Chapter 2

Background and Objectives
2.1 ORIGIN OF THE STUDY

Cervical cancer (CaCx) is the second most frequent cancer among women in urban India after breast cancer and ranks first among women in rural India. India has a population of approximately 432.2 million women above 15 years of age, who are at risk of developing CaCx. The current estimates indicate approximately 122,844 new cases diagnosed and 67,477 deaths annually in India, accounting to nearly 1/3rd of the global CaCx deaths. Indian women face a 2.4% cumulative lifetime risk and 1.4% cumulative death risk from CaCx. At any given time, about 5% of women in the general population are estimated to harbour cervical HPV infection. HPV types 16 and 18 account for nearly 82.7% of CaCx incidence in India. Unlike many other cancers, CaCx occurs early and strikes at the productive period of a woman's life. The incidence rises in 30–34 years of age and peaks at 55–65 years, with a median age of 38 years (age 21–67 years). Estimates suggest that more than 80% of the sexually active women acquire genital HPV by 50 years of age (Bruni et al, 2015). Persistent infection in the cervical mucosal epithelium with oncogenic HPV (mainly HPV type 16/18) is the major etiology of the disease. The viral oncoproteins E6 and E7 bind, degrade and inactivate various host cell regulatory proteins, especially the tumor suppressor (p53) and retinoblastoma (pRb) proteins. As a result, alteration in cell cycle regulatory pathways causes increased cell proliferation and genomic instability thereby inducing neoplasia and hence, viral propagation. However, HPV is a necessary cause of CaCx, but it is not a sufficient cause for CaCx development. Other cofactors are necessary for progression from cervical HPV infection to cancer. Long-term use of hormonal contraceptives, high parity, early initiation of sexual activity, multiple sex partners, tobacco smoking and co-infection with HIV, have been identified as established cofactors. Some other probable co-factors are co-infection with Chlamydia trachomatis and herpes simplex virus type-2, immunosuppression, low socioeconomic status, poor hygiene and diet low in antioxidants. Such persistent infections highlight the role of viral factors such as viral load, viral integration status, viral oncogene expression and genetic and epigenetic host factors in driving carcinogenesis, but the involved mechanisms are still not elucidated completely.
2.2 BACKGROUND OF THE STUDY

2.2.1 Viral factors necessary for cervical cancer pathogenesis

(A) Impact of Human Papillomavirus 16 (HPV16) load on cervical cancer development: CaCx is believed to have a co-factorial etiology in which, high-risk HPV infections are considered an essential factor and other elements play an ancillary role. Besides the importance of specific HPV genotypes, other viral cofactors such as viral load, may influence the progression likelihood. Several studies have suggested that a high HPV-DNA viral load may be a candidate marker that could help identify women at greater risk of CIN progression. In fact, some studies have pointed out that high viral load in cytological normal epithelium could also be a risk factor for neoplastic progression (Josefsson et al., 2000; Swan et al., 1999; van Duin et al., 2002; Ylitalo et al., 2000).

Many study groups have proposed viral load estimates per cell or per unit amount of genomic DNA, as a potential HPV-related biomarker, which could be used for predicting those at risk of CaCx development (Franco and Coutlée, 2009). However, there are several reports, which have failed to relate high HPV16 DNA copy number with CaCx development (Swan et al., 1999; Josefsson et al., 2000; van Duin et al., 2002; Hernandez-Hernandez et al., 2003, Abba et al., 2003; de Boer et al., 2007). Studies from our laboratory (Das et al., 2010), identified a significantly higher viral load among the CaCx cases as compared to the HPV16 positive non-malignant samples. However, the impact of viral load on the molecular progression of CaCx has not been elucidated till date.

(B) Role of HPV16 integration on cervical cancer pathogenesis: It is known that viral genome in episomal form replicates along with the differentiating epithelial cells from basal membrane to the superficial zone and is shed off along with the sloughed-off epithelial cells resulting in transient infection (zur Hausen, 2002). Persistent infection and viral genome integration into the host genome is known to mediate oncogenicity (Woodman et al., 2007). E1 and E2 are the early viral proteins needed for viral replication and translation, while E6 and E7 are the oncoproteins responsible for cellular transformation by inactivation of p53 and pRb proteins, respectively. E2 protein represses E6 and E7 expression. Integration of the viral genome into the host genome, chiefly at fragile sites (Kalantari et al., 2001; Wentzensen et al., 2004), not only affects various cellular pathways of the host cell-cycle machinery, but also disrupts the viral E2 gene most commonly in the hinge region of the...
HPV16 E2 protein. In absence of E2-driven repression, E6 and E7 are over-expressed driving infected cells toward transformation.

On the contrary, studies from our laboratory (Bhattacharjee et al., 2006) as well as a few others (Narayanan et al., 2004) have identified that a substantial proportion of individuals with CaCx harbour intact E2, which could be either purely intact or concomitant, i.e., a mixture of intact and disrupted forms (Figure 3.1). Such observations, point towards the biological plausibility of cervical carcinogenesis under the impact of HPV16 intact E2 gene or intact viral genomes, in addition to E2 disruption due to viral genome integration into the host genome.

![Diagram](image)

**Figure 2.1:** Diagrammatic representation of the various forms of existence of HPV genome in infected cells

**Role of HPV16 oncoprotein E7 in cervical cancer pathogenesis:** HPV16 encoded oncoprotein E7, plays a crucial role in the transformation of the epithelial cells leading to development of cancer. E7 is a well-studied oncoprotein and its role in oncogenesis by associating with the PDZ domain harbouring cellular proteins and pRb has been well established (Collins et al., 2005). However, such viral oncoproteins are also known to interact with multitudes of other host encoded molecules. This includes transcriptional
regulators like Polycomb repressive complex containing E2F6 (McLaughlin-Drubin, 2008; McLaughlin-Drubin and Munger, 2013). The mechanism by which these viral oncoproteins modulate host transcriptional machinery, are still not understood completely.

Interestingly, these HPV16 E7 interacting E2F6 transcription factors possess the ability of transcriptionally regulating EZH2, a member of the Polycomb Repressive complex 2 (PRC2). The PRC2 complex, on the other hand, has been reported to interact with a large number of long noncoding RNAs (lncRNAs), which have emerged as important regulators of transcription and disease progression, particularly of various cancers (Bracken et al, 2003; Hyland et al, 2011; Wapinski et al, 2011). The PRC2 complex targets a multitude of host genes and regulates their expression. The Homeobox (HOX) cluster members belong to this category of PRC2 targets (Wang et al, 2011b).

2.2.2 Host genetic and epigenetic factors in cervical cancer pathogenesis

(A) **Alteration of global gene expression in cancers:** Microarray based expression profiling has been extensively used to identify the molecular events controlling the process of cancer development, as well as for the identification of biomarkers for predicting prognosis and response to therapy. The first study on microarray based analysis was performed to distinguish the Acute Myeloid Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL) with 100% precision using a signature of 50 genes (Golub et al, 1999). Alizadeh et al, utilized expression profiling techniques to stratify DLBCL to two subtypes: germinal centre B-like DLBCL and activated B-like DLBCL. Perou et al, in 2000, also used gene expression patterns across 8,102 genes to describe the phenotypic diversity of breast cancers. The tumors clustered in two major groups which largely reflected the ER-positive and ER-negative clinical subtypes. Later, Chang et al, in 2005, identified four major breast cancer subgroups: a “luminal cell-like” group expressing the estrogen receptor (ER); a “basal-like” group expressing keratins 5 and 17, integrin β4, and laminin, but lacking ER expression; an “Erb-B2-positive” group and a “normal” epithelial group, on the basis of expression levels of 70 prognostic genes among 78 tumors. Such studies also established that prognosis of patients with ER-positive disease is largely determined by the expression of proliferation-related genes (Reyal et al, 2008). In prostate cancer, a clinical test based on prostate cancer 3 (PCA3) biomarker mRNA in urine, was recently approved in Europe under CE trademark (Sorensen et al, 2010).

Such advances in molecular biology and high throughput gene expression profiling
technologies have heralded a new era in biomarker discovery and identification of molecular targets related to carcinogenesis. Several studies have aimed to identify the gene expression changes that occur during CaCx pathogenesis. Wong et al, in 2006, aimed at identifying genome-wide characteristic differences between the cervical squamous cell carcinomas and healthy controls and identified SPP1 (Osteopontin), CDKN2A (p16), RPL39L, Clorf1, MAL, p11, ARS and NICE-1, to be the most differentially expressed genes in CaCx. Similar studies by Martin et al, in 2009, identified p16ink4a, MCM 3 and 5, CDC6, Geminin, Cyclins A-D, TOPO2A, CDCA1, and BIRC5 to be the most significantly altered genes in CaCx, which was confirmed by quantitative Real Time PCR and expression analysis at the protein level by immunohistochemistry. A recent study by Balacescu et al, in 2014, however, used gene expression profiling to identify molecular markers, which play an important role in intrinsic resistance and therapy failure in CaCx. The study identified 'DNA Replication, Recombination and Repair' to be the most important mechanisms activated in non-responsive cervical tumors, and the 'Role of BRCA1 in DNA Damage Response' was predicted to be the most significantly altered canonical pathway involved in intrinsic resistance (Balacescu et al, 2014).

Among Indian population, Rajkumar et al in 2011, focussed on identification of genes differentially expressed at various stages of CaCx development, from normal cervix samples, to CIN1/CIN2 followed by CIN3/CIS and invasive CaCx. The study identified 20 genes to be up-regulated and 14 down-regulated in CaCx and 5 up-regulated in CIN3, with UBE2C, CCNB1, CCNB2, PLOD2, NUP210, MELK, CDC20 to be the most significantly up-regulated genes among tumors and CIN3/CIS as compared to the normal samples and CIN1/CIN2. Another study among Indian women looked at CaCx across different FIGO stages of disease, from early lesion (Stage IA and IIA) to progressive stages (Stage IIB and IIIA-B), identifying distinct gene expression signatures (Thomas et al, 2013). However, none of these studies attempted for molecular characterisation of CaCx, as has been done for DLBCL and breast cancers in previous studies.

(B) **HOX cluster encoded transcription factors and long noncoding transcripts in cancers:** HOX cluster comprises of four distinct loci, located on four different chromosomes, namely HOXA, HOXB, HOXC and HOXD. HOX genes are a family of transcription factors, which act as master regulators of multiple pathways during the normal developmental process as well as during maintenance of normal adult cells. HOX genes are expressed in a tissue-specific and temporo-spatial manner (Reid et al, 2002). However, a
large number of these genes have been reported to show deregulated expression in multiple forms of carcinomas. In fact, HPV infection in case of CaCx is accomplished by invasion of the mucosal epithelia at the junction of the vagina and ectocervix, predisposing the cells to cellular transformation. Thus, the productive stage of the viral life cycle is highly dependent on host cell differentiation process, which in turn is tightly regulated by HOX genes. Not just the coding genes but also the noncoding regions are known to be transcribed from the HOX cluster leading to the synthesis of many long noncoding transcripts, which show a temporo-spatial expression pattern similar to the neighbouring HOX genes. These noncoding transcripts play a critical role in regulating not only the transcription of the HOX genes, in *cis* or in *trans*, but also regulate gene expression at the global level by associating with chromatin remodelling complexes.

Although several studies have been carried out to study the gene expression profile of the HOX genes in CaCx, most studies are based on CaCx cell lines (Hung *et al.*, 2003). So, there is a need to discover the expression profile of these genes in normal healthy cervical tissues and compare these with the CaCx, among a substantially large cohort of healthy women. The noncoding partners of these genes also remain to be identified. So also, studies on their roles in HOX gene regulation as well as any other potential functions in oncogenic transformation of healthy cervical cells, needs to be performed. Such studies will help in developing an understanding as to how these noncoding RNAs aid in CaCx progression. One such widely studied IncRNA is HOTAIR, which has reported to show deregulated expression in a variety of cancer types, being highly overexpressed in metastatic cancers. As reported, the E7 oncoprotein has some role in regulating the activity of Polycomb Repressive Complexes (HOX gene regulatory complexes), through its association with E2F6. PRC2 again is known to bind multiple long noncoding RNAs including HOTAIR, which regulates HOXD10 expression in *trans* (Tsai *et al.*, 2010). Thus it is likely that these E7 proteins and the noncoding transcripts of the HOX cluster might be acting in tandem, to regulate the epigenome of the host. HOXA11-AS1 and HOTTIP are the two other reported noncoding transcripts of the HOX cluster, both of which regulate some members of the HOXA cluster that are known to be deregulated in CaCx (Chau *et al.*, 2002; Richards *et al.*, 2015; Deng *et al.*, 2015; Zhang *et al.*, 2015). However, the activity of these long noncoding RNAs still remains to be elucidated.

*(C) Chromatin remodelling Polycomb Repressive Complex 2 (PRC2) in cancer:* Polycomb genes were originally discovered in *Drosophila* as a result of their essential role
in establishing orderly body segmentation, by silencing homeobox genes. Polycomb group proteins are now recognized as key components in repressive multiprotein complexes that modify chromatin, to silence large numbers of genes in all higher eukaryotic cells (Simon and Kingston, 2009; Margueron and Reinberg, 2011). In mammalian cells, two main Polycomb Repressive Complexes (PRCs) have been defined: PRC1 and PRC2. Both PRCs repress hundreds of genes in embryonic stem (ES) cells, including a plethora of developmentally important transcriptional regulators (Boyer et al, 2006). The majority of target genes in ES cells are jointly occupied and repressed by PRC1 and PRC2 in part, because PRC2 has the capacity to generate a histone modification, trimethyl histone 3 Lys 27 (H3K27me3) mark. This can serve to recruit PRC1 by binding H3K27me3 via a chromodomain, in a polycomb group CBX family member protein (Ku et al, 2008). Efficient H3K27 methylation requires the cooperation of several PRC2 core components, including EZH2, the enzyme that catalyses the reaction via its SET domain and cofactors SUZ12 and EED. Knockout of either of these genes in ES cells results in severe global reduction of H3K27me3.

There is compelling evidence that PRCs regulate epithelial cell differentiation as well as expansion of basal cell pools during wound healing (Shaw et al, 2009; Eckert et al, 2011; Zhang et al, 2012). HPVs may need to target both of these processes to allow for viral progeny synthesis and establishment of long-term persistent infection, respectively. Hence, PRC components and molecules that regulate their activity are attractive targets for HPV proteins. The enzymatic PRC2 component, i.e., H3K27 methyltransferase EZH2, is a bona fide human oncogene that is overexpressed and amplified in human tumors (Bracken et al, 2003; Morin et al, 2010; Chase et al, 2011). EZH2 transcription is regulated by E2F transcription factors (Bracken et al, 2013). HPV16 E7 targets pRb, p107 and p130 for proteasomal degradation and associates with E2F6 and therefore interferes with E2F transcriptional repressor activities. This causes deregulated and increased expression of E2F transcriptional targets. Consistent with this model, HPV16 E7 transcriptionally activates EZH2 through its E2F sites, and EZH2 is highly overexpressed in cervical lesions and tumors (Holland et al, 2008). Interestingly, cervical carcinoma cells appear addicted to EZH2 and its depletion causes G1 cell cycle arrest and a low level of apoptosis.

High levels of EZH2 expression in HPV-positive CaCx and HPV E7 expressing cell lines and their apparent addiction to EZH2 is particularly remarkable since the H3K27me3 mark is decreased, and not increased, in such cell lines. Different potential mechanistic
explanations have been proposed to account for this apparently paradoxical finding. There is evidence that the enzymatic activity of EZH2 in PRC2 complexes is negatively regulated by AKT mediated phosphorylation at serine residue (S) 21 (Cha et al, 2005). Since HPV16 E6 and E7 have both been reported to cause AKT activation (Menges et al, 2006; Spangle et al, 2010), it is conceivable that PRC2 complex-associated EZH2 enzymatic activity may be low, despite high-level overexpression in HPV16 positive lesions and cancers. Moreover, HOTAIR has also been shown to play an important role in regulation of PRC2 complex activity. Thus, it will be interesting to determine whether an interplay between HPV16 E7, HOTAIR and PRC2 complex exists, which is the regulator of global gene expression in CaCx cells.

With this background, this study was undertaken to fulfil the following objectives:

**2.3 OBJECTIVES OF THE STUDY**

1) Compare the transcriptome of HPV 16 positive CaCx, HPV 16 positive non-malignant samples, HPV negative control samples;

2) Identification and validation of variations in expression of the HOX clusters genes among the three groups of samples;

3) Identification and validation of expression level changes, if any, of long noncoding RNAs transcribed from the HOX cluster genes;

4) Identify the role of PRC2-bound lncRNA, HOTAIR, in modulating the global gene expression and the impact of HPV16 E7 on its function/expression;

5) Identify the role of HPV16 E7 in regulating gene expression through regulation of PcG complex activity.