2.1 SCIENTIFIC BACKGROUND

2.1.1 Origin of the study

Cervical cancer (CaCx) is the second most common malignancy in both incidence and mortality among women. It is an important public health problem for developing countries, being the most or the second most common cancer among women. Despite a decline in the incidence of CaCx in the United States, in less developed nations, CaCx remains one of the leading causes of cancer mortality. Asia alone contributes 49.4% of worldwide CaCx burden (Parkin and Ferlay 1999). This is the foremost cancer burden of Indian women. The absence of, or comparatively ineffective screening programs, has lead to relatively late diagnosis of the disease in India.

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 tumour suppressor (p53) and retinoblastoma (pRb) proteins. As a result, alteration in cell cycle regulatory pathways cause increased cell proliferation and genomic instability thereby inducing neoplasia and hence, viral propagation. It has been suggested that immune deregulation may be the other key event required for CaCx development.

The natural history of the CaCx development is intriguing, since only a minority of cervical tissues infected with HPV inevitably progress to cancer. Given that around 291 million women throughout the world are carriers of HPV DNA (de Sanjosé et al., 2007). The incidence of CaCx is relatively low (zur Hausen, 2002). Thus, the
development of CaCx occurs in a few women who cannot resolve their infection and who maintain persistent active infection for years or decades following initial exposure. However, the molecular reasons why the infection is controlled or instead progresses to subsequent stages of tumorigenesis are largely unknown. This study has therefore been drawn in order to decipher the contribution of viral and host factors (genetic and epigenetic) towards predisposition to CaCx development. This is important for distinguishing women at the highest risk of HPV infection and progression to CaCx.

2.1.2 Viral factors

Type specific HPV prevalence is known to vary in different geographical regions. Of more than 35 HPV types found in the genital tract, HPV 16 accounts for 50-60% of the cervical cancer cases in most countries, followed by HPV18 (10-20%), and HPVs 31 and 45 (4-5% each). Thus, HPV16 is the most common type in all geographic regions with HPV18 following closely, particularly being more common in Southeast Asia. While this pattern is common in squamous cell carcinomas (SCCs) worldwide, HPV18 is most predominant in case of adenocarcinomas (ADCs) followed by HPV16.

After infection of the cervical epithelium with HPV, the viral genomes are maintained as episomes in the nucleus of normal (infected) cells. In premalignant lesions, except in carcinoma in situ and severe grades of CIN, the HPV DNA always persists as a non-integrated episomal molecule. In CIN and more frequently in cancers, HPV genomes are found integrated into the host chromosome (Das et al 1992). Integration of the viral genome takes place at fragile sites and at sites of proto-oncogenes in the host genome. This event involves E1 and E2 genomic sites of the virus, disrupting some portions of these regions. Disruption of E2 results in loss of transcriptional regulation of E6 and E7 genes, leading to cellular transformation. Several studies demonstrated that HPV 16 DNA is present in an integrated form in more than 70% of CaCx (Das et al 1992). The frequency of HPV integration has also been found to increase with disease in severity correlating with disease progression.
However, absence of viral integration in about 30% of such cases indicates that integration may not be a sole pre-requisite for malignant transformation.

2.1.2 A Integration and different forms of viral genome in cervical cancer

In contrast to suprabasal differentiated keratinocytes of squamous epithelium that support vegetative HPV life cycle in benign infections, the entire epithelium in high-grade dysplasias or carcinomas comprises of basal cell-like proliferating cells due to prolonged cell-cycle deregulation by high levels of E6 and E7 (Stoler et al., 1992). These cells often harbour integrated viral genomes (Van Tine et al., 2004b) and integration increases with severity of lesions (Hudelist et al., 2004). Viral integration [by non-homologous end-joining recombination (Ziegert et al, 2003)] mainly occurs in the hinge region of E2 ORF with deletion of few nucleotides downstream, which disrupts negative feedback control of E2 repressor on oncogenic expression (Romanczuk and Howley, 1992). E6/E7 transcripts from integrated DNA capture 3’ cellular polyadenylation signals (Smotkin and Wettstein, 1986) thereby increasing viral mRNA stability (Jeon and Lambert, 1995).

In HPV16 infected CaSki cell line, about 500-600 copies of viral genome are integrated at about 12 chromosomal sites, mostly in tandem arrays (Badal et al., 2003), also called ‘concatemers’. In naturally selected cells, viral DNA gets transcribed from only one or two active center(s) localized to the perinucleolar region (Badal et al, 2003). The 3’ repeat of the tandem array, which is fused to cellular DNA producing stable mRNA (Jeon and Lambert, 1995), remains transcriptionally active, while other repeats are silenced (Baker et al., 1987) by methylation (being partially reactivated by 5-azacytidine) (Van Tine et al., 2004b). Silencing of the internal viral tandem repeats can avoid mitotic catastrophe (Van tine 2004) due to chromosomal breakage-fusion bridge cycle based on ability of E2 to associate viral origins to kinetochores during mitotic segregation of chromosomes (Van Tine et al., 2004a).
There are several reports on substantial presence of intact E2 (from either pure episomes or mixture of episomes and integrates) in CaCx specimens (Bhattacharjee and Sengupta, 2006a; Das et al., 2010; Hafner et al., 2008). Therefore, it is expected that HPV can exist within the host cell in different genomic forms, like integrated, episomal, concomitant (mixture of episomes and integrates) and concatenated (Figure 2.1).

![Figure 2.1: Different forms of viral genome within host cell](image)

2.1.2 B Viral transcription

**Viral mRNA splicing and polyadenylation:** The HPV 16 genome has two promoters, two polyadenylation sites and different splice sites and produces polycistronic mRNA (Zheng and Baker, 2006). Extensive alternative mRNA splicing produces different mRNA transcripts for different HPV early and late proteins (Figure 2.2). The promoters, p97 and p670, respectively regulate transcription of early and late proteins. The early promoter p97 is present in the LCR region just upstream of E6 and is controlled by cis-elements (like four E2-binding sites) in the LCR (Romanczuk et al., 1990). The late promoter p670 is located within the E7 ORF and is induced only in
differentiated keratinocytes (Grassman et al, 1996). The late region transcripts are detected only during vegetative HPV life-cycle in differentiated keratinocytes (Barksdale and Baker, 1993). The early and late mRNA transcripts are polyadenylated at nt 4215 (AE – early polyadenylation site) and nt 7321 (AL – late polyadenylation site), respectively. All the early primary transcripts have 3 exons and 2 introns. Each intron has three 3’ splice sites [nts 409, 526 and 742 in intron 1 (Zhao et al., 2005) and nts 2582, 2709, 3358 in intron 2 (Doorbar et al., 1990)].

Around 14 different early transcripts are formed with different coding potentials (Zheng and Baker, 2006). The early polyadenylation site is located in the intron (between nt 3632 and nt 5639) of late mRNA transcript. There is a competition between splicing of nts 3632-5639 (to produce L1) and polyadenylation at nt 4215 (to produce early protein) (Zheng and Baker, 2006). There is a splicing enhancer (a 65 nt long AC-rich sequence that stimulates early mRNA splicing at nt 3358) within the E4 coding region, which enhances early transcript polyadenylation (Rush et al., 2005). So, E1*E4 transcript can predominate over L1. E4 protein is produced only when parts of E1 and E2 are spliced (nt 880 – nt 3358). Such splicing generates repressor E2 protein (regulates p97) (Romanczuk et al, 1990) instead of the transactivator full length E2 (Lambert et al., 1987). HPV16 E6 and E7 pre-mRNAs are transcribed as bicistronic E6E7 pre-mRNA from the P97 promoter. It has an intron in the E6 coding region with one 5’ splice site and three 3’ splice sites. Alternative splicing based on these 3’ splice sites produce E6*I, E6*II, and E6^E7 mRNAs, respectively (Smotkin et al., 1989). The unspliced transcript produces E6-E7 mRNA expressing oncogenic full length E6. Splicing of the E6-E7 mRNA result in more E7 RNA templates, but overly efficient splicing can prevent E6 expression (Zheng and Baker, 2006).
Figure 2.2: Genome structure and transcription map of HPV-16. (Zheng and Baker, Front Biosci, 2006). The bracket line in the middle of the panel represents a linear form of the virus genome for better presentation of head-to-tail junction, promoters (arrows), and early (AE) and late (AL) polyadenylation sites. The open reading frames (ORFs) (open boxes) are diagramed above the bracket and the numbers above each ORF are the nucleotide positions of the first nucleotide of the start codon and last nucleotide of the stop codons in the HPV-16 genome. The E4 ORF spans two exons and formation of the intact E4 ORF requires RNA splicing (dashed lines). LCR indicates a long control region. Below the bracket line are the reported RNA species derived from alternative promoter usage and alternative splicing. Exons (heavy lines) and introns (thin lines) are illustrated on each species of the RNA, with splice site positions being numbered by nucleotide positions in the virus genome. All early transcripts driven by early promoter P97 are illustrated with a 5'-end from nt 97. Similarly, all late transcripts driven by differentiation-inducible promoter P670 are shown for convenience with a 5'-end from nt 670.

**Viral oncogenic expression in E2-intact carcinogenesis:** A substantial proportion of the CaCx cases harbor intact E2 inspite of the fact that HPV16 E2 protein negatively
regulates transcription of the E6 and E7 genes (Bhattacharjee and Sengupta, 2006a). Mechanisms of transcriptional deregulation have not yet been well-characterized in episome-associated cervical carcinogenesis (Gray et al., 2010). DNA sequence alterations due to polymorphism at LCR (Bhattacharjee and Sengupta, 2006a) or E2BS methylation (Bhattacharjee and Sengupta, 2006b) have been reported in episome-associated SCCs. Some alterations in the DNA sequence of transcription-factor binding sites in upstream regulatory region are associated with increased viral enhancer activity (Dong et al., 1994), others are not (Lace et al., 2009). All these suggest alternative mechanisms of transcriptional deregulation in episome-associated carcinogenesis (Gray et al, 2010). It has also been recently observed that viral load of E2 intact cases was significantly higher compared to those with disrupted E2 which points towards causal relevance of high viral load in E2-intact CaCx pathogenesis (Das et al, 2010).

**Detection of integrate- and episome-derived viral mRNA transcripts:** In 1999, Klaes et al described an RT-PCR based assay, amplification of papillomavirus oncogenic transcript (APOT), to differentiate between integrate-derived and episome-derived early mRNA transcripts. Integration into host cellular DNA disrupts E1 or E2 ORF along with the loss of early polyadenylation site (Doeberitz et al, 1992). So, the integrate-derived oncogenic transcripts have viral sequence at 5’ end but host cellular sequence at 3’ end. On the other hand, episome-derived oncogenic transcripts are flanked by viral sequences on both ends including the early polyadenylation site at the 3’ end (Sherman et al., 1992). All these oncogenic transcripts usually use the 5’ splice site at the beginning of the 2nd intron of the early transcripts (Schneider-Gädicke and Schwarz, 1986). The episome-derived transcripts use the 3’ splice site at nt 3358, thereby maintaining E4 ORF intact.

The assay involves a 3’ RACE, where both episome- as well as integrate-derived oncogenic transcripts are amplified using E7-specific forward primers and (dT) 17-p3 (p3 is an adaptor sequence) reverse primer. The episome-derived transcript has a specific size of 1050 bp, which is absent in integrate-derived transcripts (Klaes et al, 1999). The obtained PCR fragments are verified by Southern blot hybridization using E7- and E4-specific probes (Klaes et al, 1999).
Figure 2.3: Structure of HPV16/18 genomes, either episomal (left) or integrated (right), and derived transcripts (R Klaes et al, Cancer Res, 1999). Locations of primers used for RT [(dT)17-p3], for PCR APOT (p1, p2, and p3), and for hybridization analysis (h1, HPV E7-specific; h2, HPV E4-specific) are indicated. Integrate-derived transcripts either completely (type A) or partially (type B) lack E4 sequences.

Figure 2.4: Gel electrophoresis of APOT-PCR showing 1050bp band-size for episome-derived transcript and its absence in integrate-derived transcript (Klaes et al, Cancer Res, 1999)
Samples harbouring only episomal HPV DNA may show non-specific bands in presence of high amounts of episomal transcripts. This is due to mispriming of the oligo (dT) primer to A-rich regions. Amplification of various artifacts with different lengths than the standard episomal transcript (1050 bp) are possible. The band sizes include 250 bp, 870 bp and 1400 bp representing the transactivator E2 (unspliced form), and 530 bp, 610 bp and 1050 bp representing repressor E2.

Figure 2.5: Gel electrophoreses showing different bands generated by APOT PCR due to mis-priming of oligo(dT) primer to A-rich regions (adapted from APOT-protocol for HPV16 and HPV18, Wentzensen et al, 2002)

2.1.2 C Viral Load

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). 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; Saunier et al., 2008). However, there are several reports, which have failed to relate high HPV16 DNA copy number with CaCx development (de Boer et al., 2007; Hernández-Hernández et al., 2003). E6, which is a stable oncogene, is used to measure viral load by hydrolysis probe-based real time PCR (Peitsaro et al., 2002). A number of
studies (Cricca et al., 2007; Saunier et al., 2008) have proposed the ratio of E2 gene copy number to that of E6 (E2/E6) as a reliable indicator of the physical status of viral genomes within the host cells.

2.1.3 Host Factors

2.1.3 A HLA class I genetic variation

HLA region: Major Histocompatibility Complex (MHC) is located at 6p21.3 (Sanchez-Mazas et al., 2000). It is also known as Human Leukocyte Antigen (HLA) in human beings. It is around 3.6 Mb long. It is divided into three regions from centromere to telomere – MHC class II (1.1 Mb), MHC class III (0.7 Mb) and MHC class I (1.8 Mb) (Tsujimura et al, 2002) in addition to other loci of HLA-related, -unrelated or unknown functions (Campbell and Trowsdale, 1997). Function of the proteins coded by HLA classes I and II is to identify 'non-self' peptides, bind them and present to T lymphocytes (Zoodsma et al., 2005) to elicit immunologic response.

Many genes in this region exhibit allelic variations, because MHC is highly polymorphic with some of the genes, like HLA-B and HLA-DRB1, being the most polymorphic in the human genome (Bodmer et al., 1997). HLA alleles are multilocus haplotypes that are specified by unique combinations of sequence motifs generated by mutation, gene conversion and recombination (Marsh and System, 2005). HLA domain is different from SNP domain and in HLA this domain, each locus is defined not on nucleotide level but on HLA level (Listgarten et al, 2008). There are at the most 4 possible alleles (A/T/C/G) at each SNP locus, while in the HLA domain, each locus can have hundreds of alleles [presently, total 6403 HLA alleles; 4,946 belong to HLA class I; HLA-A, HLA-B and HLA-C have, 1601, 2125 and 1102 alleles, respectively (http://hla.alleles.org)]. HLA haplotype inference is based on the HLA loci in contrast to the haplotypes constructed on nucleotide data at SNPs. However, HLA haplotype inference can be drawn from SNP domain (Listgarten et al, 2008).
Extended regions of strong linkage disequilibrium exist in the HLA region (Bateman and Howell, 1999) interspersed with recombination hot spots (Jeffreys et al, 2001) after every 0.8 Mb, on an average (Cullen et al., 2002).

![Map of the human MHC region with the physical distances between the studied loci.](image)

**Figure 2.6:** Map of the human MHC region with the physical distances between the studied loci. The horizontal bars plotted show the pairs of loci in significant linkage disequilibrium. Significant LD exists in the region between loci A and B (A Sanchez-Mazas et al., Eur J Hum Genet, 2000)

In a study including 50 French families, it was shown that the centromeric DP-DQ interval (physical distance of around 400 bp) did not exhibit any linkage disequilibrium, while the telomeric region between loci A and B (physical distance of more than 1400 bp) exhibited significant LD, that is characterized by a low recombination rate (coldspot or frozen block) (Figure 2.6) (Sanchez-Mazas et al, 2000). The LD observed in this region extends even beyond HLA-A towards the telomere, particularly for A1-B8 haplotype (Worwood et al., 1997).

Chromosome segments within the HLA region are conserved [possibly with variations among populations (Dawkins et al., 1991; Degli-Esposti et al., 1992) not as a systematic consequence of physical linkage, but probably based on other mechanisms like natural selection (Sanchez-Mazas et al, 2000). Moreover, both significant LD
(result of directional selection) and some form of balancing selection [lower homozygosity and heterozygote advantage in response to pathogens like hepatitis B (Carrington et al., 1999) and HIV (Thursz et al., 1997)] are observed in MHC and may serve to explain the complex pattern of HLA diversity. Linkage disequilibrium maps may differ with populations.

**HLA profile in India:** A study carried out on HLA class I profile of Bengali population residing at Siliguri in West Bengal, India, revealed high frequency of A*02, A*11, A*24, A*31, B*07, B*08, B*18 and B*37 alleles (Singh et al, 2009). Reports from various parts of India have separately shown uniformly high frequency of A*02 and A*11 in North India (Mehra et al.1997; Rajalingam et al, 2002) and South India (Thomas and Banerjee 2005) as well as among Sikh (Babita and Usha, 2004), Gujarathi, Maharastrian (Kankonkar et al.2004), different castes of western India (Shankarkumar et al. 2001), Gurkha (Debnath et al.2006a) and Bangladeshi Bengali population (Ali et al., 2008), who are closely related to Bengalis of West Bengal (Singh et al, 2009). A*24, B*44, B*40 and B*51 were also found to be common among Bangladeshi Bengali population (Ali et al. 2008) and A*31 among some of the South Indian populations (Vettriselvi et al. 2006)

**Association of HLA alleles with CaCx:** There are few studies that have investigated the association of HLA class I alleles with HPV infection and squamous cell carcinoma of cervix. Nevertheless, some studies have demonstrated associations of CaCx pathogenesis with particular HLA class I alleles (Hildesheim et al., 1998; Wang et al., 2001). Brady et al (1999) reported a negative association of B15 in the Caucasian populations of Europe, but this was not supported by other studies (Krul et al 2000,). Of these, studies carried out on Caucasian population reported B44 as risk antigen for disease progression (Bontkes et al, 1998) but this was also not confirmed by other studies (Krul et al 2000). The HLA B*07 allele, which is in linkage disequilibrium (LD) with the DQB1*0602 and DRB1*1501 HLA class II alleles, has shown an association with CaCx and has been proposed to be involved in the evasion of the immune system by HPV16 (Ellis et al 1995). Epidemiological data from three populations of North and

42
Central America identified HLACw*0202 as a protective allele for cervical disease, using high-resolution genotyping methods but failed to show association of any single allele of HLA A, B, and C with disease risk.

A comprehensive analysis of HLA class I and II loci by Madeleine et al 2008, that co-occurring HLA alleles across loci seem to be more important than individual alleles in imparting risk or protection against CaCx in an HPV specific manner. A study from Sweden, reported DQB1*0603 and DRB1*1301 alleles to be protective against cervical cancer, but HLA-B locus unlikely to be genetically predisposed to the disease (Engelmark et al, 2004). On the contrary, a nested case-control study from Portland, Oregon (USA) reported association of B*07 and DQB1*0302 with increased risk and DRB1*13 with decreased risk (Hildesheim et al., 1998). DRB1*13 was also found to be protective in other studies (Apple et al., 1994; Sastre-Garau et al., 1996). In a study from the United Kingdom, B*07 was found to be associated with poorer prognosis (Duggan-Keen et al., 1996). The same group had also reported the presence of an HPV16 E6 non-wild type variant in B*07 positive cervical cancer patients, suggesting alteration of B*07 binding epitope in this variant (Ellis et al, 1995).

Conflicting associations of different HLA alleles are thus found in CaCx. This can be explained by population-specific linkage disequilibrium patterns (Maciag and Villa, 1999) and also HPV variant-specific epitope-binding with HLA (Ellis et al, 1995). Despite the importance of HLA class I-mediated CTL responses to viral and tumor antigens, very few studies have investigated role of HLA class I alleles in cervical cancer (Hosono et al., 2010). The notion about the available reports on putative associations of HLA class I alleles (Hildesheim et al, 1999; Madeleine et al, 2008) and molecules (Koopman et al., 1999) with disease-risk necessitates elaborate epidemiological research on each HLA class I allele (Hosono et al, 2010) and its underlying polymorphisms.

2.1.3 B Methods of HLA typing

There are different methods of HLA typing. Previously, serology-based typing was in practice, which depended on recognition of structural differences of HLA
molecules on the cell surface by different antibodies. This did not account for the amino acid changes in the antigen-binding groove (Shankarkumar, et al, 2008). DNA-based methods of typing include polymerase chain reactions with either sequence-specific priming (PCR-SSP) (Bunce et al., 1995) or sequence-specific oligohybridization (PCR-SSO). To maintain high accuracy in typing by these methods, the number of probes and primers needs to be upgraded with the rapidly increasing allelic diversity (Shankarkumar, et al, 2008). Sequence-based typing (SBT) of HLA alleles is the most comprehensive method to characterize polymorphisms in this region. The most polymorphic regions determining the epitope-binding specificity, that is, exons 2 and 3 for HLA class I (Kurz et al., 1999; Turner et al., 1998) and exon 2 for HLA class II are sequenced in SBT (Listgarten et al, 2008). Subsequent detailed interpretation of HLA alleles is based on comparison of nucleotide sequence with online database of allelic combinations (Robinson et al., 2000). SBT can yield ambiguities of two kinds that result – (i) if difference between two HLA alleles lie outside the sequenced region, and (ii) if more than one allele combination produces exactly same pattern of heterozygous nucleotides (Listgarten et al, 2008). Ambiguous HLA alleles can be resolved by sequencing additional exons and by PCR with group-specific primers (Rozemuller, 2006).

2.1.3 C HLA class I transcription and promoter methylation

In CaCx, downregulated expression of HLA class I molecules could be mediated through there interaction with the viral oncoproteins. There are reports showing the repression of the MHC class I heavy chain gene promoter by E7 protein of HPV16 and 18 (Georgopoulos et al., 2000) and loss of MHC class I by E5 of HPV16 (O'Brien and Saveria Campo, 2002), which might be attributed to a loss of recognition of virus-infected cells by MHC class I-restricted CTL. Frequent loss of expression of these antigens at the cell surface has been observed in many human cancers (Ruiz-Cabello et al., 1991). Loss of HLA class I antigen expression can occur due to genomic alterations, transcriptional regulation and protein transportation. Transcriptional inactivation of HLA class I genes, has been observed in case of pancreatic cancers (Ryschich et al.,
and in esophageal squamous cell carcinomas (Nie et al., 2001), primarily caused by DNA hypermethylation. A certain proportion of these cancers, like many other epithelial cancers, have also been identified to harbor HPV infections.

Loss of HLA expression in cervical cancer can occur at genetic [mutations observed in specific HLA-B and -A alleles (Koopman et al., 1999)], transcriptional or post-transcriptional levels (Kanodia and Kast, 2007). Revival of HLA class I protein expression by transcription factors like IFNγ (Qian et al., 1998) and estrogen (Rodriguez et al., 1994) in hepatocellular carcinoma and breast cancer, respectively, points towards a direct association between transcriptional downregulation and lack of HLA class I protein expression. Studies involving re-expression of HLA genes with 5'-aza-2'-deoxycytidine (demethylating agent) treatment suggests hypermethylation to be an important cause of transcriptional silencing (Serrano et al., 2001). Promoter hypermethylation of the HLA-A, -B and -C genes have been found to be associated with different human cancers like melanoma cells (Serrano et al, 2001), esophageal squamous cell carcinoma (Nie et al, 2001) and gastric cancer (Ye et al., 2010).

DNA methylation occurs at CpG motifs. CpG islands are small regions of 0.5 Kb to 5 Kb length with a GC content of more than 60% and a CpG: GpC ratio of more than 0.6 (Baylin et al., 1998). All the three classical HLA class I genes have shown high CpG content with CpG: GpC ratio of 0.83–1.95 (Nie et al, 2001). In mammals, DNA methylation by family of methyltrasferases (DNMT1, DNMT3a, DNMT3b) (Okano et al., 1998) can alter the DNA secondary structure (major groove conformation) (Nie et al, 2001; Robertson and Jones, 2000) as well as attract methylated DNA-binding proteins (MeCP) (Hendrich and Bird, 2000) which in turn attract HDAC1 (histone deacetylase) thereby causing transcriptional repression through chromosome remodeling (chromatin condensation) (Orlando and Jones, 2002; Struhl, 1998).

There is a lack of any comprehensive analysis of the relationship between the HLA class I alleles/haplotype composition, the expression of HLA class I molecules encoded by such alleles and the influence of CpG hypermethylation in the corresponding promoters on such expression, in case of CaCx.
2.2 OBJECTIVES

This study is aimed at understanding the roles played by some of the important viral as well as host factors in HPV16/18 related CaCx pathogenesis. The objectives of this study are to test the following hypotheses:

1. Certain viral factors (intactness of the HPV16 E2 gene, viral load and oncogene expression) are interdependent in determining the risk of cervical cancer.

2. Certain HLA class I alleles/haplotypes predispose towards or protect against HPV16/18 infections and CaCx development.

3. Downregulation of HLA class I transcription significantly modulate disease risk and if so, whether HLA promoter hypermethylation is associated with the phenomenon.