alteration of function or stability of the gene product.

During the course of cell cycle for cell growth and proliferation, massive amount of DNA is replicated with possible chances of mistakes (mutations). Moreover, there are mechanisms within the cell for correction and prevention of these errors and safeguard the cell against disorders like cancer. If these errors are significant and are irreparable, then the damaged cell can "self-destruct" by programmed cell death called as apoptosis. In case all these error control mechanism fail, then the error will persist and will be passed along to daughter cells after each cell division. The successive cell division of these erroneous cells results in the accumulation and propagation of genetic errors/mutations to the next cell generations. These accumulated genetic aberrations finally disrupt the regulation of cell cycle and ultimately lead to cancer.

1.2.2.1 Carcinogen Metabolizing Genes

Cancer susceptibility varies among individuals based on carcinogen metabolizing ability. Carcinogens are detoxified by many intracellular reactions in different phases catalysed by specialized enzymes. One such class of enzymes are glutathione S-transferase (GST) family gene products that are mainly phase-II detoxification enzymes playing critical role in carcinogen breakdown. Human GSTs are divided into eight distinct classes as alpha, kappa, mu, omega, pi, sigma, theta, and zeta based on the similarity of amino acid sequence and antibody cross-reactivity (Strange et al., 2001). The chromosomal gene location of *GSTM1* and *GSTTI* are on chromosome 1p13.3 and 22q11, respectively. *GSTM1* and *GSTTI* genes belong to the glutathione-S-transferases superfamily of genes, whose products are Phase II metabolizing enzymes. These enzymes act in
coordination with Phase I metabolizing enzymes in the process of carcinogen breakdown. The Phase I enzymes generally stimulate the carcinogens to reactive intermediates and the GST enzymes actively detoxify a wide variety of these potentially toxic and carcinogenic electrophiles by conjugating them to Glutathione (Strange et al., 1998).

Homozygous null alleles (deletion) of GSTM1 and GSTT1 genes will result in the lack of activity of the respective enzymes with consequent impairment of carcinogen metabolism (Bell et al., 1993). Altered frequencies of GSTT1 and GSTM1 genes are reported to modify the effect of tobacco exposure thereby increasing the susceptibility for developing cancer (Gertig et al., 1998). Many studies also found significant association between GSTM1/GSTT1 null genotypes and other lifestyle related risk factors like tobacco and betel quid chewing, smoking, and alcohol consumption in Indian population especially from the state of Assam (Mondal et al., 2013; Sharma et al., 2013).

1.3 Epigenetics

The cells in a multicellular organism usually harbour identical genetic setup or DNA sequences and consequently the similar genetic instruction sets, yet maintain diverse terminal phenotypes. This cellular memory is not regulated by genetics, rather it is epi-(above)–genetics which records developmental and environmental signals of a cell and decides which set of genes to be expressed, when to express and how much to express. Epigenetic regulations are crucial factors for a change in phenotype without a change in the genotype.

The term “epigenetics” was initially coined by Conrad Waddington to describe the interface between genes and their products that bring the phenotype into existence (Waddington, 2012). Every cell in a human body contains the same genetic material, but they do not look or behaves the same. Each cell retains its exclusive properties while it has the same master set of genes in its DNA as every other cell in the body. This is possible by the regulation of gene expression as; cells from different tissues express distinct set of genes during the course of their differentiation.
Figure 1.5 DNA methylation influences the way that genes are expressed without changing the underlying DNA sequence, and other epigenetic factors bind to histones to control when chromatin complexes open up and allow their DNA to be read (Marx, 2012).

Modifications involved in the packaging of genetic material define the accessibility of any set of genes for expression, without altering the DNA sequence termed as epigenetic modifications (Marx, 2012). These epigenetic changes persevere through numerous cell divisions to maintain a definite gene expression pattern that contributes to the cells identity. The cell identity changes during cellular differentiation finely regulated by epigenetic modifications which are crucial to guide a totipotent to a fully differentiated cell (Jones, 2007). Individual epigenetic signature determines the fate of undifferentiated cell to a fully differentiated cell. Each cell type in an organism has its own specific epigenetic signature that is affected by its genotype, developmental and environmental history and in the end leads to the phenotype of the organism (Osley, 2008).
Epigenetic alterations involve functionally relevant modifications to the genome without any change in the nucleotide sequence. These variations may persist through cell divisions, can last for multiple generations, and are considered to be epimutations (equivalent to mutations).

This heritability of gene expression patterns is mediated by epigenetic modifications and most of the times these modifications regulate gene expressions by changing DNA methylation (hypermethylation or hypomethylation) pattern, histone modifications and altering the chromosomal architecture. Each of these epigenetic alterations can regulate gene expression without altering the underlying DNA sequence.

The complement of these alterations conjointly called as the epigenome, providing a mechanism for cellular diversity by regulating what genetic information can be accessed by cellular machinery. Inability to obey and maintain these heritable epigenetic marks can result in inappropriate activation or inhibition of different signalling pathways and can result into disease states such as cancer (Egger et al., 2004). The importance of epigenetic gene silencing in cancer is also emphasized by the growing cognizance that such changes can actually predispose to permanent gene silencing or mutational events during tumour progression (Sharma et al., 2010).

The widespread distribution of hypermethylated genes across the human genome, and the evaluation of hypermethylated candidate cancer-driving tumour-suppressor genes have prompted many efforts to screen the cancer-cell genome...
for such genes (Egger et al., 2004). Epigenetic as well as genetic variations are crucial for not only the initiation but also maintenance of many human cancers (Sharma et al., 2010). Under the stressful conditions, a complex series of chromatin modification events arise due to any injury or other factors, that might ‘lock in’ abnormal heritable transcriptional repression of key genes, or even networks of genes by epigenetic mechanisms. These mechanisms could devotee cells to dependence on inappropriate activation or disruption of crucial cellular pathways. These aberrations can be considered as an ‘epigenetic gatekeeper’ step that creates a background that enables the selection of mutations in oncogenes and tumour-suppressor genes leading to tumour progression (Klein, 2005).

Understanding the molecular events for initiating and maintaining epigenetic gene silencing can lead to the development of clinical strategies for cancer prevention and formulation of improved therapy that reverse the silencing process (Egger et al., 2004). The chromatin alterations involved in epigenetic regulations like abnormal promoter methylation, can be evaluated in DNA from different biological sources for formulating promising molecular-marker to support cancer risk assessment, early detection and prognosis (Brower, 2011).

1.3.1 DNA Methylation

The covalent addition of a methyl (CH3) group at the 5th-carbon of the cytosine ring resulting in 5-methylcytosine (5-mC) referred as DNA methylation. These methyl groups remain projected into the major groove of DNA leading to transcription inhibition. In human DNA, 5-methylcytosine is found in approximately 1.5% of genomic DNA. Addition of a methyl group to the cytosine adjacent to guanine residue, linked with phosphate bond are generally known as CpG Islands (CGI) which are the principal sites for methylation in eukaryotes (Jones and Takai, 2001).

DNA methylation is the most extensively studied epigenetic modification in mammals. CGI are distinct clusters of CpG dinucleotide that are on average 1 kb long and are often found in the 5’ region of genes and almost more than half of total human genes have CGI adjacent their 5’ ends (Bird, 1986). Cytosine methylation of CpG sites in promoter-associated CpG islands is found involved in the allele-specific transcriptional inactivation of imprinted genes and genes on the inactive X-chromosome (Barlow, 1995; Sado et al., 2004). DNA Methylation
pattern is established and maintained by a family of DNA-methyltransferase enzymes (DNMTs) DNMT1, DNMT3a and DNMT3b (Kaneda et al., 2004). Two additional enzymes DNMT2 and DNMT3L may also have more specialized and related functions. DNMT1 has a high affinity for hemi-methylated DNA and is assumed to be important for maintenance of methylase activity.

![DNA methylation patterns in normal and cancer cells](image)

**Figure 1.7** DNA methylation patterns in normal and cancer cells (Gal-Yam et al., 2008)

The aberrant methylation on CGI provides a stable mechanism for gene silencing which plays an important role in regulating gene expression and chromatin architecture, in association with histone modifications and other chromatin associated proteins (Eden and Cedar, 1994; Takai and Jones, 2002). 5-mC as a major epigenetic modification for phenotype and gene expression has been widely recognized with great biological importance. Decrease in global DNA methylation has been proposed as a molecular marker in cancer and multiple biological abnormalities (Rideout et al., 1990).

There are multiple significant functions of CGI methylation as it is one of the important mechanisms to regulate the expression of genes involved in imprinting, pluripotency, cell differentiation and proliferation etc (Crea, 2011; Leung et al., 2014; Rutledge et al., 2014). DNA methylation also facilitates histones in chromatin compaction which is important and decisive for activation or suppression of gene expression (Leung et al., 2014). Moreover, it is essential for binding of DNA on nucleosome structures and also the methylation or
demethylation states assists multiple DNA binding proteins during the regulation of transcription (Miller and Grant, 2013).

![Biochemical Level](image1)

**Figure 1.8** Role of DNA methylation in different levels of eukaryotes (Franchini et al., 2012)

### 1.3.2 Other Epigenetic Modifications:

Other epigenetic modifications include histone modification or chromatin remodelling which regulates chromatin architecture and miRNA mediated gene silencing which usually take place in the post-transcriptional stage. Usually the amino acid residues present in the histone proteins are basic in nature and the amino terminals of the core histone are subjected to several types of multivalent modifications, including acetylation, methylation, phosphorylation, ubiquitination, sumoylation, etc. (Satpathy et al., 2014; Sinha et al., 2010; Yu et al., 2008). Histone modifications takes place during post-translational modifications of histone proteins which are critical for regulating chromatin structure and function and affects many DNA-associated processes, such as transcription, recombination, DNA repair and replication, and chromosomal organization (Miller and Grant, 2013; Rhie et al., 2014).
Micro RNAs (miRNAs) mediated regulation of gene expression involves translational repression and/or messenger RNA (mRNA) deadenylation and decay. This form of regulation is post-transcriptional control of gene expression and is also very crucial for switching of a transcriptionally active gene (Rudnicki et al., 2014).

1.4 CpG Island Methylation Phenotype (CIMP)

Aberrant CpG island (CGI) hypermethylation is established as one of the hallmark of cancer and is characterised by tumour-specific hypermethylation of multiple CGIs (Costello et al., 2000). A report from the year 1999, revealed identification of two distinct subgroups of colorectal cancers exhibited low and high levels of tumour-specific methylation, respectively. The latter group of tumours, with higher levels of hypermethylation was referred to as exhibiting a “CpG island Methylator Phenotype” (CIMP) (Toyota et al., 1999a). The characteristic of CIMP group was defined to be concordant tumour specific DNA methylation and could be clearly distinguished due to displaying higher CGI methylation levels in comparison to non-CIMP tumours that showed only low levels of tumour-specific methylation. These findings delineate a distinct subgroup of CIMP positive cancers with a characteristic hypermethylation in multiple tumour suppressor genes assumed to be involved in the evolution of cancer (Issa, 2004). In other words, CIMP refers to the concept that a subset of tumours has extensive methylation of CpG islands leading to epigenetic inactivation of tumour suppressor genes by promoter methylation.

After the first report, CIMP was also demonstrated in several other cancers, (Garcia-Manero et al., 2002; Hoque et al., 2006; Strathdee et al., 2001; Yuen et al., 2010) and confirmed the original findings using similar markers and technology (van Rijnsoever et al., 2002; Whitehall et al., 2002). However, some other findings do not support the same (Eads et al., 2001; Esteller et al., 2000). However, Genome wide scan for changes in DNA methylation has also reconfirmed the presence of a hypermethylator phenotype in various human cancers (Costello et al., 2000). Thus, CIMP appears to be involved in an important new tumourogenesis pathway that leads to cancer progression by simultaneously inactivating multiple genes.
Chapter 1  **Introduction**

Most of the CIMP studies were done using distinct panel of markers and with diverse criteria to define CIMP. To evaluate CIMP status without genome-wise screening, a CIMP panel consisting of *CDKN2A* *or* *p16*, MINT1, MINT2, MINT31and *MLH1* selected markers were used also called as “classical” CIMP panel to distinguish high and low level methylation (Park *et al.*, 2003). Another CIMP panel was proposed with different set of markers referred as “New” panel consisting of *CACNA1G, IGF2, NEUROG1, RUNX3* and *SOCS1* genes based on stepwise screen of 195 markers (Weisenberger *et al.*, 2006). Although screening of CIMP status rely on methylation markers and criteria used, each of these authors claimed their CIMP marker panel to exhibit better efficiency for distinguishing CIMP in cancers.

The exact molecular mechanism responsible for causing CIMP positive tumours is not precisely known; however, CIMP-positive cancers exhibit different genetic and clinical features, probably due to their distinct tumorigenic pathway. With the reports of CIMP in other cancers like lung, liver, gastric, ovarian, etc. it has been realized that CIMP can play a vital role in other cancer types (Shen *et al.*, 2002; Strathdee *et al.*, 2001; Toyota *et al.*, 1999b). These findings indicate that CIMP is not restricted to specific tumour types, instead consistent aberrant DNA methylation can be a usual phenomenon in cancer development, however may involve different genes depending on the tumour type.

According to some studies, CIMP-high tumours are found exhibiting worse prognosis probably due to their increased epigenetic plasticity (Arain *et al.*, 2010). Therefore it will be significant to identify the molecular source of CIMP and also evaluation of the existing set of methylation markers for detection of CIMP. However, CIMP-high tumours can be easier to diagnose at an early stage since aberrant DNA methylation can be detected with high sensitivity. Moreover, they are usually susceptible to demethylating agents as DNA methylation is reversible in nature. There is also a possibility that certain DNA damaging drugs themselves may induce CIMP during chemotherapy, which an important concern needed to be resolved (Huh, 2011). Undoubtedly, further extensive research in this field is required to reach a conclusion, but this is an issue with potentially huge implications for current chemotherapy regimes.
1.5 Statement of the Problem

Northeastern part of India is among one of the high incident region of OSCC and ESCC. According to both population-based and hospital-based cancer registries in India, it was observed that the highest incidence of esophageal cancer occurs in Assam (Nandakumar et al., 2005). The specific etiology of this high incidence in this region is not precisely known.

Interestingly, the population under study is with typically distinct lifestyle, where tobacco consumption habit both smoked and smokeless forms are very common practice in this region. Other than tobacco, chewing of betel quid is customary among the population (Lee et al., 2011; Phukan et al., 2001b). Moreover, there are reports which assume that tobacco is a potent modulator of DNA methylation. Therefore, association between tobacco consumption (smoked and smokeless) and promoter methylation of tumour-related genes/loci (CIMP) in oral and esophageal cancer is essential to be studied.

1.5.1 Rationale of the Study

Oral and esophageal squamous cell carcinoma is amongst the most fatal forms of cancers in the developing countries especially in India, with high incidence and mortality rates. These malignancies are assumed to develop as a consequence of complex interactions between environmental, genetic and epigenetic factors. However, these interactions are not well understood in oral and esophageal squamous cell carcinoma. According to the recent research, epigenetic events like aberrant DNA methylation are considered as important factors in development, progression and even prognosis of various gastrointestinal tract cancers like gastric and colorectal cancer. Based on methylation patterns of CpG islands, distinct tumour sub-groups have been identified in such cancers, the subgroup with high level of methylation was referred to as exhibiting a “CpG island methylator phenotype” (CIMP). The CIMP high subgroups exhibited distinct clinicopathological, genetic and prognostic characteristics. However, only a few studies are available on OSCC and almost no significant study was done on ESCC. The experimental evidences in this direction are also elementary and require in-depth studies and extensive validation of the importance of DNA methylation in the genesis of OSCC and ESCC.
Furthermore, evidences are growing that several environmental risk factors like tobacco smoke associated carcinogens and carcinogen metabolising gene polymorphisms are capable of modulating DNA methylation in cultures, animal models as well as certain tobacco-related cancers like lung cancer (Jin et al., 2010; Lin et al., 2010; Mani et al., 2012; Pulling et al., 2004). Cigarette smoke has also been found to induce promoter methylation of certain genes in esophageal epithelial and ESCC cell lines, however, no study involving human subjects were carried out (Huang et al., 2011; Meng et al., 2012). Further, null genotype of GSTM1 gene was associated with an increased susceptibility of CpG island hypermethylation in gastric-mucosa (Tahara et al., 2011). Although, OSCC & ESCC are amongst the most important tobacco-related cancers, but the interaction of smoked and smokeless tobacco, carcinogen metabolizing gene polymorphisms and aberrant DNA methylation in these cancers has remained largely unexplored. Furthermore, most of the studies considered only few genes which were tested on models/cultures or a very limited number of cases. Hence, detailed studies with extensive validation in large number of samples are lacking.

The importance of the current study is further enhanced by the fact that, it is conducted on a unique population of Northeast-India, where tobacco related habits like tobacco chewing; beedi and cigarette smoking are common. Moreover, consumption of a combination of areca nut, betel leaf, slaked lime with or without tobacco, called ‘betel quid’ or locally as ‘pan’ or ‘tambul’ is customary in this concerned population. The Assam and Mizoram states of NE-India are amongst the highest incidence region of esophageal cancer, with an age-adjusted rate of around 17/100000 to 27 per/100000 population (2010). Although, previous studies on the risk factors of ESCC in Northeast-Indian population specify the association of tobacco and betel quid chewing with its carcinogenesis, but very little is established about the environmental, genetic or epigenetic risk factors (Phukan et al., 2001a). Moreover, no studies were conducted on DNA methylation signatures of the O&ESCC patients in this population.

In addition, recent findings also associated HPV as a risk factor with a probable causal role in oral and esophageal squamous cell carcinoma (Crow, 2012) and studies have indicated towards a mechanism involving aberrant DNA methylation of host genome in response to HPV infection (Fertig et al., 2013;
Worsham et al., 2013a). HPV positive (+) and negative (-) host genome exhibits large differences in their methylation pattern interfering key cellular pathways (Lechner et al., 2013; Worsham et al., 2013b). Although, HPV presence is known as a predictive marker for better survival in oropharyngeal cancers (Petrelli et al., 2013; Tahtali et al., 2013; Worsham et al., 2013c), its prognostic role in oral and esophageal cancers are not well studied.

![Diagram](image.png)

**Figure 1.9** Schematic representation of the proposed study

Hence, keeping in view the above mentioned status of the disease, the present study is important to understand and establish the CpG island methylation status of squamous cell carcinoma of oral cavity and esophagus and its probable modulation by various environmental carcinogens.

### 1.5.2 Novelty of the Study

As per our knowledge, this is the first study defining CIMP in ESCC as no such study conducted on ESCC previously. This study might be among one of the precursor works in assessing the role of smokeless tobacco and HPV in modulating aberrant DNA methylation, responsible for epigenetic origin of OSCC and ESCC. In addition, this might be the first study to determine the crucial epimutations accountable for cancer development and progression in clinical samples from a high incident region of India.
1.6 Specific Research Objectives

- To study the association of tobacco and alcohol related habits and genotyping of carcinogen metabolizing genes in oral and esophageal squamous cell carcinoma development.

- To study the CpG Island methylator phenotype (CIMP) in oral and esophageal squamous cell carcinoma patients and its association with habit related carcinogen exposures.

- To study the association of CpG Island methylator phenotype (CIMP) and response to treatment in these cancer patients.

- To study the association of viral infections and CIMP status in the cancer patients.