Identification of Specific Genetic Alterations in Cervical Cancer by Genome wide LOH and Copy Number Analysis

By

POULAMI DAS
LIFE09200604005

Tata Memorial Centre
Mumbai

A thesis submitted to the
Board of Studies in Life Sciences

In partial fulfilment of requirements
For the Degree of

DOCTOR OF PHILOSOPHY

Of

HOMI BHABHA NATIONAL INSTITUTE

March, 2013

Homi Bhabha National Institute

Recommendations of the Viva Voce Board

As members of the Viva Voce Board, we certify that we have read the dissertation prepared by Ms. Poulami Das entitled "Identification of Specific Genetic Alterations in Cervical Cancer by Genome wide LOH and Copy Number Analysis" and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____**Date:**

Chairman – Dr.K. Mohandas

_____**Date:**

Guide / Convener – Dr. R. Mulherkar

_____**Date:**

Member 1 – Dr. S.N. Dalal

_____ **Date:**

Member 2 – Dr. N. Shirsat

_____**Date:**

Member 2 - Dr. S. Galande

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date:

Place:

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfilment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Poulami Das

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

Navi Mumbai Poulami Das

March 2013

CERTIFICATE

I certify that the thesis titled “Identification of Specific Genetic Alterations in Cervical Cancer by Genome wide LOH and Copy Number Analysis” submitted for the degree of Doctor of Philosophy by Ms. Poulami Das is a record of the research carried out by her during the period 2006 to 2012 under my supervision. This work has not formed the basis for the award of any degree, diploma, associateship or fellowship at this or any other institute or university.

Navi Mumbai,

March 2013

Rita Mulherkar

Dedicated to

BABA & MA

Acknowledgments

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

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

I am thankful to Dr. Rajiv Sarin (Director, ACTREC) and Dr. Surekha Zingde (Deputy Director, ACTREC) for providing me an opportunity to work in this institution and the excellent infrastructure. I also thank DST and IRG for funding the project and ACTREC for my fellowship. I am fortunate to have Dr. K. Mohandas (TMH), Dr. S.N. Dalal (ACTREC), Dr. N. Shirsat (ACTREC) and Dr. S. Galande (IISER) as my doctoral committee members. Their expert comments, critical analysis of the results and helpful suggestions have contributed significantly to the work. A special thanks to Dr. Amit Dutt (ACTREC) for his extremely precious suggestions regarding the 'next generation' sequencing data analysis. I am extremely thankful to our clinical collaborators Dr. Umesh Mahantashetty (TMH) and Dr. Shyam K. Shrivastava (TMH) for providing us with the cervical cancer samples, patient follow-up and valuable clinical inputs, as well as Dr. Kedar Deodhar (TMH) for pathological examination of the biopsies. A special thanks to Dr. D. Bandopadhyay and Dr. Jamal who had actually collected all the samples beautifully. I would also like to thank the team from Vimta Labs, Hyderabad for the Luminex array experiment and Genotypic Technologies (P) Ltd. as well as Philips Research Asia, Bangalore for their help with the NGS data.

I am extremely thankful to the sequencing facility, the common instrument facility, photography, library, IT, administration and accounts department of ACTREC for their constant help and support. I am highly grateful to Ms. S. Kannan for her invaluable help in statistical analysis of the data.

I want to express my warmest thanks to all Mulherkar lab members Dr. Ujjwala, my seniors Ashwin, Aditya, Vishal, Deepika and especially Asha who helped me in innumerable ways during the entire period. A special thanks to my batch mate Tabish who has always been there for me. Thanks to Abhishek, Rashmi, Abhay and Meenal for all their help. I cannot forget to thank Nirmala, Mansi, Ganeshji, Jadhavji and Raneji, for their excellent technical support. I would also like to thank all my trainees Rashmi, Mithila, Priyanka, Puja and especially Akansha. All these people have kept the lab environment lively during these years and made it a second home.

I owe a lot to all my friends Amit S, Tabish, Ajit, Manoj, Amit R, Amit F, Sapna, Pallavi, Lalit, Atul Amit V, Manoj Bhonsle and the entire student's community of ACTREC who kept the tough and tense days of my Ph.D as joyous as possible. I am extremely thankful to some of my oldest friends Moumita, Sanjukta and Jayeeta for being patient listeners and keeping my days fun filled.

And now, the most important people in my life, my family. Baba and Ma, you always have been my greatest strength and source of inspiration. Whatever I have achieved today is all because of your blessings and sacrifices, I owe everything to you. I feel so proud and blessed to be your daughter. A special thanks to my aunts, uncles and cousins for all their love and affection. Papa and Mummy, I am so proud to be your daughter-in- law; Manish in you I found a brother and a true friend. Finally, Amit, my husband, you are my biggest critic and source of perennial support. The best in me is always a reflection of you. With you by my side, nothing seems impossible. Thank you for being in my life.

Contents

Synopsis

Homi Bhabha National Institute

Ph. D. PROGRAMME

1. Name of the Student: Ms. Poulami Das

2. Name of the Constituent Institution: Tata Memorial Centre**,** Advanced Centre for Treatment Research and Education in Cancer

3. Enrolment No. : LIFE09200604005 (01/09/2006)

4. Title of the Thesis: Identification of Specific Genetic Alterations in Cervical Cancer by Genome wide LOH and Copy Number Analysis

5. Board of Studies: Life Sciences

INTRODUCTION

With about 134000 new cases and 72800 deaths annually, cancer of the uterine cervix accounts for the most frequent cancer among women in India. Globally, it is the third most common cancer and significantly contributes to cancer burden (WHO/ICO). Based on histology cervical cancer can be broadly divided into 3 subtypes- squamous cell carcinoma or cancer of the flat epithelial cells with the highest incidence (80-90%) rate followed by adenocarcinoma (10-20%) arising from glandular epithelium and mixed carcinoma (1-2%) with features of both types. HPV infection has been shown to play a critical, though not sufficient role in the etiology of cervical cancer. The relationship between exposure to HPV and the disease might vary with the infecting HPV type, viral integration status and viral load; these considerations might have profound implications on patient prognosis (Josefsson et al., 2000; Woodman et al., 2007).

Identification of individual HPV types in a population is important not only to investigate the epidemiology and clinical behavior of particular type, but as a primary screen for cervical cancer detection. Till date several HPV types have been reported; 24 of the most common types have been divided into three groups based on the severity: 15 high-risk types (16, 18, 31, 33, 35, 39, 45, 51, 52,56, 58, 59, 68, 73, and 82), 3 putative high-risk types (26, 53, and 66), and 6 low-risk types primarily found in genital warts and low-grade cervical lesions (6, 11, 42, 43, 44 and 70) (Munoz et al., 2003; Schmitt et al., 2006). Most of the high risk HPV (HR-HPV) infections regress spontaneously and only in about 10% cases the infection persists and progresses to high-grade cervical intraepithelial neoplasia. This generally occurs through integration of the HPV genome into the host chromosome with associated loss or disruption of E2 (Wentzensen et al., 2004). In the absence of E2 the transcriptional control on E6 and E7 is lost, which leads to immortalization and transformation of the cells (Romanczuk and Howley, 1992). It

has been reported that transcripts derived from the integrants are more stable than those from episomal viral DNA, hence the event has been associated with a selective growth advantage for affected cells (Jeon et al., 1995; Jeon and Lambert, 1995). The integration event in most cases results in the generation of mRNAs that comprises of the viral oncogenes E6 and E7 and cotranscribed cellular sequences (Klaes et al., 1999). Another important factor associated with HPV infection is the viral load. The relationship between viral load and the disease is quiet complex; while some studies report an increase in viral load with increasing disease severity, others found either no association or a higher viral load in women with low-grade squamous intraepithelial lesion than in those with high-grade squamous intraepithelial lesion (Lilo et al., 2005; Swan et al., 1999; Woodman et al., 2007). Moreover, while considering the relationship between viral load and disease prognosis, the physical status of the virus also becomes important. All these parameters taken together could therefore provide valuable insight into the natural history of HPV infections and their relationship to disease.

Apart from infection by HPV, genetic instability is also responsible for the pathogenesis of cervical cancer. The genetic instabilities may range from point mutations, copy number changes, chromosomal rearrangements to widespread aneuploidy. Somatic alterations (present only in tumour) and not in germline (present in blood) would shed light on tumorigenesis as well as response to therapy. Next generation sequencing techniques which include whole genome sequencing, exome sequencing and transcriptome sequencing, have made such studies much more feasible. Though exons constitute only about 1% of the genome (37.6 Mb), sequencing them can yield significant information as they have been reported to harbor most variations. Also most frequent type of disease mutations are those that cause amino acid substitutions resulting from variations in exons. Further, approaches involving targeted sequencing

provide increased sequence coverage of a particular region of interest at high throughput and lower cost.

The present study involves investigation of various factors involved in cervical carcinogenesis. This included genotyping of 24 HPV types along with identification of site of viral integration in a cohort of Indian women with locally advanced cervical cancers (FIGO Stage IIB and IIIB), and determining viral copy number of two high risk HPV types - HPV16 and 18. Further, whole exome sequencing has been carried out to identify the genetic alterations observed in cervical cancer biopsies compared to matched blood in some cases. The data has been validated using different techniques.

AIMS AND OBJECTIVES

1. Determination of HPV status in cervical cancer biopsies.
2. Identification of integration site of the virus by Amplification of Papillomavirus Oncogene Transcripts (APOT) assay.
3. Identification of genomic alterations in cervical cancer biopsies by exome sequencing
4. Correlation of the data from the study with clinical data.

MATERIALS and METHODS

Clinical Sample accrual: Pretreatment cervical tumour biopsies, predominantly from FIGO stage IIIB as well as blood samples were obtained from patients undergoing radiotherapy alone or concomitant chemo-radiation at the Radiation Oncology Department, Tata Memorial Hospital, Mumbai, after obtaining IRB approval. A generic consent for basic research was obtained prior to obtaining the biopsies. However for the current study a consent waiver was obtained from the Hospital Ethics Committee since the samples were collected more than 10 years ago. The samples were collected in liquid

nitrogen and stored at -80°C freezer until further use. All the samples were assigned a laboratory code to maintain confidentiality.

Processing of tumour samples: Cryosectioning of cervical biopsies was done to determine tumour percentage as well for isolation of DNA and RNA. First, two 5µm sections of tissues were mounted on lysine coated slides for H&E staining and determining tumour percentage. Next, seven 30µm sections were collected in RLT buffer for isolation of RNA and five 30µm sections collected in STE buffer for isolation of DNA.

Genomic DNA extraction: DNA was extracted using standard phenol/chloroform method and quantitated using Nanodrop. For the exome sequencing study isolation of genomic DNA was done by DNA Mini kit (Qiagen) according to the manufacturer's protocol. The integrity of DNA was checked on 0.8% Agarose gel.

Extraction of RNA: RNA was extracted using RNeasy Mini Kit. DNAse treatment of the RNA was done to remove any DNA contamination. RNA was quantitated using Nanodrop and its integrity was checked on 1.2% Formaldehyde gel.

Genotyping of HPV by high throughput Luminex assay: Genotyping of 24 HPV types which included 15 high-risk types (16, 18, 31, 33, 35, 39, 45, 51,52, 56, 58, 59, 68, 73, and 82), 3 putative high-risk types (26, 53, and 66), and 6 low-risk types (6, 11, 42, 43, 44, and 70) was carried out in 270 cervical biopsies using Multiplex HPV Genotyping assay based on Luminex xMAP technology. As per the manufacturer's instructions, PCR was carried out using sets of biotinylated broad range primers in a total volume of 50 µL containing 3.5 mM $MgCl_2$, 200 µM deoxynucleotide triphosphates, 0.75 unit of Taq DNA polymerase and 1µl primer mix. The amplification steps included an initial DNA denaturation at 94°C for 5 min, followed by 40 cycles of denaturation for 20 s at 94°C, annealing for 30 s at 38°C, and extension for 1 min 20 s at 71°C, before a final

extension for 4 min at 71°C. PCR positive samples were then subjected to the Luminex run. Ten microlitres of the PCR product was mixed with the Luminex bead mix containing distinct bead populations coupled to 24 HPV types. After thermal denaturation the target sequences were hybridized to bead-bound probes. The hybridized PCR products were labeled by binding to R-phycoerythrin conjugated streptavidin. The read-out was obtained in the Luminex bioanalyzer. HPV types were discerned according to the unique bead signature, whereas the presence of PCR products was determined by phycoerythrin fluorescence. An analytical sensitivity cut-off was calculated based on the negative control which was deducted from each of the read-out.

HPV genotyping by PCR using MY09/11 and SPF1/2 primers: The samples that tested HPV negative by luminex array (n=92) were further screened for HPV by PCR using MY09/11 L1 (Gravitt et al., 2000) and SPF1/2 (Kleter et al., 1998) primers.

Association of HPV16, HPV18 and HPV16/18 infection with clinical outcome: The genotyping data for the two HR-HPV types, HPV16, HPV18 and HPV16/18 together, where adequate follow-up data was available was compared with the clinical outcome of the patients. Kaplan-Meier analysis was done to determine association between infection with these HR-HPV types and recurrence of disease.

Identification of integration site by APOT assay: Site of viral integration was identified in a subset of cervical tumour samples (n=86) positive for HPV16, HPV18 or both and with a follow up data of 3 years using Amplification of Papillomavirus Oncogene Transcripts (APOT) assay as described by Klaes *et al* (Klaes et al., 1999). . The assay is based on a 3'-rapid amplification of cDNA ends PCR and involves the following steps-

Reverse transcription-1µg RNA was reverse transcribed using SuperscriptTM first strand synthesis and an oligo (dT)17 primer coupled to a linker sequence referred to as

Frohman primer (Frohman et al., 1988). Integrity of the cDNA was checked by β-actin PCR.

Nested PCR- For the PCR reactions HPV E7 primers was used as forward primers, an adapter primer complementary to the linker sequence in the Frohman primer as first reverse primer, and the Frohman primer as the second nested reverse primer. The PCR amplification was carried out in a total volume of 50µL containing 2.5mM $MgCl_2$, 10mM deoxynucleotide triphosphates, 25µM of each primer and 1 unit of Taq DNA polymerase. The PCR comprised of an initial DNA denaturation step of 94°C for 2min, followed by 35 cycles of denaturation for 30s at 94°C, annealing for 30s at 58°C, and extension for 4min at 72°C. A final extension for 20min at 72°C was given. 7µl of the PCR product was used as template for nested PCR. The PCR conditions were same as that of first PCR except the annealing temperature was 66°C.

Cloning and sequence analysis- The amplified products thus obtained were visualized by 1.2% agarose gel electrophoresis. Amplicons other than the major episomal transcripts (~1050 bp for HPV 16 and ~1000 bp for HPV 18) were suspected to be derived from the integrated HPV genomes. These were excised from the gel and DNA isolated using GFX PCR DNA and Gel Band Purification Kit. The isolated DNA was either sequenced directly or after cloning into pTZ57R/T vector using the InsTA PCR Cloning Kit on the DNA automated sequencer. The chromosomal integration loci were determined using National Centre for Biotechnology Information (BLAST) and/or the University of California, Santa Cruz (BLAT) hg19 (Feb. 2009) human genome assemblies. Further the integration sites were checked for the presence of fragile sites and any genes of known identity by using NCBI fragile site map viewer and the UCSC Blat tool respectively.

Validation of the recurrent integration sites: Some of the recurrent integration sites were checked at the genomic level by carrying out genomic DNA PCR with HPV E7 primers as the forward primer and primers specific to a given chromosomal region as the reverse one.

Association of viral integration with clinical outcome: The data obtained was compared with the clinical outcome of the patients. Kaplan-Meier analysis was done to determine the association of the viral state (episomal/integrated) with recurrence of the disease. Disease free survival was considered from start of radiation therapy to the time when recurrence occurred or till last follow-up. Statistical significance was evaluated using the log-rank test. Further, where viral integration was observed at a particular chromosomal locus in ≥4 cases, Kaplan-Meier analysis was performed to check whether the site of integration in the host genome had any bearing on the disease prognosis.

***Detection of HPV copy number by quantitative real time PCR (qRT-PCR)*:** Copy number of two high-risk HPV types- HPV16 and HPV18 was assessed in the 86 samples where the physical status of the virus was already known by SYBR green based qRT-PCR. A 100 bp sequence located in the E7 gene of HPV16 or HPV18 was used for amplification, while a DNA fragment in the Tata Binding protein (TBP) gene which is a single copy gene was used as reference for relative quantitation. PCR conditions were 15 min at 95 ºC and 45 cycles (15 s at 95ºC, 30 s at 62ºC, and 1 min at 72ºC). HPV copy number was estimated by the $2^{-\Delta CT}$, with the TBP sequence as a reference for two DNA copies as describedby Peter *et al*. (Peter et al., 2010).

Identification of genetic alterations by Exome Sequencing: Exome sequencing of six cervical tumour tissues (all of stage IIIB and with more than 70% tumour) and three matched controls (respective blood sample from three patients) was carried out at

Genotypic Technology (P) Ltd, Bangalore. Following steps were involved in the process:-

Paired End library preparation- This included shearing of the genomic DNA to less than 800bp fragments, ligation of adapters, gel extraction of ligated products and finally amplification of adapter-ligated library.

Hybridization- Hybridization of the DNA library to biotinylated RNA library was carried out, followed by selection of DNA-RNA hybrids by streptavidin coated magnetic beads and digestion of RNA.

Post-Hyb processing- This included PCR amplification, purification and finally sequencing of the captured library on Illumina GAIIX sequencing platform.

Analysis for identification of single nucleotide variations and indels:- The first level of analysis involved filtration of the reads by the software SeqQC. The filtered reads were then aligned against the *Homo sapiens* genome (Hg19 UCSC Build) by the program BWA 0.5.7 followed by variant calling from the alignment by Samtools 0.7.1a. In order to further shortlist the potential somatic variations (mutations) that could be important in cervical carcinogenesis, data from only the 3 paired samples was considered. The variations were first filtered out for the somatic and nonsynonymous ones and then subjected to Gene Ontology (GO) classification and KEGG pathway analysis to cluster the genes harboring the variations into groups based on their molecular function and biological processes. Important members were selected from this based on their relevance to cancer and the variations within them were subjected to prediction tools such as SIFT (Kumar et al., 2009) and Polyphen 2 (Adzhubei et al., 2010) for identifying whether a given amino acid substitution affects protein function.

Validation of the data by Sanger sequencing: Seventy one variations that included SNVs as well as short indels, in 69 genes were selected for validation by PCR followed

by sequencing. These variations were selected mainly based on their molecular function and biological processes using Gene Ontology (GO) classification and KEGG pathway analysis. The SIFT and Polyphen score were also considered.

Validation of the data by Customized SNP Array: Apart from validation by Sanger sequencing, we attempted to validate the exome sequencing data by customized SNP array. Accordingly, 700 known (reported in dbSNP version131) and 3178 novel nonsynonymous variations across the samples were selected for validation on an 8x60K customized Agilent oligonucleotide microarray. Validation of the few known variations was done as a check for the quality of the exome sequencing data.

Analysis for detection of copy number variation (CNV) and loss of heterozygosity (LOH): Identification of LOH and CNV was done from the exome sequencing only for the paired sample where the coverage between blood and tumour was comparable using 'ExomeCNV' package as described by Sathirapongsasuti *et al*.(Sathirapongsasuti et al., 2011).

RESULTS

High-throughput HPV Genotyping by Luminex bead-based array: Using $GP5^{+}/GP6^{+}$ primer set provided in the multiplex HPV genotyping kit:

- 178/270 samples were found to be positive for HPV.
- 169/178 samples were positive for different HPV types whereas 9 samples were negative. These 9 samples could have HPV infection not included in the 24 types detected by the kit.
- Infection with HPV16 and/or HPV18 infection was most common - 114 samples being positive for HPV16 alone, 6 samples for HPV18 alone and 16 samples for both HPV16 and HPV18.

- Frequently reported high-risk HPV type 31, 33 and 35 infection was found to be low.

HPV genotyping by PCR using MY09/11 and SPF1/2 primers: In order to estimate the true HPV positivity in the 270 cases, the 92 cervical cancer biopsies negative for HPV by luminex array, were subjected to PCR using MY09/11 and SPF1/2 primers.

- 25/92 samples were positive for HPV by MY09/11 PCR.
- 79/92 samples were positive by SPF1/2 primers. These 79 samples also included the 25 samples that tested HPV positive by MY09/11 PCR.
- The overall HPV positivity was therefore 95% (257/270).
- Further genotyping of the 79 samples, using HPV 16/18 specific primers, showed 49 samples to be positive for HPV16.
- Overall HPV16/18 positivity in this cohort was 69% (185/257).

Association with clinical outcome*:* Kaplan-Meier survival analysis data for 125 patients with HPV type16, 18 and dual infection and with adequate clinical follow-up revealed that there was no significant difference between infection with these two HR-HPV types in terms of disease outcome.

Identification of viral integration sites by APOT assay: In order to detect physical state and/or site of integration of the virus in 86 cases, APOT assay was performed. Sequencing data revealed that –

- 79% (68/86) of the cervical cancer samples showed HPV integration
- Rest 21% (18/86) had HPV in the episomal form
- In 18% of the patients (n=12) integrated as well as episomal form of HPV was observed
- The site of integration could be predicted with a high score in 48 cases, for the remaining 20 cases the score was low. Only those cases where the integration site was

predicted with high score (n=48) were analyzed further for different features associated with the same.

- Only 1 sample showed HPV integration at two chromosomal loci simultaneously
- Using NCBI Fragile site Map Viewer it was observed that 60% of integrations (29/48) were located in or close(~5Mb) to a common or rare fragile site
- Using the UCSC Blat tool 58% of the sequences (28/48) were seen to have homology with protein coding genes. These genes belonged to various categories ranging from oncogenes, transcription factors, and tumour suppressor genes

Validation of the recurrent integration sites: For most of recurrent integration sites such 1p36.23, 3q28, 3q23, 6q22.31, 6q23.3, 8p11.21 11q13.1 and 13q22.1 PCR amplification was observed with genomic DNA, confirming genomic integration.

Association of viral integration with clinical outcome: Survival data revealed that 16 out of 18 patients with only episomal form of HPV(16/18), had disease free survival as compared to those with integrated form of the virus, indicating a good clinical outcome ($p=0.067$, representing a borderline significance). Further, in case of viral integration, patients with integration at chromosomal loci 1p (7/7), 6q (4/4) and 11q (4/4) were disease free, while most of the patients with integration at the chromosomal loci 3q (5/8), 13q (4/4) and 20q (2/4) showed recurrence of the disease in the form of either loco-regional or distant metastasis.

***Detection of HPV copy number by quantitative real time PCR (qRT-PCR)*:** Viral copy number as high as 443 to as low as 1 was detected across the 86 samples with known viral physical status. Comparison with survival of the patients as well as physical status revealed that cases where the virus was in the episomal form, the survival had no association with copy number. But those with integrated form of the virus showed reduced survival when the copy number was high as compared to when it was low.

Therefore, copy number of the virus combined with the physical state might serve as a good prognostic marker for the disease.

Exome capture, sequencing and analysis: After variant calling, about 14400 unique sequence variants including single nucleotide variations and indels across all 9 samples were detected. Of all these variations, a total of 7407 SNVs, 94 small deletions (1-5bp) and 71 small insertions (1-3 bp) were not reported in dbSNP131 and were represented as putative novel sequence variants. These variations could be unique to the Indian population. Further analysis revealed that about 6114 of the SNVs to be nonsynonymous. Analysis with the paired data identified 874 novel, potential somatic SNVs, 30 small deletions (1-5 bp) and 22 small insertions (1-3 bp). Out of these, 395 SNVs, 12 small deletions and 5 small insertions were found to have defined/ important functions. SIFT analysis predicted 167 variations to have a potential deleterious effect on the protein function while with Polyphen 2, 128 variations were predicted to be damaging. Both the prediction tools predicted 73 common variations to be deleterious.

Validation of the data by Sanger sequencing: So far, we have tried validating 71 different variations in 69 genes by PCR followed by sequencing. Most of the variations did not get validated by this method. This may be due to the fact that Sanger can accurately detect 1:2 or 1:3 representations of heterozygotes and since almost all of these variations were heterozygous, there is a chance of missing out the mutant allele. However, certain variations in genes like- RNASEL (R462Q), PTPRJ (R1222H), ZFP64 (Y210X), NFAT5 (Y1188C) and FBN1 (delAGG), were validated by sanger sequencing. But when we checked for these variations in the blood, they were detected in the blood as well. Therefore, these cannot be considered as somatic as given by exome sequencing data. These could however represent variations that might predispose women to cervical cancer.

Validation of the data by Customized SNP Array: In the SNP array 1455/3178 novel nonsynonymous and 192/700 known single nucleotide variations were validated. The overall concordance rate between the exome sequencing and SNP array data was found to be 33%. However, non confirmation in the array did not mean that a particular variation call was incorrect since a very high threshold was used as cut off while analyzing the array data. Although, most of the SNVs that were validated were germline in nature, a set of 22 novel, nonsynonymous and somatic variations were obtained which included important candidates such as CD97, YWHAH and CD52. These somatic SNVs might help in understanding the mutational landscape of cervical cancer.

CNV and LOH identification: CNV in chromosomes 3, 8 and 17 and LOH in chromosomes 2, 3, 5 were observed for one of the paired samples. Though this data has not been validated, this is in agreement with previous published reports.

DISCUSSION

We have tried to obtain a broad overview of cervical cancer by addressing various factors – type of HPV infection, viral copy number, viral integration and genetic alterations that might play a role in cervical carcinogenesis in a cohort of Indian women with locally advanced cervical cancer.

The HPV genotyping study demonstrated 95% HPV positivity using three different primer sets. It is apparent from the results that a single set of primers is not sufficient to estimate the true HPV infectivity. Of the various HPV genotypes HPV16 was most common (60%), followed by infection with HPV18 alone (2%). Dual infection with HPV16/18 (6%) and HPV16/45 (3%) was also observed. These results are in concordance with other studies from the Indian subcontinent which reports 57-

65% HPV16 positivity, followed by HPV18, 45, and 33 in cervical neoplasia (Basu et al., 2009; Bhatla et al., 2008).

For the viral integration study, we chose APOT assay in order to limit the analysis to integration sites with a transcriptionally active viral genome. Also, APOT assay allows detection of integrated viral genome in clinical lesions even in the presence of a large excess of nonintegrated episomal form of viral genomes (Klaes et al., 1999; Vinokurova et al., 2008). The frequency of viral integration into the host genome in cervical carcinomas has been reported to be as high as 100% in $HPV18^+$ tumours (Corden et al. 1999) and up to 80% in $HPV16^+$ tumours (Melsheimer et al., 2004) . In our study we found 4 $HPV18^+$ samples where the virus was integrated and one $HPV18^+$ sample where the virus was episomal. The incidence of integration in $HPV16^+$ samples was 94%. The preference of integration for the chromosomal loci 1p, 3q, 6q, 11q, 13q, 6q and 20q indicate that integration might not be a random event. Most of the integrations (28/48) were found to be located within or near certain genes. This could indicate that the virus prefers transcriptionally active regions for the integration event. Such genes included oncogenes such as myc, transcription factors like TP63, MECOM, etc. The importance of genes like myc and TP63 has already been demonstrated in the context of HPV integration (Wentzensen et al., 2002). Again, integration of the virus near or within fragile sites is a frequently reported phenomenon (Kraus et al., 2008; Thorland et al., 2000). In our study 29/48 integrations were located within or near a common or rare fragile site which is in concordance with previous reports. Comparison of the physical state of the virus (episomal/integrated) with the clinical outcome after radical radiotherapy revealed that patients with episomal form of the virus had increased disease free survival compared to those with integrated form. This observation is supported by various reports which state that the integration event is associated with a

decreased disease free survival (Vernon et al., 1997). However, there are contrasting reports as well, according to which physical state of the virus does not correlate with disease free survival. This needs to be studied further.

Though infection by HPV plays a major role in the etiology of cervical cancer, possible role of several genetic alterations, such as Point mutations, Chromosomal aberrations etc. cannot be completely ruled out. Especially among older women genetic mutations may be a reason for cervical cancer acting in synergy with a lower immune response. Studies have identified genes such as LKB1, FGFR3, PI3KA, Ras, Smad, etc. to be frequently mutated in cancer of the cervix (Cui et al., 2009; Maliekal et al., 2003; Wong et al., 1995; Yee et al., 2000). However, these studies were mostly based on candidate gene approach and till date there are no reports describing the mutational landscape of the disease. Therefore, we attempted to identify the alterations that might play role in cervical carcinogenesis by carrying out whole-exome sequencing of 6 locally advanced cervical biopsies (3 paired and 3 unpaired). Apart from helping understanding the mutational landscape of cervical cancer, these novel variations would be a valuable source of information about the Indian genome which the dbSNP currently lacks. We obtained about 6114 novel nonsynonymous variations unique to this population. A proportion of this data was also validated by customized DNA microarray and sanger sequencing. The significance of the exome sequencing analysis lies in the fact that it provides significant information about the Indian genome with special reference to cervical cancer. To the best of our knowledge, this is the first report where we have applied whole exome sequencing for identifying single nucleotide variations, short indels, LOH and CNV in cervical cancer in Indian women.

REFERENCES

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.

Basu, P., Roychowdhury, S., Bafna, U.D., Chaudhury, S., Kothari, S., Sekhon, R., Saranath, D., Biswas, S., Gronn, P., Silva, I., *et al.* (2009). Human papillomavirus genotype distribution in cervical cancer in India: results from a multi-center study. Asian Pac J Cancer Prev *10*, 27-34.

Bhatla, N., Lal, N., Bao, Y.P., Ng, T., and Qiao, Y.L. (2008). A meta-analysis of human papillomavirus type-distribution in women from South Asia: implications for vaccination. Vaccine *26*, 2811-2817.

Corden, S.A., Sant-Cassia, L.J., Easton, A.J., and Morris, A.G. (1999). The integration of HPV-18 DNA in cervical carcinoma. Mol Pathol *52*, 275-282.

Cui, B., Zheng, B., Zhang, X., Stendahl, U., Andersson, S., and Wallin, K.L. (2009). Mutation of PIK3CA: possible risk factor for cervical carcinogenesis in older women. Int J Oncol *34*, 409-416.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A *85*, 8998-9002.

Gravitt, P.E., Peyton, C.L., Alessi, T.Q., Wheeler, C.M., Coutlee, F., Hildesheim, A., Schiffman, M.H., Scott, D.R., and Apple, R.J. (2000). Improved amplification of genital human papillomaviruses. J Clin Microbiol *38*, 357-361.

Jeon, S., Allen-Hoffmann, B.L., and Lambert, P.F. (1995). Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. J Virol *69*, 2989-2997.

Jeon, S., and Lambert, P.F. (1995). Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. Proc Natl Acad Sci U S A *92*, 1654-1658.

Josefsson, A.M., Magnusson, P.K., Ylitalo, N., Sorensen, P., Qwarforth-Tubbin, P., Andersen, P.K., Melbye, M., Adami, H.O., and Gyllensten, U.B. (2000). Viral load of human papilloma virus 16 as a determinant for development of cervical carcinoma in situ: a nested case-control study. Lancet *355*, 2189-2193.

Klaes, R., Woerner, S.M., Ridder, R., Wentzensen, N., Duerst, M., Schneider, A., Lotz, B., Melsheimer, P., and von Knebel Doeberitz, M. (1999). Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res *59*, 6132-6136.

Kleter, B., van Doorn, L.J., ter Schegget, J., Schrauwen, L., van Krimpen, K., Burger, M., ter Harmsel, B., and Quint, W. (1998). Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. Am J Pathol *153*, 1731-1739.

Kraus, I., Driesch, C., Vinokurova, S., Hovig, E., Schneider, A., von Knebel Doeberitz, M., and Durst, M. (2008). The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. Cancer Res *68*, 2514-2522.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc *4*, 1073-1081.

Lillo, F.B., Lodini, S., Ferrari, D., Stayton, C., Taccagni, G., Galli, L., Lazzarin, A., and Uberti-Foppa, C. (2005). Determination of human papillomavirus (HPV) load and type in high-grade cervical lesions surgically resected from HIV-infected women during follow-up of HPV infection. Clin Infect Dis *40*, 451-457.

Maliekal, T.T., Antony, M.L., Nair, A., Paulmurugan, R., and Karunagaran, D. (2003). Loss of expression, and mutations of Smad 2 and Smad 4 in human cervical cancer. Oncogene *22*, 4889-4897.

Melsheimer, P., Vinokurova, S., Wentzensen, N., Bastert, G., and von Knebel Doeberitz, M. (2004). DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. Clin Cancer Res *10*, 3059-3063.

Munoz, N., Bosch, F.X., de Sanjose, S., Herrero, R., Castellsague, X., Shah, K.V., Snijders, P.J., and Meijer, C.J. (2003). Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med *348*, 518-527.

Peter, M., Stransky, N., Couturier, J., Hupe, P., Barillot, E., de Cremoux, P., Cottu, P., Radvanyi, F., and Sastre-Garau, X. (2010). Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. J Pathol *221*, 320-330.

Romanczuk, H., and Howley, P.M. (1992). Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. Proc Natl Acad Sci U S A *89*, 3159-3163.

Sathirapongsasuti, J.F., Lee, H., Horst, B.A., Brunner, G., Cochran, A.J., Binder, S., Quackenbush, J., and Nelson, S.F. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics *27*, 2648-2654.

Schmitt, M., Bravo, I.G., Snijders, P.J., Gissmann, L., Pawlita, M., and Waterboer, T. (2006). Bead-based multiplex genotyping of human papillomaviruses. J Clin Microbiol *44*, 504-512.

Swan, D.C., Tucker, R.A., Tortolero-Luna, G., Mitchell, M.F., Wideroff, L., Unger, E.R., Nisenbaum, R.A., Reeves, W.C., and Icenogle, J.P. (1999). Human papillomavirus (HPV) DNA copy number is dependent on grade of cervical disease and HPV type. J Clin Microbiol *37*, 1030-1034.

Thorland, E.C., Myers, S.L., Persing, D.H., Sarkar, G., McGovern, R.M., Gostout, B.S., and Smith, D.I. (2000). Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. Cancer Res *60*, 5916-5921.

Vernon, S.D., Unger, E.R., Miller, D.L., Lee, D.R., and Reeves, W.C. (1997). Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer. Int J Cancer *74*, 50-56.

Vinokurova, S., Wentzensen, N., Kraus, I., Klaes, R., Driesch, C., Melsheimer, P., Kisseljov, F., Durst, M., Schneider, A., and von Knebel Doeberitz, M. (2008). Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. Cancer Res *68*, 307-313.

List of Abbreviations

µg	: microgram
µl	: microlitre
ADC	: Adenocarcinoma
APOT	: Amplification of Papillomavirus Oncogene Transcripts
bp	: Base pairs
BWA	: Burrows-Wheeler Aligner
CFS	: Common fragile sites
CIN	: Cervical Intraepithelial Neoplasia
CNV	: Copy number variation
COSMIC	: Catalogue of somatic mutations in cancer
dbSNP	: Database SNP
DEPC	: Diethylpyrocarbonate
DNA	: Deoxyribonucleic acid
DNase	: Deoxyribonuclease
DTT	: Dithiothreitol
E6-AP	: E6 - associated protein
EDTA	: Ethylenediaminetetraacetic acid
Epi	: Episomal
EtBr	: Ethidium Bromide
FBN1	: Fibrillin-1
FGF7	: Fibroblast growth factor 7
FIGO	: International Federation of Gynaecology and Obstetrics
GAPDH	: Glyceraldehyde-3-Phosphate Dehydrogenase
H&E	: Hematoxylin and eosin
HGMD	: Human Gene Mutation Database
HR	: High risk
HPV	: Human Papillomavirus
Indels	: Insertions and deletions
Int	: Integrated
KEGG	: Kyoto Encyclopedia of Genes and Genomes
LB	: Luria Bertani broth
LOH	: Loss of heterozygosity
LR	: Low risk
M	: Molecular weight marker
MAP3K3	: mitogen-activated protein kinase kinase kinase 3
min	: Minutes
mM	: millimolar
ND	: Not determined
NFAT5	: Nuclear factor of activated T-cells 5
ng	: nanogram

NGS	: Next generation sequencing
OMIM	: Online Mendelian Inheritance in Man
ON	: Overnight
ORF	: Open reading frame
PCR	: Polymerase chain reaction
Polyphen-2	: Polymorphism Phenotyping v2
PTPRJ	: Receptor-type tyrosine-protein phosphatase eta
Rb	: Retinoblastoma protein
RFS	: Rare fragile sites
RNaseA	: RibonucleaseA
RNASEL	: Ribonuclease L
Rpm	: Revolutions per minute
RT	: Room Temperature
RT-PCR	: Reverse transcription polymerase chain reaction
SAM	: Sequence Alignment/Map
SCC	: Squamous cell carcinoma
SDS	: Sodium Dodecyl Sulphate
sec	: Second
SIFT	: Sorting Intolerant From Tolerant
SNP	: Single nucleotide polymorphism
SNV	: Single nucleotide variation
SS II	: Superscript II
TAE	: Tris acetate EDTA
TP53	: Tumour protein 53
UCSC	: University of California, Santa Cruz
UV	: Ultraviolet
ZFP64	: Zinc finger protein 64 homolo

List of Tables

Chapter 1

Introduction and Review of Literature

1.1 Introduction

Dr. Hutchison's words "***Lets us not make the cure of the disease more unbearable than the disease itself***" had been one of the most enduring and inspiring lines in medicine. Inspite of having being spoken almost a century ago, these profound words still find a haunting echo in today's context, particularly when it comes to diagnosing and treating cancer. Keeping in agreement with its Latin meaning – a 'crab', the disease is too hard to crack just like the Crustacean. This is clearly reflected in the statistics that indicates the disease as one of the principal causes of deaths worldwide, including India, where it accounts for approximately 6% of total human deaths (Dikshit et al., 2012; GLOBOCAN, 2008a)

Cancer of the cervix is one of the most prevalent forms of cancer worldwide, the major burden of the disease being felt in developing countries like India. Although, early detection and advancement in diagnostic and treatment modalities have led to improved disease management and increased survival of patients in developed countries, in India, cervical carcinoma still continues to be the most common cancer among women and accounts for the maximum cancer deaths each year (GLOBOCAN, 2008b; IARC, 2009). Persistent infections with high-risk (HR) Human Papillomaviruses, such as HPV 16, 18, 31, 33 and 45 have been identified as a major, although not sufficient etiological factor in the development of the disease. The relationship between exposure to HPV and the disease might vary with the infecting HPV type, viral integration status and viral load; these factors could have profound implications on patient prognosis (Josefsson et al., 2000; Woodman et al., 2007).

Till date 120 official HPV types have been identified. However, the prevalence and distribution of HPV types in cervical neoplasias vary with geographic regions and by grade of disease (Clifford et al., 2005; Insinga et al., 2008; Smith et al., 2007).

Therefore, studies involving detection and genotyping of the virus in a population have become extremely important. At least 50% of sexually active men and women get HPV at some point in their lives, but in most cases the infection regresses spontaneously and only in about 10% cases it persists and progresses to high-grade cervical intraepithelial neoplasia. This generally occurs through integration of the HPV genome into the host chromosome with associated loss or disruption of E2 (Wentzensen et al., 2004). In the absence of E2 the expression of E6 and E7 increases eventually leading to immortalization and transformation of the cells (Romanczuk and Howley, 1992).

Besides infection by HPV, genetic instability has been reported as a significant event in disease pathogenesis. Genetic instability may range from point mutations, copy number changes, chromosomal rearrangements to widespread aneuploidy. Identification of such alterations through high throughput techniques such as Next Generation Sequencing would not only provide a detailed picture of the genomic landscape of the disease but would also be significant for understanding disease prognosis and response to therapy.

Studies addressing all the major aspects involved in pathogenesis of cervical cancer such as incidence of HPV infection and contribution of its different genotypes; physical state and site of viral integration combined with viral load in cervical lesions as well as genetic alterations would help in better understanding the disease and be a step forward towards identifying biomarkers and newer treatment modalities for management and cure of cervical cancer.

1.2 Review of Literature

With about 134000 new cases and 72800 deaths annually, cervical cancer ranks as the 1st most frequent cancer among women in India. Characterized by abnormal bleeding, pelvic pain and unusual heavy discharge, the disease develops in the tissues of the cervix, a part connecting the upper body of the uterus to the vagina. It comprises of *endocervix* or the upper part which is close to the uterus and covered by glandular cells; and the *ectocervix,* the lower part which is close to the vagina and covered by squamous cells. The two regions of the cervix meet at the **'transformation zone'** (Fig. 1.1). It is this region where most cervical cancers begin to develop.

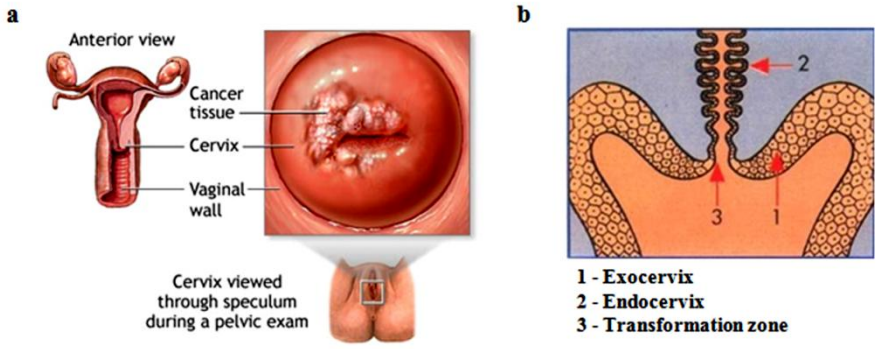

Fig 1.1: Cervical cancer- facts. *Section through the cervix showing malignant growth (a) (A.D.A.M. Medical Encyclopaedia). Diagram showing different stratas of the cervix and their characteristic cell types (b). The zone of transformation marks the site of initiation of malignancy in most cases (Viernes, 2010).*

Based on the histology, carcinomas of the cervix may be classified into the following groups, with each group further having several morphological variants:

- **Squamous carcinomas** - Carcinomas arising from the ectocervical epithelium and characterized by the highest incidence rate, 85-90%.
- **Adenocarcinomas** - Carcinomas arising from endocervical columnar/glandular epithelium and constituting the remaining 10-15%

- **Adenosquamous carcinomas or mixed carcinomas** - Cancers with features of both squamous cell carcinomas and adenocarcinomas, constitute a small percentage (<1%) of the disease burden

1.2.1 Incidence of cervical cancer

Cervical cancer is the third most common cancer in women worldwide, and the seventh overall. Majority of this global burden is felt in low & middle income developing countries (WHO, 2009) and in low socio-economic groups within countries (Kurkure and Yeole, 2006).

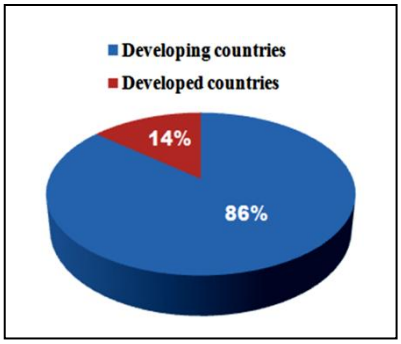

Fig. 1.2: Estimated cervical cancer cases in developed and developing countries. *The incidence of cervical cancer in developing countries is approximately 72% higher than that of developed countries. (Globocan, 2008).*

Five year survival rates of less than 21% are reported from developing countries whereas rates as high as 70% are reported from developed regions like United States. India has a disproportionately high burden of cervical cancer, with the highest rate of incidence and mortality among Indian women (GLOBOCAN, 2008b; IARC, 2009).

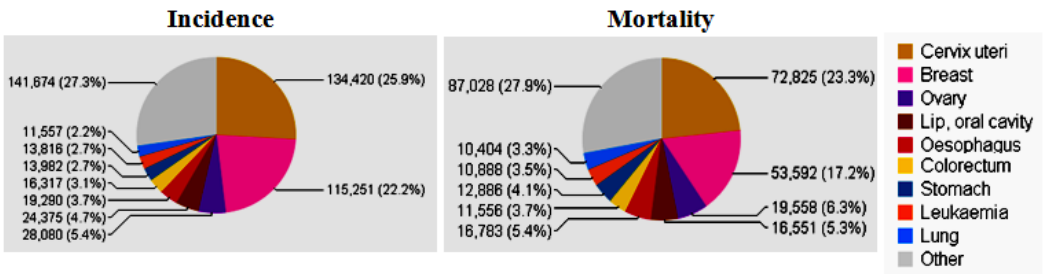

Fig. 1.3: Incidence and Mortality of cervical cancer in India. *Cancer of the cervix has the highest incidence and mortality rate in Indian women (Globocan, 2008).*

The incidence of the disease begins to rise at ages 30–39 and peaks in the fifth or sixth decade of life (WCR, 2008). While the cumulative risk of incurring the disease before the age 64 is 1.19% for a female if world population is considered, it is 2.10% for an Indian female, making them highly susceptible group. The age-adjusted incidence is highest in Chennai and lowest in Thiruvanathapuram. There is also a high incidence belt in the north eastern districts of Tamil Nadu, as well as in two districts in the North-Eastern region of the country. Also, compared to the world and Southeastern Asia, the age specific mortality from the disease is highest in India, the five year overall survival rate being only 48% (Sankaranarayanan et al., 1998).

With the given rate of incidence and mortality, the estimated number of new cervical cancer cases and deaths in India is projected to increase by a large fraction by 2025 (WHO, 2009).

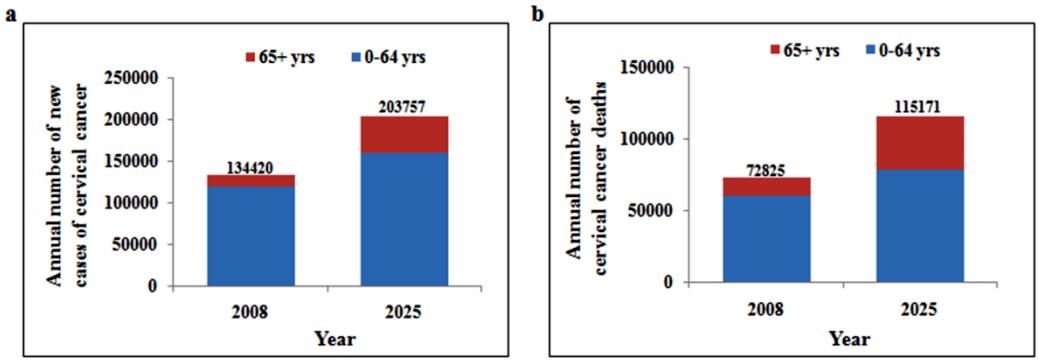

Fig. 1.4: Estimated number of cervical cancer cases and deaths in India projected in 2025. *The estimated number of cervical cancer cases and deaths is expected to increase by 86% and 87% in the age group of 65+ by 2025 (IARC Globocan, 2008).*

1.2.2 Natural History of Cervical Cancer

Cancer of the cervix marks its beginning with the development of pre-cancerous, benign lesions. WHO classification describes the first stage of development as mild dysplasia, which can then progress to moderate dysplasia, severe dysplasia, and carcinoma in situ (CIS) or invasive cervical cancer, with increasing degrees of severity. Mild dysplasia usually regresses spontaneously without treatment. However, in a small percentage of

women it progress to more severe forms. Women with moderate to severe dysplasia are at high risk of developing invasive cancer, although the progression may take several years. There are two other systems of classification-

The Cervical Intraepithelial Neoplasia (CIN) system (Buckley et al., 1982) is based on the degree of involvement of epithelial thickness by the atypical cells. According to this system, mild to moderate dysplasia are classified as CIN1 and involves lower one third of the mucosa, intermediate dysplasia involving two thirds of the mucosa as CIN2, and severe dysplasia and carcinoma in situ are together classified as CIN3, where the whole epithelial layer is replaced but with no disruption of the basal membrane.

The Bethesda system simplifies it further, by classifying CIN1 as Low Grade Squamous Intraepithelial Lesion (LSIL), and both CIN2 and CIN3 as High Grade Intraepithelial Lesion (HSIL).

The invasive form of carcinoma is further classified into various stages, as per the International Federation of Gynaecology and Obstetrics (FIGO) (Sankaranarayanan and Wesley, 2003). This classification is often used to decide the treatment options.

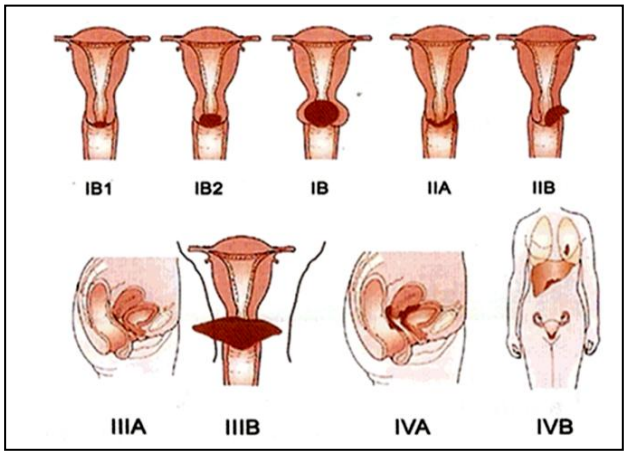

Fig 1.5: FIGO staging of cervical cancer. *The FIGO stages I-IV of cervical carcinoma has been depicted* (temanggunggaul12.blogspot.com; Sankaranarayanan and Wesley, 2003).

FIGO staging of cervical cancer

Stage I: Carcinoma strictly confined to the cervix.

- Stage IA: Invasive cancer identified only microscopically. Invasion is limited to measured stromal invasion with a maximum depth of 5 mm and no wider than 7 mm.
- Stage IA1: Measured invasion of the stroma no greater than 3 mm in depth and no wider than 7 mm diameter.
- Stage IA2: Measured invasion of stroma greater than 3 mm but no greater than 5 mm in depth and no wider than 7 mm in diameter.
- Stage IB: Clinical lesions confined to the cervix or preclinical lesions greater than Stage IA.
  - Stage IB1: Clinical lesions no greater than 4 cm in size.
  - Stage IB2: Clinical lesions greater than 4 cm in size.
- **Stage II:** Carcinoma extends beyond the cervix but does not extend into the pelvic wall. The carcinoma involves the vagina, but not as far as the lower third.
- Stage IIA: No obvious parametrial involvement. Involvement of up to the upper two thirds of the vagina.
- Stage IIB: Obvious parametrial involvement, but not into the pelvic sidewall.
- **Stage III:** Carcinoma that has extended into the pelvic sidewall. On rectal examination, there is no cancer-free space between the tumour and the pelvic sidewall. The tumour involves the lower third of the vagina.
- Stage IIIA: No extension into the pelvic sidewall but involvement of the lower third of the vagina.
- Stage IIIB: Extension into the pelvic sidewall or hydronephrosis or non-functioning kidney.
- **Stage IV:** Carcinoma that has extended beyond the true pelvis or has clinically involved the mucosa of the bladder and/or rectum.
- Stage IVA: Spread of the tumour into adjacent pelvic organs.
- Stage IVB: Spread to distant organs.

1.2.3 Etiology of cervical cancer

Infection by HR-HPV has been identified as the key etiological factor for cervical carcinogenesis. The infecting HPV type, viral integration and viral load – all these have profound implication on the disease. Besides HPV infection, the disease is also associated with a genetic component. Genetic variations ranging from point mutations, single nucleotide polymorphisms and chromosomal abnormalities have been identified to contribute to disease development. Apart from these, several exogenous cofactors, such as hormonal contraceptives, smoking, parity, and co-infection with other sexually transmitted agents also play an important role. Each of these major factors is described:

1.2.3.1 Human papillomavirus (HPV) Infection

The association between HPV infection and carcinoma of the cervix has its origin in the works of Prof. zur Hausen in the early 1980s. Since then, several studies have established that persistent infection with HR- HPV is the major risk factor for the development of high-grade precancerous and cervical carcinoma (de Villiers et al., 1981; Gissmann and zur Hausen, 1980; Nobbenhuis et al., 1999). More than 30 to 40 types of HPV are typically transmitted through sexual contact and at least 50% of sexually active men and women get HPV at some point in their lives. Most HPV infections in young females are transitory and have little long-term significance. However, in 5-10% cases the infection persists and can progress to invasive cervical cancer. Besides being a key etiological factor for cervical cancer, there is growing evidence for the central role of HPV in oral carcinoma as well as cancers of other anogenital sites (Fig. 1.6).

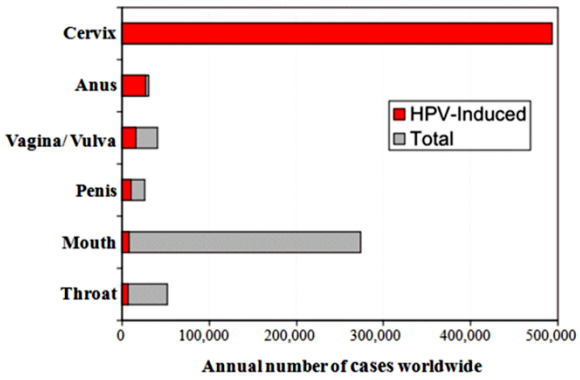

Fig. 1.6: Frequency of HPV contribution in several HPV related cancer. *The annual number of cases worldwide for each of these is depicted. The fraction of cancers estimated to be induced by HPV is shown in red (Wikipedia).*

1.2.3.1.1 HPV Genome

Papillomaviruses belong to the Papovaviridae family, characterized by a small non-enveloped DNA genome with a virion size of ~55 nm in diameter. The virus consists of a double stranded DNA genome (~7800-7900 bp) and an icosahedral capsid of 72 capsomers, which contain at least two capsid proteins, L1 and L2. The HPV genome can

be divided into three regions - the noncoding long control region (LCR) or the upper regulatory region (URR), and the protein encoding early (E) and late (L) gene region (Fig. 1.7). The three regions in all papillomaviruses are separated by two polyadenylation (pA) sites: polyA Signal 1 and polyA Signal 2.

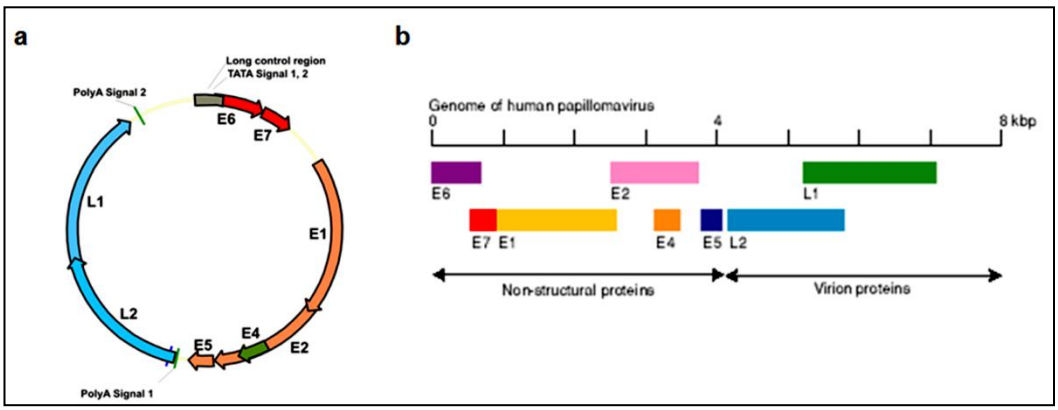

Fig. 1.7: Organization of the HPV genome. *The circular form (a) and the simplified linearised form (b) are depicted. The 'E' and 'L' regions represent the early and late gene regions (Wikipedia and Expert Reviews in Molecular Medicine).*

Most of the *cis*-responsive elements that influence viral transcription and replication are located in LCR, a 400 to 1,000 bp genomic segment. The LCR contains a number of recognizable short motifs that show high conservation across the papillomavirus family. It regulates the production of viral proteins and particles by controlling viral transcription from the early and late regions. This region also contains overlapping binding sites for many different transcriptional activators and repressors such as AP1, glucocorticoid receptor, progesterone receptor NFI TEF-1 Oct-1, etc. which in turn can modulate viral gene expression. Immediately downstream the of the LCR, lies the early region that contains six open reading frames, E1, E2, E4, E5, E6 and E7 and is involved in viral replication and oncogenesis. This is followed by the late region genes. The major function of each of these is summarized in Table 1.1.

Table 1.1: Function of the HPV genes

Viral region	Function
E1	Functions in origin recognition and exhibit both ATPase and 3'-5' helicase activities. Works in conjunction with E2 protein and unwinds the DNA strands for viral replication; also linked to genome maintenance.
E2	Regulates viral replication and transcription by enabling E1 protein to bind to the viral origin of replication in the LCR; encodes a LCR-binding protein that regulates transcription of the early region and facilitates the correct segregation of genomes during cell division. High levels of E2 also have the ability to repress E6 and E7.
E4	Although E4 is part of the early region, it is expressed later in the virus life cycle, when complete virions are being assembled. The exact role played by E4 is not clear, but it is believed to interact with cytokeratin in the epithelial cells.
E5	Interacts with cell membrane growth factors such as EGF and PDGF and increases cellular proliferation and DNA synthesis. Also associated with an increase in mitogen-activated protein kinase activity and down regulation of major histocompatibility complex (MHC) class I molecules.
E6	Viral oncogene. Transforming potential is attributed to its ability to bind to p53 targeting its rapid degradation via ubiquitin ligase; and induction of telomerase activity.
E7	Viral oncogene. Has the ability to bind to hypophosphorylated form of Rb thereby promoting S phase entry; also interacts with inhibitors of cyclin dependent kinases
L1	Encodes the major capsid protein; functions in self-assembling into virus-like particles.
L2	Encodes the minor portions of the capsid and is expressed before L1 to allow for proper construction of the capsid. Has the ability to improve infectivity and packaging.

1.2.3.1.2 Viral life cycle

HPVs are host-specific and show distinct tropism for squamous epithelial cells. The virus enters the epithelial basal layer through mild abrasion or microtrauma and infects the dividing basal cells. The nature of the receptor that facilitates viral entry is still a subject of debate, but it is generally agreed that heparan sulphate proteoglycans might function in the initial binding and/or virus uptake (Joyce et al., 1999; Shafti-Keramat et al., 2003). Besides heparan sulphate proteoglycans, efficient HPV infection is also believed to require secondary receptors such as α6 integrin, laminin 332 and syndecans (Abban and Meneses, 2010; Culp et al., 2006). After being taken into the cell, papillomavirus particles disassemble in late endosomes and/or lysosomes, eventually

transferring the viral DNA to the nucleus with the help of the minor capsid protein L2 (Day et al., 2004). Following infection, the early HPV genes E1, E2, E4, E5, E6 and E7 are expressed, and the virus establishes itself as a stable episome, maintaining its genome as a low copy number, in the basal cells of the epithelium (Fig. 1.8). This stage is mediated by the viral early genes E1 and E2. (Francis et al., 2000). In the suprabasal layers of the epithelium, the virus switches to a rolling circle mode of DNA replication and begins its productive stage (Flores and Lambert, 1997). This involves amplification of the viral genome to higher copy number by the E4 and E5 protein (Wilson et al., 2005), expression of the late genes encoding L1 and L2 that form the capsid proteins in upper epithelial layers and production of viral progeny (Fig. 1.8). In cervical lesions, the viral genome generally gets integrated with associated loss or disruption of E2 (Durst et al., 1985; Jeon et al., 1995) . In absence of functional E2, the transcriptional control on the viral oncogenes, E6 and E7 is lost, thereby increasing their expression (Jeon and Lambert, 1995) and in turn proliferation of the suprabasal epithelial cells (Fig. 1.8).

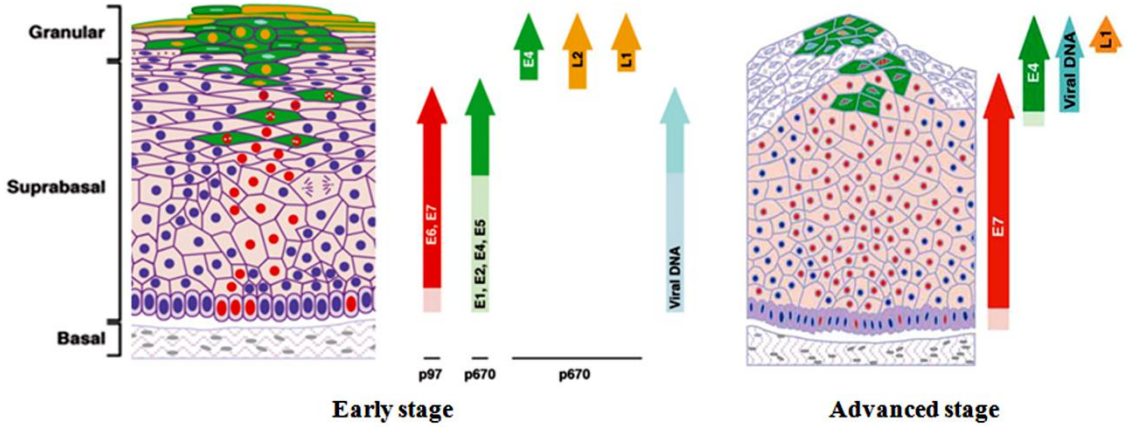

Fig. 1.8: HPV life cycle. *HPV infects the basal layer of epithelium. In the suprabasal layers E6 and E7 drive continued proliferation as cells differentiate and other early genes increase HPV gene expression and genome amplification. In the epithelial granular layers, the late L1 and L2 genes are expressed, forming infectious virions. With increase in disease progression, proliferation of the suprabasal epithelial cells increases mediated by increased expression of E6 and E7 (Doorbar, 2006).*

1.2.3.1.3 Molecular Pathophysiology of HPV infection- Role of E6 and E7

The viral oncogenes, E6 and E7 interfere with the normal cell cycle machinery through biochemical interactions with several important cellular molecules.

Viral E6 binds E6 - associated protein (E6 - AP), an endogenous E3 ubiquitin ligase in the epithelial cells. This complex has the ability to bind the critical cell cycle protein p53 thereby targeting it for degradation by the 26S proteasomal pathway (Scheffner et al., 1993). In absence of p53, epithelial cells fail to recognize senescence or apoptosis signals and continue to divide throughout the stratified squamous epithelial layers. Besides, E6 has also been associated with degradation of the pro-apoptotic protein Bak and proteins containing PDZ domain such as human Scribble, MUPP - 1, and MAGI - 1, 2, and 3, as well as induction of the telomerase activity in the cervical cancer cells (Klingelhutz et al., 1996) (Fig. 1.9).

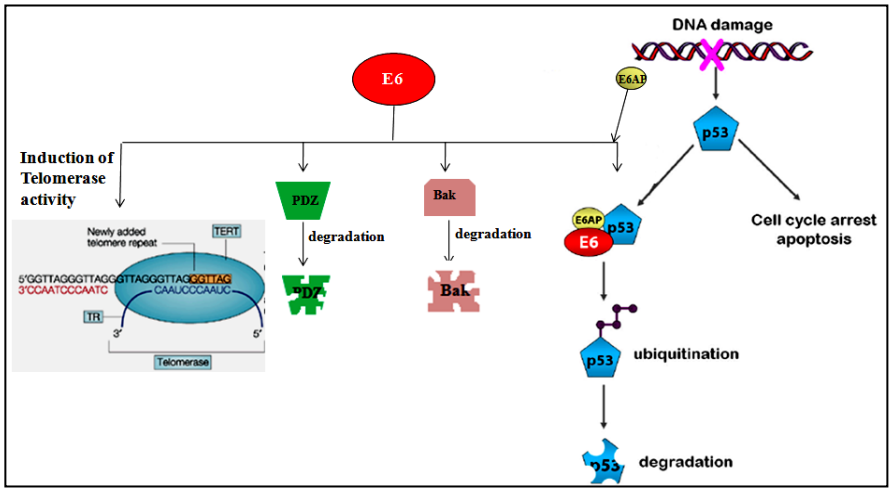

Fig 1.9: Role of E6 in pathophysiology of HPV infection. *E6 targets p53, Bak and PDZ domain - containing proteins, for degradation. It also activates telomerase via induction of expression of the catalytic subunit, hTERT (modified from Jo et al., 2005).*

E7 has been reported to function in tandem with E6 to drive continuous division of differentiated epithelial cells. It has the ability to bind to members of the 'pocket protein' family such as retinoblastoma protein (Rb) and target it for ubiquitin-mediated

degradation by the proteasome (Boyer et al., 1996). Degradation of pRb releases E2F, a transcriptional activator from the pRB/E2F complex, which then activates expression from S - phase promoters resulting in DNA replication and cellular division (Fig. 1.10). Besides, E7 can also associate and interfere with the activity of several endogenous molecules such as cyclin dependent kinase inhibitors $p21^{cip1}$ and $p27^{kip1}$, histone deacetylases and members from the AP1 family of transcription factors. HR- E7 can also activate the proto-oncogene DEK, which may be critical in HPV mediated malignant progression (Wise-Draper et al., 2005). E6 and E7 have also been reported to cooperatively disturb chromosome duplication and segregation during mitosis, thereby inducing severe chromosomal instabilities (Duensing and Munger, 2001).

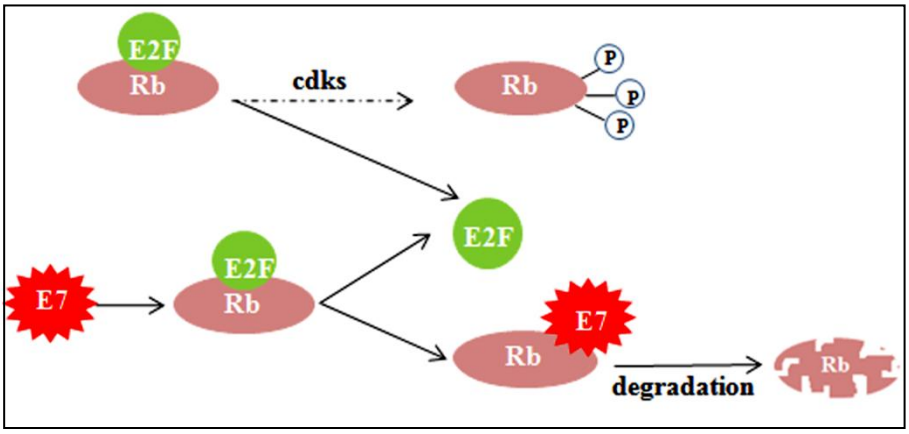

Fig. 1.10: Role of E7 in pathophysiology of HPV infection *E7 targets retinoblastoma protein, for proteasomal degradation, releasing E2F which promotes S - phase entry of the cell cycle (modified from Jo et al., 2005).*

The relationship between exposure to HPV and the disease might vary with the infecting HPV type, viral integration status and viral load (Josefsson et al., 2000; Woodman et al., 2007). Each of these parameters might have profound implications on patient prognosis and are described in greater detail.

1.2.3.1.4 HPV types

Currently, 120 different HPV types are officially recognized (Bernard et al., 2010). These can be classified into five evolutionary groups – Alpha, Beta, Gamma, Mu and Nu, based on whether L1 nucleotide sequence of one is at least 10% dissimilar from that of any other papillomavirus type. HPV types that infect the cervix belong to the group Alpha with over 60 members. The Alpha group also include cutaneous viruses such as HPV2, which cause common warts, and are very rarely associated with cancers. Those belonging to the Beta, Gamma, and Mu and Nu groups are primarily related to cutaneous infections (Fig. 1.11).

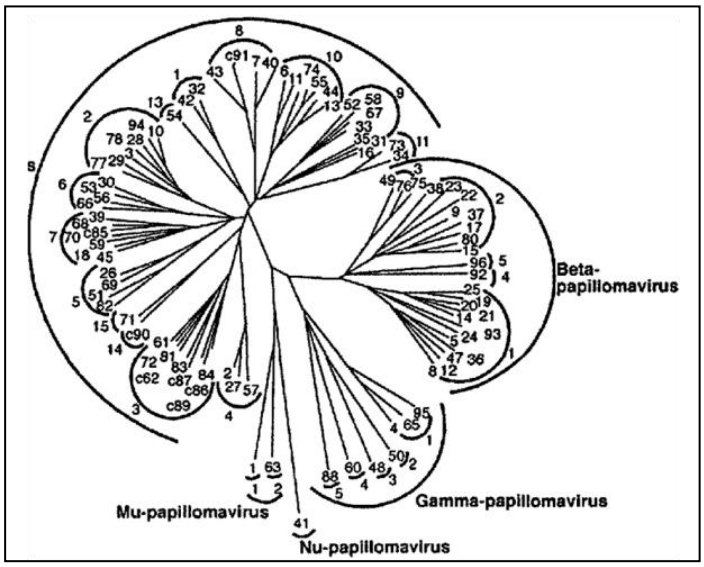

Fig. 1.11: HPV cladogram. *Different HPV types are classified based on their sequence similarities. The genus alpha papillomaviruses primarily infect mucosal epithelium, while genus beta, gamma, mu and nu papillomaviruses primarily infect the skin (Doorbar, 2006).*

Twenty four of the most common members of the Alpha family of papillomaviruses can further be divided into three categories high risk, putative high-risk types and low risk or cutaneous (Munoz et al., 2003; Schmitt et al., 2006) that include:

15 high-risk (HR) types - 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, 82

3 putative high-risk types - 26, 53, and 66

6 low-risk (LR) types - 6, 11, 42, 43, 44 and 70

However, the classification of HPV types into high-risk, putative high-risk and low-risk groups is often confusing especially for the weakly carcinogenic and rare HPV types, and most importantly due to the co-infection of multiple HPV types within the cervical epithelium (Schiffman et al., 2009; Wentzensen et al., 2009).

Methods for HPV detection and genotyping

Detection of HPV infection is quiet challenging since they cannot be cultured in conventional cell cultures, and serological assays for the detection of anti-HPV antibodies have limited accuracy. Moreover, classical direct virological diagnostic techniques, such as electron microscopy and immunohistochemistry, lack sensitivity as well as specificity for routine detection. Therefore, almost all HPV detection techniques that are presently being used, rely on the detection of HPV nucleic acids in a specimen (Poljak and Kocjan, 2010). The most common method for detection as well as HPV genotyping till date is PCR followed by sequencing. The primers used for the PCR can be either type specific primers (Baay et al., 1996) or universal primers that include degenerate MY09/11 and its modified version PGMY 09/11 (Gravitt et al., 2000; Manos et al., 1989), CP65/70, and internal primers, CP66/69 (Berkhout et al., 1995), consensus GP 5/6 and the modified version GP5+/6+ (Jacobs et al., 1997; van den Brule et al., 1990), OBI/II (Jenkins et al., 1991), CPI/CPII (Tieben et al., 1993), SPF1/2 (Kleter et al., 1998), and FAP primers specific for the cutaneous HPV types (Forslund et al., 1999). As opposed to type specific primers, general or consensus PCR primers have the ability to detect a broad spectrum of HPV genotypes. These general primers, also known as universal primers are mostly designed in the L1 region of the viral genome (except for CPI/II primers that are based in the E1 region) which is well-conserved across all HPVs.

Some of the commonly used general primers for detection of mucosal HPV infections are depicted in Fig 1.12.

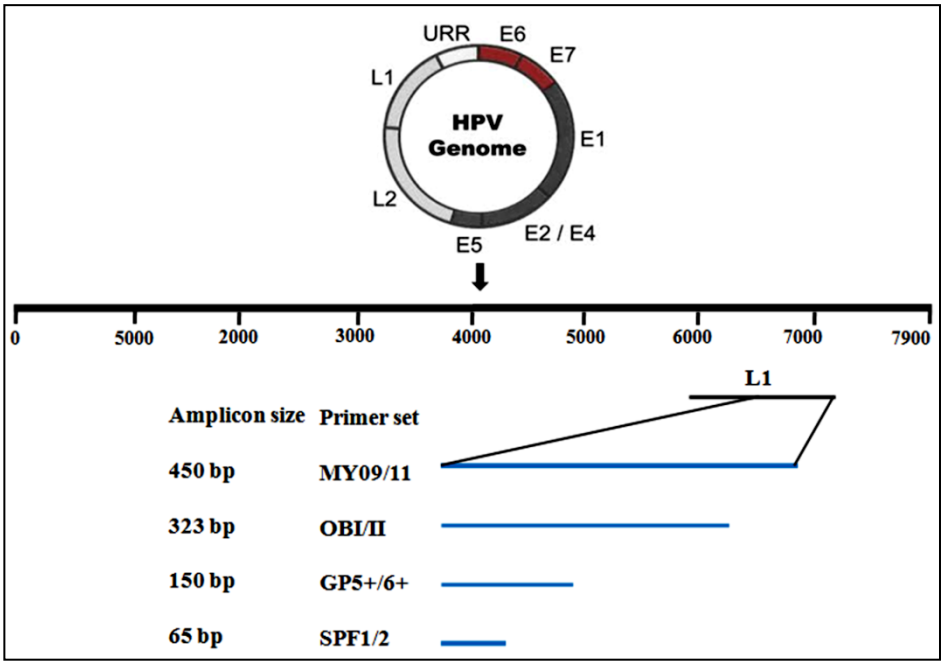

Fig. 1.12: Diagrammatic illustration of the position of the different general primer sets on the HPV genome*. The circular HPV DNA genome, ~8 kb in size is divided into early (E) and late (L) genes. The general primers are usually designed in the L1 region as this is well conserved across the HPV types. The positions of the amplification targets of these primer sets along with the expected amplicon sizes are depicted (modified from Kleter et al., 1999).*

However, with advancement in technologies, multiplex detection and genotyping methods have become popular. Most of these technologies are based on PCR coupled with other high-throughput methods for detection as well as genotyping of multiple HPV types. Some of the important multiplex detection methods available for the alpha-HPVs are summarized in Table 1.2.

Table 1.2: Important currently available methods for multiplex detection of HPVs

Assay	Underlying principle	HPV types detected
Hybrid Capture 2 (hc2) HPV DNA Test (Digene Corporation, USA)	Nucleic acid hybridization assay. Most frequently used HPV diagnostic assay. Does not allow exact determination of HPV type	13 HR-HPVs (cocktail B): 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 68 and five low- risk HPVs (cocktail A): 6, 11, 42, 43 and 44.
Cervista HPV HR Test (Third Wave Technologies USA)	PCR amplification coupled to fluorescence detection. Does not determine the exact HPV types.	14 HPVs using 3 oligonucleotide probe sets: A5/A6 (51, 56 and 66), A7 (18, 39, 45, 59 and 68) and A9 (16, 31, 33, 35, 52, 58)
The Amplicor HPV Test (Roche Molecular Systems, USA)	PCR amplification and detection on microwell plates. Does not determine the exact HPV types	same 13 HPV types as hc2
CareHPV Test (Qiagen)	Based on hc2 technology, allows quick detection (~3 h)	13 HPV types included in the original hc2 plus HPV66,
Cervista HPV 16/18 Test (Hologic)	Real-time PCR. Currently the only FDA approved HPV genotyping assay.	HPV16 and HPV18
INNO-LiPA HPV Genotyping.	Combines PCR amplification and reverse line-blot.	Allows identification of 17–28 (depending on the assay version) HPVs: 6, 11, 16, 18, 31, 33–35, 39, 40, 42–45, 51–54, 56, 58, 59, 66, 68, 70, 73 and 74 (26, 69/71 and 82)
digene HPV Genotyping RH Test RUO	PCR amplification combined with reverse line-blot hybridization.	18 HPV types: 6, 18, 26, 31, 33, 35, 39, 45, 51–53, 56, 58, 59, 66, 68, 73 and 82.
The PapilloCheck HPV-Screening Test (Greiner Bio-One GmbH, Germany)	PCR combined with microarray.	Identifies 24 alpha-HPV types: 6, 11, 16, 18, 31, 33, 35, 39, 40, 42–45, 51–53, -56, 58, 59, 66, 68, 70, 73 and 82.
The Multiplex HPV Genotyping Kit (Progen/Multimetrix, Germany)	PCR combined with bead-based xMAP technology	Detection and identification of 24 HPV types: 6, 11, 16, 18, 26, 31, 33, 35, 39, 42–45, 51–53, 56, 58, 59, -66, 68, 70, 73 and 82.
PreTect HPV-Proofer (NorChip, Norway)	Based on detection of viral E6/E7 mRNA transcripts by nucleic acid sequence-based amplification (NASBA).	5 most common HR-HPV types: 16, 18, 31, 33 and 45.
INFORM HPV (Ventana, USA)	*In situ* hybridization (ISH)	12 HPVs: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 66.

Using one or combination of these methods, a huge number of studies have been carried out across the globe that shed considerable light on the association of HPV infection

with cervical cancer as well as contribution of different genotypes to disease pathogenesis.

Incidence and prevalence of different HPV types in cervical cancer

The prevalence and distribution of HPV types in cervical neoplasias vary with geographic regions and by grade of disease (Clifford et al., 2005; Insinga et al., 2008; Smith et al., 2007). Also, the magnitude of risk of cancer following infection is virus type specific (Chan et al., 1995; Munoz and Bosch, 1997; Munoz et al., 2003). Some of the recent major studies that highlight the global as well as Indian scenario of incidence of HPV infection and the prevalence of different HPV genotypes are summarized in Table. 1.3.

Table 1.3: Incidence and prevalence of HPV – Global and Indian scenario

Worldwide studies				
Study	**No. of cases**	**Method of detection**	**Prevalence (%)**	**Major HPV types (%)**
(Varela et al., 2011)	30,848	Meta-analysis of 243 studies published from 1990 to 2010.	89.9	16 (57), 18 (16), 58 (4.7), 33 (4.6), 45 (4.5), 31(3.8), 52 (3.4), 35 (1.7), 59 & 39 (1.3), 51 (1)
(de Sanjose et al., 2010)	10,575	PCR (SPF10 primers) and enzyme immunoassay Reverse hybridisation line probe assay (LiPA25)13	85	16 (61), 18 (10), 31 &33 (4), 45(6), 52 (3)
(Ciapponi et al., 2011)	5540	Meta-analysis of 62 studies of invasive cervical carcinoma from Latin America and the Caribbean	89	16 (53.2), 18 (13.2), 31 (7.5), 58 (3), 33 (4.3), 45 (4.6), 52 (3.2).
Studies from India				
(Deodhar et al., 2012)	113	Multiplex E7 PCR/APEX assay	92	16 (76), 18 (10.6), 59 (4), 33, 45 (2.7), 58 (4.4), 59 (3.5)
(Bhatla et al., 2008)	558	Meta-analysis of 9 studies from India	94.6	16 (63.3), 18 (15.6),45 (6), 33 (5.4), 35 (5), 58

				(3.6), 59 (2), 31 (1.9), 56 (1.8), 51 (1.3), 73 1.1)
(Franceschi et al., 2003)	191	PCR (GP5+/6+ primers) and enzyme immunoassay	99.4	16(59), 18(12), 33(2.8), 45(2.2), 31(1.7), 51,52,58,59,66,73≤1 %
(Grace Nirmala and Narendhirakannan, 2012)	119	Type specific PCR	-	16 (57%), 18 (18%)
(Basu et al., 2011)	278	PCR (MY09/11 primers) and genotyping; PreTect HPV-Proofer	91.7	16 (59.4), 18 (13.3), 33 (4%), 31 (2.5%), 45 (1.8%), 56 (1.4%), 52, 53, 59, 62, 67, 69, 73 ≤1%
(Basu et al., 2009)	192	MassARRAY; Sequenom	94.8	16 (57.5), 18 (10.4), 16+ 18 (7.3), 33 (3), 59 (2), 45, 56 & 58(1.6),
(Peedicayil et al., 2006)	119	PCR (PGMY 09/11 primers) and the line blot assay	95	16 (60), 18 (14), 31 (3), 33& 58 (6), 35& 45 (5), 42, 51, 56 & 61(2), 62, 64, 81 & 82 (1)
(Pillai et al., 2010)	667	PCR (PGMY09/11 primers) and reverse dot blot assay	92.1	16 (67.8), 18 (7.6), 45 (2.7), 73 (1.8), 31 (1.3), 56 (1.2), 52, 58, 59, 33, 68, 51, 35, 26 & 39<1
(Singh et al., 2009)	110	PCR and line blots	93.6	16 (76.4), 18 (13.7), 45 (3.9), 31 (2.9), 73 (1.96), 33, 56, 58 & 59 <1
(Das et al., 2012)	270	The Multiplex HPV Genotyping Kit; PCR (MY09/11, SPF1/2 and HPV16, 18 Type specific primers)	95	16 (63), 18 (2), 16+18 (6), 16+45 (3.5)
(Saranath et al., 2002)	337	PCR and Southern hybridization	-	16 (73), 18 (16), 16+18 (12)

The above reports clearly bring out the fact that the incidence of HPV infection in carcinoma of cervix is more in the Indian subcontinent as compared to other countries. However, HPV16 and 18 are most the most common types universally and hence clinically more important.

Rather than simply detecting the presence of virus, studies involving further characterization of the infection status such as viral load and its physical status would be much more informative.

1.2.3.1.5 Viral load

A high viral load has been found to be directly associated with the persistence of HPV infection (Dalstein et al., 2003; Ho et al., 1998). Also, there are reports stating direct correlation of viral titre with disease stage, being gradually increased from mild dysplasia to cervical cancer (Abba et al., 2003; Hernandez-Hernandez et al., 2003; Josefsson et al., 2000; Swan et al., 1999; van Duin et al., 2002). However, most of these findings are in context of HPV16 and it has been observed that the viral load of other HPVs such as HPV18, 31 and 45 does not increase with increasing disease severity (Gravitt et al., 2003; Ho et al., 2005; Swan et al., 1999), indicating a type specific relation between copy number and disease prognosis. Further, in the study by Singh et al. in a North Indian cohort, although significant increase in viral loads for HPV16 and 18 was observed from controls through SILs to tumours, but no significant differences was detected between different stages of cancer (Singh et al., 2009). In this context it has also been argued that it is indeed the E6 and E7 expression and not the viral load that is associated with disease prognosis (de Boer et al., 2007). An altogether different finding was presented by Kim et al. wherein they showed that patients with lower HPV viral load showed worse disease free survival (Kim et al., 2009) after radiation therapy. These conflicting results might be due to the variation in sampling techniques and different methods used to calculate viral load. Nonetheless, the relationship between viral load and disease is far more complex than was initially thought. The picture is further complicated by another important characteristic of the virus – viral integration.

While with increasing disease severity, the integrated forms of the virus become dominated, integration itself is followed by a decrease in viral load. In most of the studies that have measured viral load, integration status is not defined. Considering these facts, it becomes apparent that studies involving robust quantitation of type specific viral load in samples with known physical status of the virus could provide useful insights into the pathophysiology of HPV infections and their relationship to disease.

1.2.3.1.6 Viral integration

An important event in cervical carcinogenesis is the integration of the HPV into the host genome. Viral integration has been looked upon as a significant event in progression of cervical cancer from precancerous lesions to invasive carcinoma (Durst et al., 1985; Kalantari et al., 2001; Klaes et al., 1999). The frequency of viral integration into the host genome in cervical carcinomas has been reported to be as high as 100% in HPV18 positive tumours (Corden et al., 1999; Cullen et al., 1991; Pirami et al., 1997) and up to 80% in HPV16 positive tumours (Cullen et al., 1991; Melsheimer et al., 2004; Pirami et al., 1997). However, recent reports have confirmed presence of only episomal form of the virus in advanced cervical squamous cell carcinomas, thereby establishing that integration might not be absolutely mandatory for the process of carcinogenesis (Gray et al., 2010; Vinokurova et al., 2008).

Studies have also been undertaken to study the association between physical state of the virus and disease prognosis. The reports are quiet conflicting; while some report that the integration event is associated with a decreased disease free survival (Kalantari et al., 1998; Vernon et al., 1997) there are others according to which physical state of the virus does not correlate with disease free survival (Holm et al., 2008; Nambaru et al.,

2009). Whatever the case may be, the event of viral integration by itself is a significant episode in cervical carcinogenesis.

Methods of detecting HR-HPV integration

Several strategies have been used to study viral integration. Existing methods can be divided into two broad categories (Pett and Coleman, 2007):

a) Detection of viral integrants those are transcriptionally active by techniques such as Amplification of Papillomavirus Oncogene Transcripts assay (APOT) (Klaes et al., 1999), based on 3' RACE-PCR and RNA ISH (Van Tine et al., 2004)

b) Detection of integrated viral DNA irrespective of its transcriptional status by Southern blotting (Cullen et al., 1991), Quantitative real-time PCR (Peitsaro et al., 2002), Ligation-mediated PCR (Luft et al., 2001), Restriction site-PCR (Thorland et al., 2000) and DNA ISH (Adler et al., 1997; Evans and Cooper, 2004)

Significance of viral integration

The event of integration is characterized by deletion of viral genes essential for synthesis of an infectious virion, and hence is not a normal part of the HR-HPV life cycle. It generally occurs in the E2 region and results in complete or partial disruption of the open reading frame (ORF) for E2 (Woodman et al., 2007). However, the viral oncogenes E6 and E7 oncogenes together with the viral upstream regulatory region are always retained. The integrant derived transcripts usually comprise of viral sequences at their 5'- ends and cellular sequences at their 3'- ends (Type A transcript). In certain rare cases, however, the viral genome either gets disrupted within the E4 region resulting in viral-cellular fusion transcripts that comprises of E6-E7-E1 sequences at their 5'-ends followed by E4 sequences and cellular sequences at their 3'-ends (Type B transcript); or

directly read through from viral to cellular sequences within the E1 gene (Type C transcript) (Fig. 1.13). As opposed to this, transcripts derived from the episomal form mostly comprise of the E1-splice donor signal spliced to the E4-splice acceptor site and terminated at the viral polyadenylation site (Fig. 1.13).

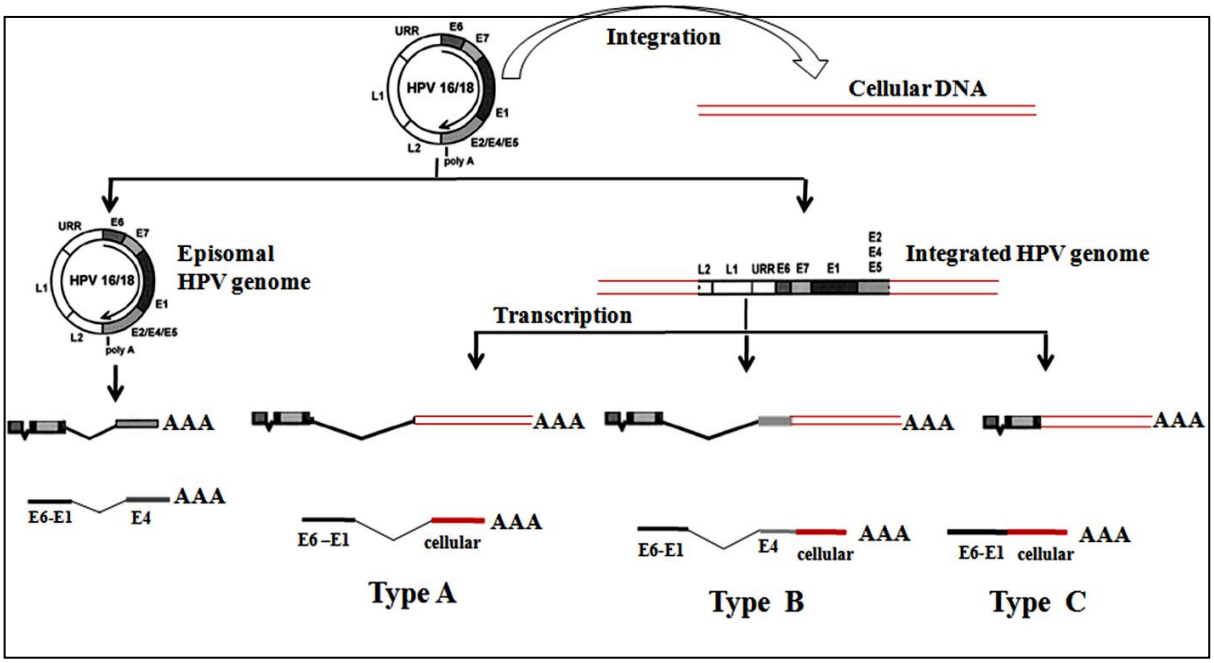

Fig. 1.13: Common types of viral transcripts generated from viral - cellular fusion. *Integration may occur in any of the three ways-Type A, B and C; type A being most common (modified from Klaes et al., 1999).*

The significance of viral integration in the context of cervical carcinogenesis may be due to the following reasons:

- Loss of viral E2 not only inhibits transcription from the integrated viral promoter (Dowhanick et al., 1995; Hwang et al., 1993), but also releases the transcriptional control on E6 and E7, thereby resulting in their increased expression.
- Since hTERT expression is inhibited by E2 (Lee et al., 2002) and activated by E6 (Veldman et al., 2001), HR-HPV integration activates telomerase and, along with E7, brings about immortalization of epithelial cells (Kiyono et al., 1998).

- Disruption of the viral genome results in the failure of early gene transcription from the viral early polyadenylation signal. This leads to the use of host poly(A) signals and generation of stable virus–host fusion transcripts with a longer half life (Couturier et al., 1991; Di Luca et al., 1986; Jeon and Lambert, 1995), which imparts the cells with a selective growth advantage (Jeon et al., 1995).

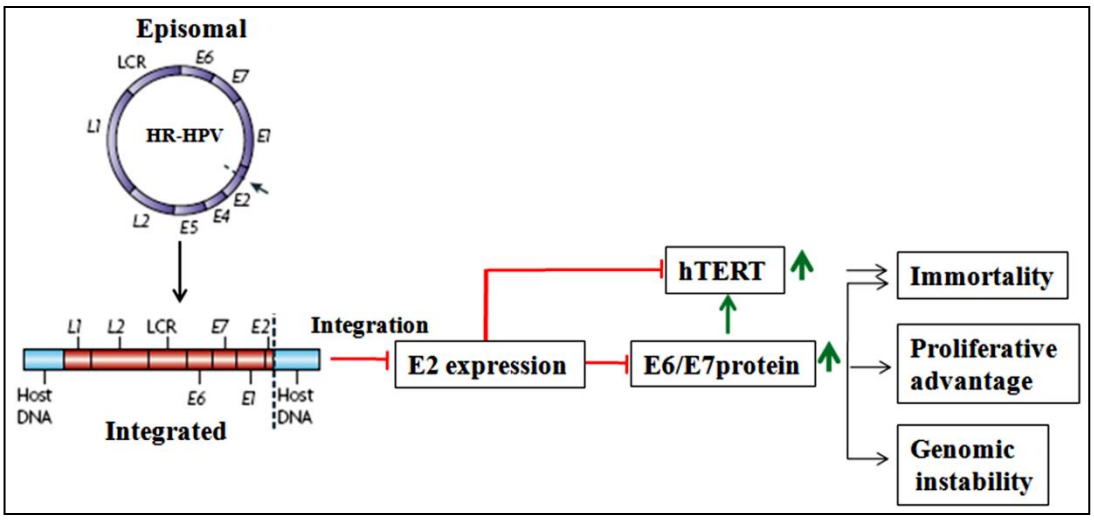

Fig. 1.14: Significance of HR-HPV integration in cervical cancer. *Viral integration results in the disruption of E2 ORF. In absence of E2, expression of the oncogenes E6 and E7 increases, imparting a proliferative advantage to the cells. Loss of E2 combined with overexpression of E6 activates telomerase, shifting the cellular balance towards immortality. Increased Levels of these oncogenes can also bring about genomic instability (adapted from Pett and Coleman, 2007).*

In addition to the effects mentioned above, integrated viral genes may activate cellular oncogenes or inactivate tumour suppressive genes though insertional mutagenesis, eventually leading to alterations in cellular growth and proliferation. For example, APM-1, a putative tumour suppressor has been reported to get inactivated in a number of cervical carcinoma cells lines as a result of insertional mutagenesis combined with the deletion of its second allele (Reuter et al., 1998). Also, in a number of tumours, viral integration has been observed near the MYC locus (Ferber et al., 2003b; Peter et al., 2006) and into the human telomerase reverse transcriptase (Ferber et al., 2003a). Besides, viral integration can render both viral coding genes as well as the

cellular genes susceptible to epigenetic changes which could regulate their expression. Overall, the process may be looked upon as an event that promotes cellular immortalization, deregulated proliferation, and increased genomic instability, all of which are the cellular hallmarks of cancer.

An additional important and largely overlooked situation arises in case of a cell harbouring mixed forms of the virus, i.e. both integrated and episomal. In such cases, E2 may be available in trans to modulate the expression levels of oncogenic E6 and E7, and that overcoming this inhibition would represent an important event on part of the virus for selection of its integrated form (Arias-Pulido et al., 2006) Therefore, as demonstrated by Pett et al., loss of episomes is as much important as integration of the virus into the host genome for progression of lesions to cervical neoplasia (Pett et al., 2006).

'Hotspots' of viral integration- whether a random event

Studies on viral integration have identified integration sites to be distributed throughout the host genome, with no specific preference for any chromosomal loci (Wentzensen et al., 2004). However, according to more recent reports, although the integration event encompasses almost all the chromosomal loci, certain hotspots can be identified such as 3q28, 4q13.3 , 8q24.21, 13q22.1, 1p36.23, 1p36.33, 3q26.33, 3q28, 6q22.31, 6q23.3, 8p11.21, 9q22.32, 13q22.1 and 20q11.21 (Das et al., 2012; Kraus et al., 2008; Schmitz et al., 2012). Besides, the virus has been shown to prefer fragile sites, translocation break points and transcriptionally active regions for its integration (Koopman et al., 1999; Thorland et al., 2000; Wentzensen et al., 2004; Ziegert et al., 2003). Fragile sites are specific regions in the chromosomes that nonrandomly undergoes break in response

to certain stress, making them as susceptible targets for foreign DNA integration. These can be divided into two types-

- Rare fragile sites (RFSs) – present in less than 5% of the population and are associated with expanded CCG repeats or expanded AT-rich minisatellite repeats (Richards, 2001).
- Common fragile sites (CFSs) - present in all individuals and the mostly induced in response to treatment with the DNA polymerase inhibitor, aphidicolin (Thorland et al., 2003)

A high correlation has been reported between the fragile sites and the HPV integration sites, owing mainly to their accessibility to the viral DNA (Dall et al., 2008; Kraus et al., 2008; Matovina et al., 2009; Thorland et al., 2003; Thorland et al., 2000; Wentzensen et al., 2002). Apart from the fragile sites, several tumour related genes such as myc, *TP63, NR4A2, APM-1, FANCC, TNFAIP2,* and *hTERT* have been shown to be involved in the viral integration process (Wentzensen et al., 2002). Further, a recent study has highlighted an association of site of HPV integration with micro RNAs (Schmitz et al., 2012).

1.2.3.2 Genetic Alterations

Though infection by HPV has been established as a major etiological factor for the genesis of cervical cancer, it may not be sufficient for tumour development. The evidence for this is based on the fact that the disease develops in a small proportion of women who have been infected with HPV and generally arises decades after the initial HPV exposure. Therefore, the possible role of genetic aberrations contributing to malignant transformation and tumour progression cannot be ruled out. These alterations may include chromosomal abnormalities, point mutation (both somatic and germline),

single nucleotide polymorphisms and allelic discrimination that might either be involved directly in causing the disease or increase the susceptibility to developing it. The commonly reported alterations in cervical carcinoma are summarized in Table 1.4 and 1.5.

Table 1.4: Major genetic alterations associated with cervical cancer

Gene	Nature of alteration	Alterations reported	Study
HLA class II antigen	Allelic variation	DQB1*03, DRB1*1501 and DQB1*0602, DRB*13 and DQB1*0603 alleles	(Hildesheim et al., 1998; Madeleine et al., 2002).
LKB1/ STK11	Missense and Nonsense point mutation; Deletion Frameshift	S19X,P6Q, E57X, 140611_1420 del CTC TGTCCCAGGG AAATTCAACTACT H107R ,P294fs, F298L, R304W	(Wingo et al., 2009); COSMIC
CDKN2A	Missense point mutation; Deletion	codon 74 (CTG→ATG); codon 129 (ACC→ATC); del exon 2–intron 2 region, etc.	(Nakashima et al., 1999; Park et al., 1999); COSMIC
TP53	Point mutation, Deletion; Frameshift Polymorphism	Codon 234 (TAC→TGC), Codon 273 (CGT→TGT); Codon 170 (ACG →CCG); Codon 154 (GGC→AGC); Codon 297 (CAC-CCC); Codon (AGC→ACC); Codon 273GCT→TGT; codon 166 (A insertion, frameshift); codon 175 (CGC→CAC); codon 192 (CAG→TAG); Codon 213 (CGA→CGG); R72P	(Andersson et al., 2006; Busby-Earle et al., 1994; Crook et al., 1992; Ikenberg et al., 1995; Kim and Kim, 1995; Liu et al., 1994; Miwa et al., 1995; Storey et al., 1998; Tenti et al., 1998); COSMIC
PTEN	Nonsense point mutation, Deletion	R233X, 800delA, 415delTATT, 1038del16 bp, etc.	(Harima et al., 2001; Minaguchi et al., 2004; Yaginuma et al., 2000); COSMIC
CTNNB1	Point mutation, Frameshift	Codon 45 (TCT→TTT); Codon 47 (AGT→GGT); Codon 37 (TCT→ACT); Codon 45 TCT→TT Frameshift ; Codon 41 (ACC→AAC); Codon 37 (TCT→TTT).	(Shinohara et al., 2001); COSMIC
FGFR3	Point mutation	S249C	(Yee et al., 2000); COSMIC
PI3KA	Missense point mutation	R88Q; E542K; E545K; Q546E; H1047R; M1043I	(Cui et al., 2009; Miyake et al., 2008); COSMIC
Ras (H-ras, K-ras, N-ras)	Point mutation	Codon 12, 13 and 61	(Grendys et al., 1997; Huang et al., 1996; Kang et al., 2007; Lee et al., 1996; Pappa et al.,

			2006; Parker et al., 1997; Wong et al., 1995); COSMIC
Smad	Insertion and frameshift	Ins A- codon 122	(Maliekal et al., 2003)
EpCAM	Non-synonymous polymorphism	M115T	(Hu et al., 2012)
CCND1	Polymorphism	G/A polymorphism in rs9344	(Wang et al., 2012)
CD83	Nonsense point mutation	F138 X	(Zhang et al., 2007)
RNASEL	Germline point mutation	T→C (rs3738579) present in 5'UTR of exon 2, 95 bp upstream from the translation initiation codon.	(Madsen et al., 2008)
BRAF	Point mutation	G595D; G468E ; L596V; G465V	(Kang et al., 2007); COSMIC

Table 1.5: Major chromosomal aberrations associated with cervical cancer

Nature of alteration	Chromosomal region	Study
Amplifications Gains	3q21, 3q26–q29, 5p, 7q22, 8q23–q24, 8q24.3, 9p22, 9p23–24, 10q21, 11q13, 11q21, 11q22–23, 12p13, 14q12, 17q12, 17q25, 19q13.1, 20q11.2, 20q13.1 1q22–q23, 1q25.3–q32.1, 3q24–29, 5p12–p13, 6p, 7q11.22–q11.23, 7q31.1–q31.2, 8q24.13–q24.22, 9q33.2–q34.3, 17q25.1–q25.2, 17q21–q22, 19q13.3, 20q11.22, 20q11.21–q13.33, 20q11.21–q12, 20q13.12–q13.31, 20q12, Xp11.2–p11.3, Xq12, Xq22, Xq28	(Heselmeyer et al., 1996; Hidalgo et al., 2005; Lockwood et al., 2007; Narayan and Murty, 2010; Scotto et al., 2008a; Scotto et al., 2008b; Wilting et al., 2008; Yang et al., 2001)
Losses LOH	2q33–q37, 3p12–23, 4p16.3–p16.1, 4, 4q, 4q28.3–q32.1, 4q13.3, 4q35.2, 4q28, 6q, 8p23.3, 8p12–21.3, 8q23.2–q23.3, 9p, 11p15.5, 11q13.3, 11q22.3–25, 3q12.11–13q14.3, 13q14.3–q21.33, 13q31.1–q31.3, 17p13.3, 18q11.2–18q23 1q21–42, 2q33–q37, 3p13–21.1, 3p14.1–p24, 4p, 4q34–qter, 5p15.1–15.33, 6p21.2, 6p25–21.3, 6q16.3–q27, 8p23–p21, 8q22.2, 9p21, 10p15–p14, 10q23.3, 11p15.5, 11p15.3–p15.11, 11p13–p12, 11q13, 11q23.1–q24.1, 13q12–14, 17p13–p12, 17q12–q21, 17q11.2, 18q21–q22, 19p13.3, 19q13.4, Xq21–q22	(A et al., 2004; Acevedo et al., 2002; Bethwaite et al., 1995; Bhattacharya et al., 2004; Choi et al., 2007; Chuaqui et al., 2001; Herrington et al., 2001; Kersemaekers et al., 1998; Manolaraki et al., 2002; Miyai et al., 2004; Narayan et al., 2003; O'Sullivan M et al., 2001; Pulido et al., 2000; Sherwood et al., 2000; Tsuda et al., 2002)

The mutations involved in cervical cancer, described so far are mostly based on candidate gene approach and till date there are no reports describing the entire genomic landscape of the disease. With advances in technology, it is now possible to carry out such high-throughput studies.

1.2.3.2.1 Next generation sequencing - a new face of cancer research

During the past 5 years, 'next generation' sequencing technologies or NGS have surfaced as a promising tool for genomics research. With the advent of NGS, cancer genomics have moved from focused approaches based on single-gene sequencing and arrays to comprehensive genome-wide approaches. Contrary to Sanger, NGS is based on sequential imaging of the stepwise addition of nucleotides, resulting in the generation of hundreds of megabases to gigabases of nucleotide sequence output in a single instrument run, depending on the platform. Some of the popular NGS platforms available today include Roche 454 GS FLX, Illumina Genome Analyzer, Applied Biosystems SOLiD, Helicos BioSciences HeliScope, etc.

NGS technologies broadly include whole genome sequencing, whole exome sequencing and transcriptome sequencing. Of these whole exome sequencing has gained popularity over others in identifying disease mutation. Though exons constitute only about 1% of the genome (37.6 Mb), sequencing them can yield significant information as they have been reported to harbor most variations (Botstein and Risch, 2003). Also most frequent type of disease mutations are those that cause amino acid substitutions resulting from variations in exons. Further, approaches involving targeted sequencing provide increased sequence coverage of a particular region of interest at high throughput and lower cost compared to whole genome sequencing making it highly suitable for understanding the genetic landscape of cancer (Meyerson et al., 2010). Some of the

major recent studies where exome sequencing has successfully been employed for identifying novel mutations are summarized in Table 1.6. In addition to these genetic alterations, studies have identified an epigenetic component in cervical cancer. Promoter CpG island hypermethylation in genes such as p16, DAPK, MGMT, APC, HIC-1, E-cadherin RARß, FHIT, GSTP1and hMLH1have been associated with cervical carcinogenesis (Dong et al., 2001; Virmani et al., 2001).

Table 1.6: Mutations in different cancers as identified by whole exome sequencing

Type of cancer	Gene harbouring mutation	Study
Myelodysplasia	RNA splicing machinery genes-U2AF35, ZRSR2 and SRSF2	(Yoshida et al., 2011)
Head and Neck Squamous cell Carcinoma	Notch1	(Agrawal et al., 2011)
Prostrate	No. of mutations in 20 genes including TP53, PDZRN3, SDF4, etc.	(Kumar et al., 2011)
Renal cell Carcinoma	PBRM1	(Varela et al., 2011)
Acute Monocytic Leukemia	DNMT3A	(Yan et al., 2011)
Primary Colon cancer	BMPR1A	(Timmermann et al., 2011)
Multiple Endocrine Neoplasia Type 2A	RET Germline mutation	(Qi et al., 2011)
Uveal Melanomas	BAP1	(Harbour et al., 2010)
Kaposi Sarcoma	STM1	(Byun et al., 2010)

1.2.4 Molecular biomarkers in cervical cancer – from diagnosis to prognosis

'Cancer can be cured, if detected early'- this paradigm is apt in the context of cervical carcinoma. Although, early detection combined with improved treatment modalities can help managing the disease and control the death toll to a very large extent, the existing screening and treatment modalities face issues with specificity and sensitivity. Hence, in recent times the main focus of cervical cancer research have shifted towards

identification of molecular biomarkers, which, combined with the existing screening and treatment procedures, are expected to improve diagnosis, prognosis, prediction of response or recurrence, and disease monitoring. A number of biomolecules including HPV oncogenes E6 and E7 are effectively being used as diagnostic biomarkers. E6 and E7 interfere with the cell cycle regulators and bring about a change in the expression pattern of a large number of molecules, thereby contributing to neoplastic progression. Some of the important biomarkers of cervical carcinoma include **mini chromosome maintenance** (Stoeber et al., 2002), c**ell division cycle protein 6, p16INK4A** (Murphy et al., 2003), **squamous cell carcinoma antigen (**serum markers of squamous cell carcinoma) (Duk et al., 1990; Farghaly, 1992) and c**ell proliferation markers- PCNA and Ki-67** (Konishi et al., 1991; Mittal et al., 1993).

Besides these diagnostic biomarkers, molecules such as VEGF and EGFR are being used as potential therapeutic targets in cervical carcinoma. VEGF inhibitors including monoclonal antibodies, such as bevacizumab and small-molecule tyrosine kinase inhibitors (TKIs), such as sunitinib, could be successfully used in clinics either in the form of combinatorial therapy or monotherapy for treating locally advanced or recurrent disease (Monk et al., 2009). Similarly treatment regimen involving targeting of EGFR by small molecules such as cetuximab, gefitinib, erlnotinib or anti-EGFR antibody matuzumab, in combination with standard radiotherapy and chemotherapy protocols are in various phases of clinical trial (Goncalves et al., 2008).

The discovery of these and other plentiful molecular biological markers for diagnostics, therapeutics and prognosis has paralleled advances in high-throughput molecular biologic techniques in the genomic, transcriptomic, and proteomic fields. The use of Next generation sequencing technologies in this context has already been described. Some of other high-throughput technique includes Array Comparative

Genomic Hybridization for identification of chromosomal copy number changes, Single Nucleotide Polymorphism (SNP) Profiling, Gene Expression Profiling, Proteomics and Metabolomics. Of all the high-throughput technologies, the field of microarray technology have witnessed the maximum growth. It has been used extensively for classification, subclass identification, identification of prognostic biomarkers and predictive signatures in the field of cancer biology. Identification of HER2/neu (ERBB2) in metastatic breast cancer and EGFR in metastatic colorectal cancer as potential drug targets as well as determination of gene signatures to predict aggressiveness in prostate cancer and survival in colon cancer are some of the recent achievements of microarray. Not only this, two diagnostic arrays, Oncotype DX (Genomic Health, USA) and MammaPrint assay (Agendia,The Netherlands) have even found their way to the clinics for prediction of breast cancer recurrence and response to therapy.

In the field of cervical cancer, a plethora of biomolecules have been identified from a large number of studies that focuses on diagnostic, predictive and prognostic biomarkers. Differential gene expression pattern has been reported between squamous cell carcinoma and adenocarcinoma of the cervix (Chao et al., 2006; Contag et al., 2004). Besides, overexpression of EGFR, ERBB2, CDKN2A, KRAS, MYCN, KIT, TOP2A have been associated with malignancy of cervical epithelium. Apart from this, microarray has also been employed in cervical cancer to predict treatment response to radiotherapy or concurrent chemo-radiotherapy as well as prognosis after therapy. A number of studies have come up in recent time with different gene signatures in important pathways such as MAPK, apoptosis, metastasis, hypoxia, b-catenin, etc. that can predict response to therapy as well as prognosis (Harima et al., 2003; Huang et al., 2011; Huang et al., 2012; Iwakawa et al., 2007; Rajkumar et al., 2009).

Overall, the availability of high-throughput technologies has vastly broadened the potential for biomarker discovery, however, establishing validity and maintaining quality control throughout each phase of biomarker discovery and development remains the primary concern.

RATIONALE OF THE STUDY

Considering the overwhelming increase in the rate of cervical cancers, particularly in developing countries, approaches addressing all the major aspects of disease pathogenesis are the need of the hour. HPV has been recognized as the major etiological factor for the disease, however, the risk of disease development, progression and prognosis of the infected individuals depend to a great extent on the infecting viral genotype, physical state of the virus (i.e., integrated or episomal) and the viral load. HPV by itself can also bring about several genetic alterations thereby destabilizing the genome. Identifying the association of these viral factors as well as the genetic alterations in the tumour genome, in context of disease development, is important in understanding cervical carcinogenesis.

Pretreatment cervical cancer biopsies as well as blood samples from a cohort of Indian patients with a good follow up were available in the laboratory. This formed an ideal set to address most of the questions related to the pathogenesis of the disease. We undertook a multifaceted approach to study the contribution of several viral cofactors as well as the genetic alterations that are believed to play a role in the genesis of the disease. Through this study we aim towards identification of a set of genetic biomarkers for genesis, progression and prognosis of the disease.

AIMS AND OBJECTIVES

1. Determination of HPV status in cervical cancer biopsies.
2. Identification of integration site of the virus by Amplification of Papillomavirus Oncogene Transcripts (APOT) assay.
3. Identification of genomic alterations in cervical cancer biopsies by exome sequencing
4. Correlation of the data from the study with clinical data.

Chapter 2

Determination of HPV status in cervical cancer biopsies

2.1 INTRODUCTION

Human papillomavirus (HPV) is the major etiologic cause of nearly all carcinomas of the uterine cervix (Bosch and Munoz, 2002; Crum et al., 2003; Walboomers et al., 1999; zur Hausen, 2002). Different HPVs have been categorized as 'types' based on whether L1 nucleotide sequence of a particular type is at least 10% dissimilar from that of any other papillomavirus type. Currently, 120 different HPV types are officially recognized, ranging from HPV1 to HPV124, with HPV46, HPV55, HPV64 and HPV79, initially identified as distinct types, now classified as subtypes as these did not yet meet the criteria of unique HPVs (Bernard et al., 2010). In addition to the 120 official HPVs, several new types were characterized during 2009–2010, the last being HPV152 (Nomenklatura., 2010). Clinically most important HR-HPVs are HPV16 and HPV18, being associated with 50–65% and 7–20% of cases of cervical cancer, respectively (Li et al., 2011). The prevalence and distribution of HPV types in cervical neoplasias vary with geographic regions and by grade of disease (Clifford et al., 2005; Clifford et al., 2003; Insinga et al., 2008; Smith et al., 2007). Also, the magnitude of risk of cancer following infection is virus type specific (Chan et al., 1995; Munoz and Bosch, 1997; Munoz et al., 2003; Van Ranst et al., 1992). Considering the above facts, there is a strong need to identify HPV genotypes in order to investigate the epidemiology and clinical behaviour of particular types. This information would not only serve as the basis to assess the protection offered against cervical cancer by the currently available vaccines, but would enable to identify the potential types that should be included in the vaccine regime in near future so as to offer maximum protection.

Although several studies have been reported from the Indian subcontinent on the prevalence and distribution of HPV genotypes in invasive cervical carcinoma (Bhatla et al., 2008), considering the fact that the major global burden of the disease is felt in this

country, more number of such studies is essential. Several strategies have been effectively used to detect and type different HPVs, ranging from simple PCR using general / type specific primers (Basu et al., 2009; Gravitt et al., 2000; Hart et al., 2001; Soderlund-Strand et al., 2009) followed by sequencing, to high throughput assays such as INNO-LiPA HPV Genotyping, Linear Array HPV Genotyping Test, *Digene* HPV Genotyping RH Test RUO, Multiplex HPV Genotyping Kit, Roche AMPLICOR HPV, etc. Most of the high-throughput methods combine PCR using general primers with techniques like reverse line-blot hybridization, luminex bead array and solid phase microarray.

2.2. MATERIALS and METHODS

MATERIALS

Chemical/ Reagent	***Company***
Ethanol, KOH, KCl, NaOH, Glacial acetic acid	*SD Fine-Chem. Ltd., Mumbai, India*
Lysine, Proteinase K, Phenol, Agarose, Bromophenol blue, Xylene cyanol, SDS, R-phycoerythrin conjugated streptavidin	*Sigma, St. Louis, MO, USA*
Tris , EDTA	*USB corp., Cleveland, OH, USA*
NaCl, Sodium acetate, Chloroform, Isoamyl alcohol, *Glycerol*	*Qualigens* fine chemicals, Mumbai, India
Ethidium bromide, dNTP mix, RNase DNase free water	*Invitrogen, Carlsbad, CA, USA*
PCR reagents, DNA molecular weight markers	*MBI Fermentas Canada Inc., Burlington, Ontario, Canada*
Multiplex HPV Genotyping Kit	*Multimetrix, Heidelberg, Germany*
Hybridization plate, Filter plate	*Luminex Corporation, Austin, TX, USA*
PCR tubes, Tips, Glass slides	*Axygen , Union city CA, USA*

METHODS

2.2.1 Clinical Sample accrual

Pretreatment cervical tumour biopsies, predominantly from FIGO stage IIB and IIIB as well as blood samples were obtained from patients undergoing radiotherapy alone or concomitant chemo-radiation at the Radiation Oncology Department, Tata Memorial Hospital, Mumbai after obtaining IRB approval. A generic consent for basic research was obtained prior to acquiring the blood and biopsies. However, for the current study a consent waiver was obtained from the Hospital Ethics Committee since the samples were collected more than 10 years ago. The biopsies were obtained from histologically proven, primary cervical tumour, before the start of radical radiation therapy and were coded for de-identification by the physician prior to testing. The samples were collected in liquid nitrogen and stored at -80°C freezer till use.

2.2.2 Processing of tumour samples

For most of the cervical biopsies, cryosectioning was performed for determining tumour percentage by H&E staining and for DNA/ RNA isolation. Frozen cervical tumour tissues were attached to cryostat (Leica CM1100) chuck by applying small amount of RNase free water around the tissue and immediately freezing it inside the cryostat. First two 5µm sections of tissues were mounted on lysine coated slides for H&E staining. The stained slides were examined by a pathologist and the tumour percentage was arrived at. Next five 30µm sections were collected for isolation of DNA and seven, 30µm sections for RNA, as described below.

2.2.3 Genomic DNA isolation

Isolation of genomic DNA from mammalian cells is usually done by using a hypotonic lysis buffer containing a detergent (SDS), EDTA and Proteinase K. SDS and Proteinase K lyses the cell membrane while EDTA inhibits the action of DNase by chelating divalent cations like Mg^{2+} required for the activity of the enzyme. This is followed by extraction with phenol chloroform and alcohol precipitation. Phenol helps in removal of protein contaminants by denaturing them, while chloroform removes any lipid components from the lysate leaving nucleic acids in the aqueous phase. Finally DNA is salted out by the effect of 100% alcohol.

For extraction of genomic DNA from the cervical cancer biopsies, five 30 µm sections were collected in 700 µl STE buffer (0.1 M NaCl, 0.05 M Tris pH 7.5, 1 mM EDTA pH 8.0, 1% SDS) containing Proteinase K (20 mg/ml) and incubated at 37°C ON. The lysate was centrifuged at 12000 rpm/10 min/RT. To the supernatant, 700 µl Tris equilibrated phenol was added, mixed vigorously, centrifuged at 12000 rpm/10 min/RT and the upper aqueous phase was carefully collected in a fresh tube. Equal volume of phenol:CIA mix (24:1 v/v) was added to the aqueous phase, mixed vigorously and centrifuged at 12000 rpm/10 min/RT. The upper aqueous phase was collected in a fresh tube and extraction was repeated with CIA alone. Phases were separated by centrifugation as mentioned earlier. To the aqueous phase, two volumes of chilled absolute ethanol and $1/10^{th}$ volume of 3M sodium acetate (pH 7.0) were added, followed by centrifugation at 12000 rpm/10min/RT to precipitate DNA. The precipitated DNA was washed with 1 ml 70% ethanol; air dried and resuspended in ~200 µl TE (pH 8.0). Quantity and quality of DNA was assessed by measuring concentration and OD_{260}/OD_{280}; OD_{260}/OD_{230} ratio on NanoDrop UV-Vis

spectrophotometer. The integrity of DNA was ascertained by 0.8% agarose gel electrophoresis.

2.2.4 Agarose gel electrophoresis

Agarose gel electrophoresis is a routinely used method for separation of nucleic acids. The method uses naturally occurring polymer obtained from an Alga in order to achieve a semi solid gel conformation on which nucleic acids are separated based on their size and shape and visualized as fluorescent bands by intercalating dyes like Ethidium Bromide (EtBr) exposure to UV.

Agarose gel was made in the buffer TAE (0.484% w/v Tris, 0.15% v/v glacial acetic acid and 10% v/v 0.5M EDTA in distilled water) and contained EtBr. About 1µg DNA or 5-7 µl of PCR product was mixed with 6x loading dye containing 0.25% Bromophenol blue, 0.25% xylene cyanol and 30% glycerol in distilled water, and loaded in to the well. The percentage of agarose gel varied (0.8-2%) depending on the size of the DNA to be analyzed. Standard DNA markers of known fragment sizes were run in parallel to the samples in order to have a known reference. The gel was run at a constant voltage not exceeding 100V, and visualized on a Gel documentation system.

2.2.5 Genotyping of HPV

HPV genotyping was determined in 270 cervical cancer biopsies by:

- High throughput Luminex bead based assay using biotinylated $GP5^{+}/GP6^{+}$ primers and allows genotyping of 24 HPV types.
- Polymerase chain reaction with MY09/11 and SPF1/2 primers followed by PCR with HPV16/18 type specific primers.

Each of these methods is described in detail below.

2.2.5.1 Genotyping of HPV by high throughput Luminex assay

Genotyping of 24 HPV types which included 15 high-risk types (16, 18, 31, 33, 35,39, 45, 51,52, 56, 58, 59, 68, 73, and 82), 3 putative high-risk types (26, 53, and 66), and 6 low-risk types (6, 11, 42, 43, 44, and 70) (Munoz et al., 2003) was carried out in 270 cervical biopsies using Multiplex HPV Genotyping kit based on Luminex xMAP technology. In the assay, the 150 bp region of the HPV L1gene was first PCR amplified using biotinylated $GP5^{+}/GP6^{+}$ primers (primer set 1) concurrently with the human β-globin gene, serving as an internal control. The resulting amplicons were then genotyped using 26 distinct fluorescence labelled polystyrene bead populations - 24 HPV type specific, one β-globin specific and one hybridization control specific.

PCR Amplification: PCR reactions were carried out in sterile thin walled 0.6ml capacity microfuge tubes in 50 µl reaction volumes containing 300 ng of DNA template, 2 µl of primer set 1(provided in the kit), 200 µM of dNTP mix, 2U of Taq polymerase, 3.5 mM $MgCl_2$ and PCR buffer.

Table 2.1: Program for PCR amplification by Multiplex HPV Genotyping Kit

Step	Temperature (ºC)	Duration (min:sec)
Initial denaturation	94	15:00
Denaturation	94	0:20
Annealing	38	0:30
Elongation	71	0:80
40 cycle step 2-4		
Final elongation	71	4:00

Luminex array: The biotinylated PCR products which showed amplification on a 2% agarose gel were subjected to genotyping for 24 HPV types in the luminex array, following manufacturer's protocol. Briefly, 10 µl of PCR product was mixed with 40 µl

of bead mix and 1 µl of hybridization control for internal control of hybridization efficiency in each well of a 96 well hybridization plate. A negative control was also set up containing 10 µl water and 1 µl of hybridization control. The plate was incubated at 95°C for 10 min to allow denaturation of the amplified products, followed by 1 min incubation on ice and 30 min incubation at 41°C under agitation (370 rpm) for hybridization. During hybridization, each well of a filter plate was equilibrated with 100 µl wash buffer and incubated 30 min at room temperature. The Wash Buffer was removed by vacuum filtration. The hybridized 'bead mix PCR samples' were mixed thoroughly and transferred to the filter plate. This was followed by washing with 100 µl wash buffer and incubation with 70 µl staining solution at room temperature for 30 min under slight agitation (250 rpm). After twice washes (each with 100 µl wash buffer), 100 µl wash buffer was added, mixed thoroughly for 5 min. and the plate was read in the Luminex analyzer (Luminex Corporation, Austin, TX, USA).

Note: *Staining solution was prepared by diluting the Conjugate (R-phycoerythrin conjugated streptavidin) 1:10 in staining buffer. For each well 75 µl staining solution was prepared. Bead Mix was resuspended by vortexing before adding into each well of the Hybridization plate.*

Data analysis: The Luminex analyzer analyzes the reporter fluorescence signal of at least 100 beads of each individual bead set per well. The fluorescence signals were evaluated by the software of the Luminex analyzer. The fluorescence intensities of each sample was calculated as the median fluorescence intensity (MFI). Accordingly, HPV types were discerned as per the unique bead signature, whereas the presence of PCR products was determined by phycoerythrin fluorescence. For each run and each HPV type a background and cut-off value was calculated based on the signal of the negative control as follows:

Background = 1.25 x negative control value (MFI)

Cut-off value = Background + 15 MFI

For HPV type 16, 59 and 82, cut-off = Background + 30 MFI

For a given sample, infection with a particular HPV type was considered positive if the MFI was above the background value.

2.2.5.2 Genotyping of HPV by using MY09/11 and SPF1/2 primers

The samples that tested HPV negative by luminex array (n=92) were further screened for HPV by PCR using MY09/11 and SPF1/2 primers. Prior to this, all 92 samples were subjected to PCR using β-actin primers to check the quality of DNA. PCR reactions were carried out in 25 µl reaction volume containing 100 ng of DNA template, 10 µM of each forward and reverse gene specific primers (Table 2.2), 200 µM of dNTP mix, 2U of Taq polymerase, 1.5-2.5 mM MgCl2 and PCR buffer. The cycling profile for β-actin, MY09/11 and SPF1/2 is given in Table 2.3.

Table 2.2: Primers used for HPV detection and genotyping

Name	Sequence (5'-3')	PCR product size (bp)
MY09/11	F-CGTCCMARRGGAWACTGATC	450
	R- GC**M**CAGGG**W**CTATAA**Y**AATGG	
SPF1/2	F-GCiCAGGGiCACAATAATGG	65
	F-GCiCAGGGiCATAACAATGG	
	F-GCiCAGGGiCATAATAATGG	
	F-GCiCAAGGiCATAATAATGG	
	R-GTiGTATCiACAACAGTAACAAA	
	R-GTiGTATCiACTACAGTAACAAA	
HPV16	F-AAGGCCAACTAAATGTCAC	217
	R-CTGCTTTTATACTAACCGG	
HPV18	F-ACCTTAATGAAAAACCACGA	100
	R-CGTCGTTTAGAGTCGTTCCTG	
β-actin	F-ACACTGTGCCCATCTACGAGGG	621
	R-AGGGGCCGGACTCGTCATACT	

Table 2.3: PCR programs for MY09/11, SPF1/2, HPV16/18 and β-actin

Primer	Denaturation	Annealing	Elongation	No. of cycles
MY09/11	95^{0}C 30 sec	55 ^{0}C 30 sec	72^{0}C 30 sec	35
SPF1/2	94^{0}C 1 min	55 ^{0}C 1 min	72^{0}C 1 min	40
HPV16	95^{0}C 30 sec	55 ^{0}C 30 sec	72^{0}C 30 sec	35
HPV18	95^{0}C 30 sec	55 ^{0}C 30 sec	72^{0}C 30 sec	35
β-actin	94^{0}C 1 min	55^{0}C 1 min	72^{0}C 1 min 30 sec	30

For all PCR reactions initial denaturation was done at 95°C for10 min and final elongation was done at 72^{0}C for 5 min. About 7 µl of the reaction product was analyzed on 1.2-3% agarose gel electrophoresis as mentioned earlier. The samples which tested positive for HPV either by MY09/11 or SPF1/2 or both were further genotyped for the two most common HR-HPV types- HPV16 and 18 using HPV16/18 specific primers.

2.2.6 Statistical analysis

The genotyping data for the two HR-HPV types, HPV16, HPV18 and HPV16/18 together, where adequate follow-up data was available was compared with the clinical outcome of the patients. Kaplan-Meier analysis (SPSS 15.0) was done to determine association between infection with these HR-HPV types and recurrence of disease. Disease free survival was considered from start of radiation therapy to the time when recurrence occurred or till last follow-up. Statistical significance was evaluated using the log-rank test (SPSS 15.0).

2.3 RESULTS

2.3.1 Genotyping of HPV by high throughput Luminex assay

Detection and genotyping of HPV was carried out in 270 locally advanced cervical cancer biopsies using Multiplex HPV Genotyping Kit that allows simultaneous detection and genotyping of 24 HPV types - 15 high-risk types (16, 18, 31, 33, 35, 39,

45, 51,52, 56, 58, 59, 68, 73, and 82), 3 putative high-risk types (26, 53, and 66), and 6 low-risk types (6, 11, 42, 43, 44, and 70). The kit combines PCR with biotinylated, consensus primer set with Luminex xMAP technology for this purpose. Using the $GP5^{+}/GP6^{+}$primer set for PCR, 178 out of 270 samples were found to be positive for HPV (Fig. 2.1).

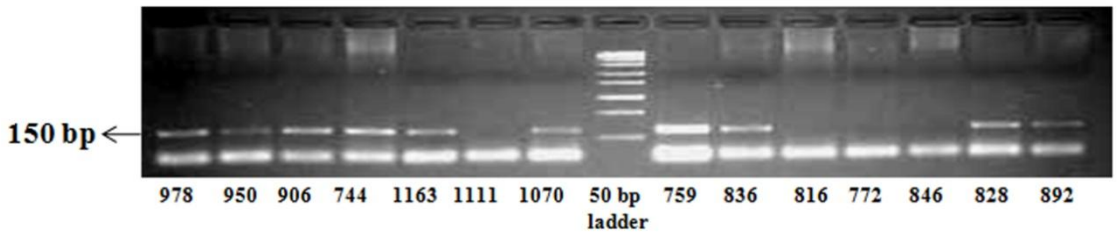

Fig. 2.1: HPV positivity in cervical cancer biopsies by $GP5^{+}/6^{+}$ primers set*. DNA from cervical cancer biopsies were PCR amplified using $GP5^{+}/6^{+}$ consensus primers and resolved on 2% agarose gel. Presence of 150 bp band confirms HPV positivity.The the sample names begin with 'CT' followed by the lab code. M= 50 bp ladder.*

The HPV positive samples were further subjected to hybridization to bead-bound probes by luminex array as described earlier. Following were the observations:

- 169 samples were found to hybridize to the different HPV probes, whereas 9 samples were negative. These 9 samples could have HPV infection not included in the 24 types detected by the kit (Fig. 2.2).
- 168/169 samples were positive for one or more HR-HPV types indicating a high association of cervical cancer with HR-HPV infection.
- Of the 24 HPV genotypes included in the HPV Genotyping Kit, the following 14 HPV types were present in this cohort of 270 patients: HPV 11, 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 73 (Fig. 2.2).
- HPV16 and/or HPV18 infection were most common - 114 samples being positive for HPV16 alone and 6 samples for HPV18 alone (Fig. 2.2).

- Dual infection by HPV16 and HPV18 (n=16) and HPV16 and HPV45 were also common (n=9) (Fig. 2.2).
- Triple infection by HPV 16, 39 and 59 was observed in one sample.
- HPV type 31, 33 and 35, high-risk viruses were detected at a very low frequency.

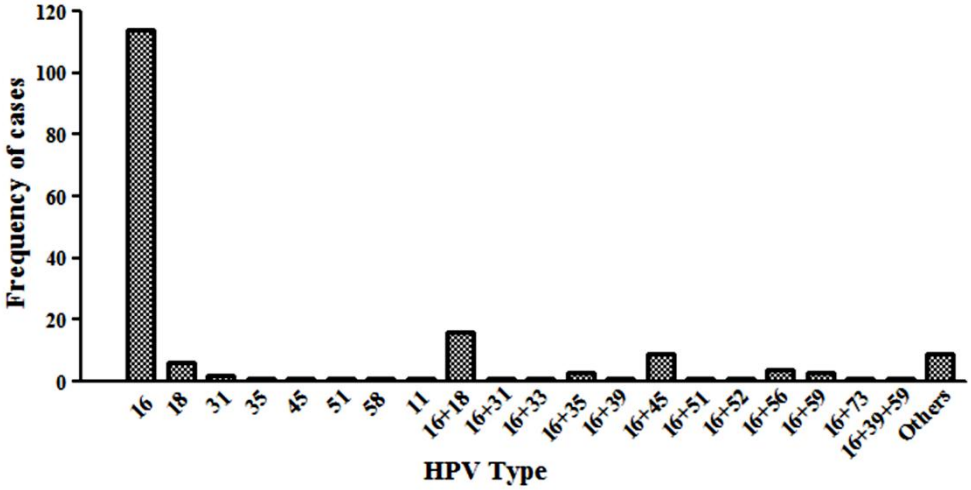

Fig. 2.2: Frequency of 24 HPV types as detected by Luminex array. *The graph depicts frequency of 24 HPV types in 178 cervical cancer biopsy samples which were found to be positive $GP5^{+}/6^{+}$ primers. HPV16 infection predominated in the samples. Each bar represents different HPV types.*

2.3.2 Genotyping of HPV by using MY09/11 and SPF1/2 primers:

In order to estimate the true HPV positivity in the 270 cases, the 92 cervical cancer biopsies negative for HPV by luminex array, were first subjected to PCR using β-actin primers to check the quality of DNA. All were found to be positive for β-actin. Next they were subjected to PCR using MY09/11 and SPF1/2 primers. These primer sets amplify a 450 bp and 65 bp fragments in the L1 region of the viral genome respectively. Twenty five out of 92 samples were found to be positive for HPV by MY09/11 PCR (Fig. 2.3a). Since the amount of DNA was limiting, SPF 1/2 PCR could not be carried out in 4 of the 92 samples. Screening with SPF1/2 primers revealed 79 samples to be HPV positive (Fig. 2.3b). These 79 samples also included the 25 samples that tested HPV positive by MY09/11 PCR.

The HPV positivity was therefore calculated taking into account the results from luminex array, MY09/11 and SPF1/2 PCR. The overall HPV positivity in this cohort was found to be 95% (257/270). Since the prevalence of HPV type 16 and 18 were highest from the luminex data, the 79 samples that tested positive by MY09/11 and SPF1/2 primer sets, were further genotyped using HPV 16/18 specific primers. Forty nine samples were found to be positive for HPV16 (Fig. 2.3c). None of these 79 samples were positive for HPV18. Therefore, the overall prevalence of HPV16 and/or HPV18 in this cohort was 72% (185/257).

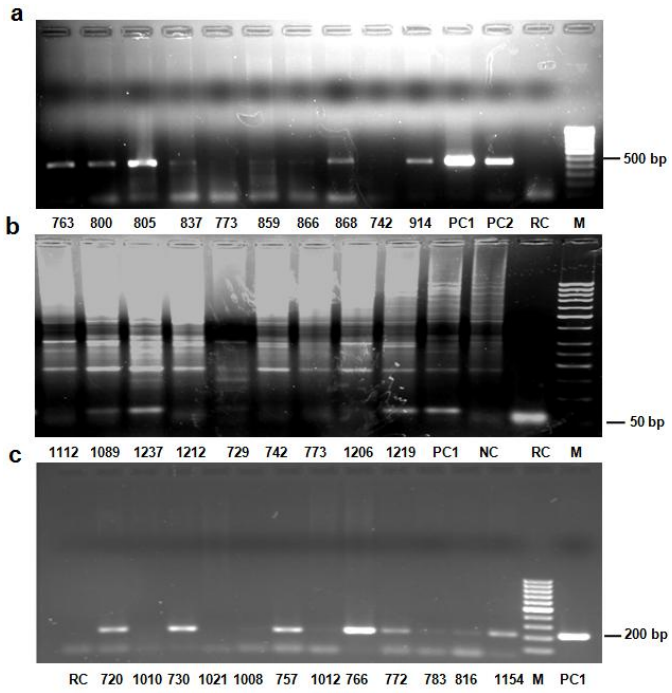

Fig. 2.3: HPV positivity in cervical cancer biopsies by MY09/11 and SPF1/2 primer sets*. 92 samples that were found to be HPV negative by GP5^{+}/6^{+} primers were tested for DNA integrity by β-actin PCR. Samples that were positive for β-actin were further amplified by MY09/11 and resolved on a 1.5% agarose gel. A 450 bp band confirms presence of HPV (a). Also, the negative samples were checked by SPF1/2 primers and resolved on a 3% agarose gel wherein a 65 bp amplicon shows HPV positivity (b). HPV positive samples were checked for HPV16 and HPV18 infection by type- specific primers. 217 bp and 100 bp amplicon confirms presence of HPV16 and HPV18 respectively. 49 samples were positive for HPV16 (c). The sample names begin with 'CT'. PC= positive control, PC1= DNA from Hela cells, PC2= DNA from Caski cells, NC= negative control, DNA from C33A cells, RC=reagent control, M= molecular weight marker*

2.3.3 Correlation with clinical outcome

Of all the cases positive for the two HR-HPV types - HPV16, HPV18 and HPV16/18 (n=185), adequate follow-up data was available for 125 cases (median follow-up for 125 cases was 54 months). The detail clinicopathological data of these 125 cases is given in Appendix I. The survival data of these 125 cases, in terms of recurrence of the disease, was compared with presence of the above- mentioned HPV genotypes. Kaplan-Meier survival analysis data revealed that there was no significant difference between infections with these two HR-HPV types in terms of disease outcome (Fig. 2.4). However, HPV16 alone positive cases showed a trend for better survival as compared to the other two groups.

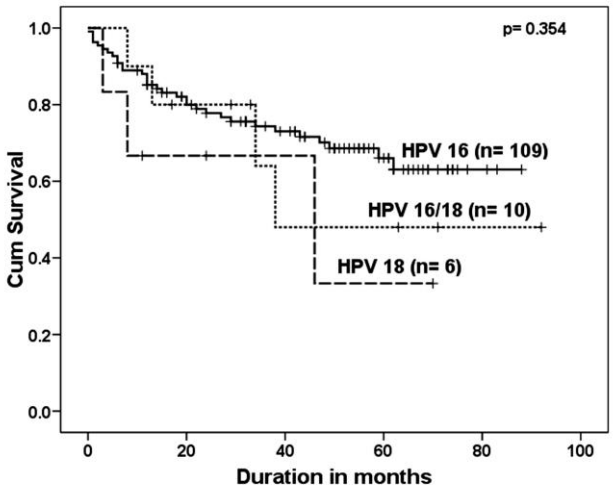

Fig. 2.4: Kaplan-Meier analysis for two HR-HPV types - HPV16 and/or HPV18 and disease outcome. *Kaplan-Meier survival analysis for HPV16 and/or HPV18 in 125 patients who had a good clinical follow up was carried out. Patients with HPV16 infection alone showed a trend towards better disease free survival as compared to HPV18 infection alone and dual infection with HPV16/18.*

2.4 DISCUSSION

The association of HPV with cancer of the cervix has been proven beyond doubt. Reports from different parts of the sub-continent indicate a prevalence of HPV ranging from 73 to as high as 99% (Basu et al., 2011; Basu et al., 2009; Bhatla et al., 2008;

Chatterjee et al., 2003; Deodhar et al., 2012; Franceschi et al., 2003; Grace Nirmala and Narendhirakannan, 2011; Peedicayil et al., 2006; Pillai et al., 2010; Saranath et al., 2002; Singh et al., 2009; Sowjanya et al., 2005; Travasso et al., 2008). The main concern that remains with HPV detection is that since the virus still cannot be cultured efficiently and the available serological assays are not sensitive enough, diagnosis is mainly based on the detection of HPV nucleic acids in a specimen (Poljak and Kocjan, 2010). The most common method for detection as well as HPV genotyping till date is PCR (with general or type specific primers) followed by sequencing. As opposed to type specific primers, general or consensus PCR primers have the ability to detect a broad spectrum of HPV genotypes. These general primers, also known as universal primers are mostly designed in the L1 region of the viral genome (except for CP primers that is based in the E1 region) which is well-conserved across all HPVs, and include MY09/11 (Gravitt et al., 2000; Manos et al., 1989), OBI/II (Jenkins et al., 1991), CPI/CPII (Tieben et al., 1993), $GP5^{+}/6^{+}$(Jacobs et al., 1997) and SPF1/2 (Kleter et al., 1998). Although MY09/11 and $GP5^{+}/6^{+}$ primer sets are widely used for routine diagnosis of HPV infections, there are several issues with their sensitivities and performance (Kleter et al., 1998). SPF primers have been considered to be the most sensitive, mainly due to the small amplicon size. Recent studies have shown that the overall prevalence of HPV can be underestimated considerably if only a single DNA detection method is used (Baay et al., 1996; Karlsen et al., 1996; Smits, 1996; Smits et al., 1995). Therefore, in order to estimate the true HPV positivity in our cohort of 270 advanced cervical cancer biopsies, detection of HPV was carried out using 3 primer sets - $GP5^{+}/6^{+}$, MY09/11 and SPF1/2. Using these 3 primer sets, we observed 95% HPV positivity and it is apparent from this study that a single set of primers is not sufficient to estimate the true HPV infectivity. However, we could genotype 24 HPV types only in the samples those were positive by

$GP5^+/6^+$ primers (n=178) since the Multiplex HPV Genotyping Kit relies on amplification by $GP5^+/6^+$ primer set. For the remaining 92 samples, genotyping for the 2 most common HR-HPV types - HPV16 and HPV18 was done (Fig. 2.5). Overall observation was that infection with HPV16 was most common (60%) followed by infection with HPV18 alone (2%). Dual infection with HPV16/18 (6%) and HPV16/45 (3%) was also seen.

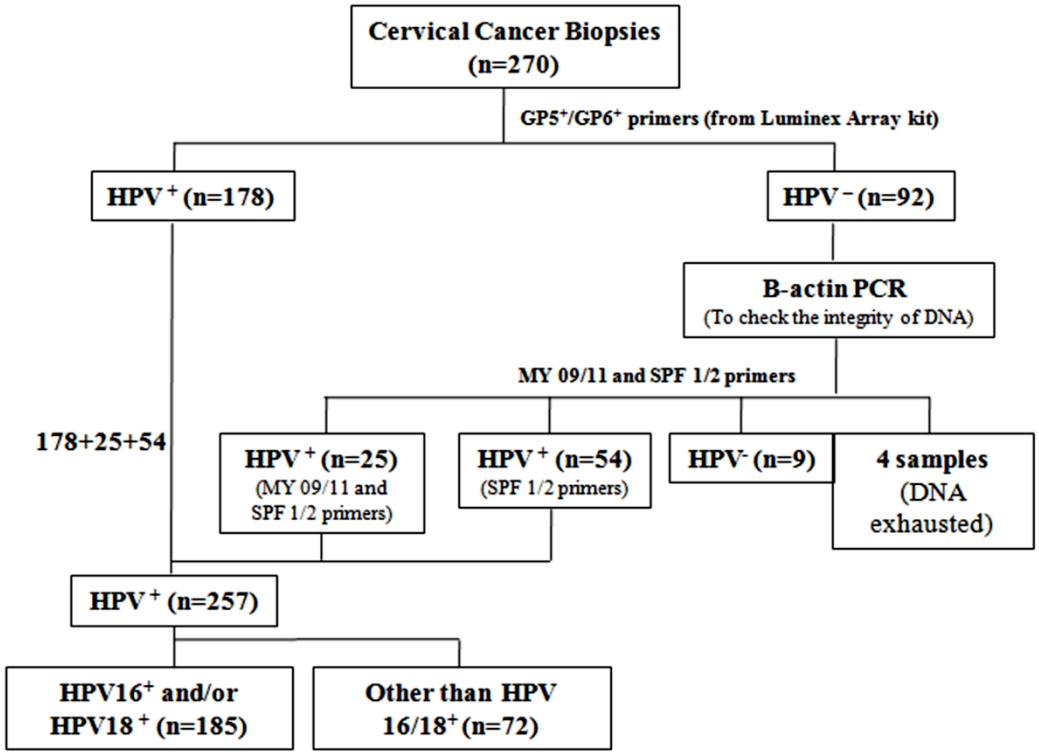

Fig. 2.5: Flowchart depicting summary of the genotyping study. *Genotyping was carried out on 270 advanced stage cervical cancer samples by high-throughput, $GP5^+/6^+$ primers based luminex array; consensus MY09/11 and SPF1/2 primers. HPV positivity was 95% (257/270) with infection by HPV16 and 18 being most common.*

Major studies, particularly from the Indian subcontinent indicate 57-65% HPV16 positivity followed by HPV18, 45, and 33, in cervical neoplasias. One of these studies by Bhatla *et al.* comprising of meta-analysis of 6 Indian studies provides a clear picture of the prevalence of a broad spectrum of HPV types apart from 16 and 18 in invasive cervical cancers (Bhatla et al., 2008). Five of these studies were based in South India

with one of them including some patients from the eastern region; the sixth was conducted in North India (Bhatla et al., 2006; Franceschi et al., 2003; Munirajan et al., 1998; Peedicayil et al., 2006; Sathish et al., 2004; Sowjanya et al., 2005). A more recent study by Basu *et al*., that determined HPV genotypes in cervical cancer cases across four cities representing different regions of the Indian subcontinent, also reports similar distribution of HPV genotypes, with HPV16, 18 infection followed by infection with HPV33 (Basu et al., 2009). The distribution of HPV genotypes in our study was in concordance with these reports except for the fact that infection by HPV33 was<1%, which might be an underrepresentation, owing to technical limitations.

Although, reports have confirmed the prognostic value of HPV genotypes in cervical cancer treated with radiotherapy (Wang et al., 2010), in our study HPV16 positive cases showed a trend towards better disease free survival after radiation therapy, as compared to HPV18 and dual infection by HPV16 and 18, but this was not statistically significant. This disparity may be attributed to a difference in the size of study population between the two studies.

2.5 CONCLUSION

The association of HPV with carcinoma of the cervix in this cohort of patients was found to be as high as 95%, with HPV type 16 (66%), 18 (2%) and dual infection by 16+18 (6%) being most common. Other HR-HPV types infecting the cohort included 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 73. Association with clinical outcome showed that HPV16 positive cases had better prognosis in terms of response to radiation therapy as compared to the HPV18 positive ones.

Chapter– 3

Identification of Integration site of the virus by Amplification of Papillomavirus Oncogene Transcripts (APOT) assay

3.1 INTRODUCTION

Integration of HPV DNA into the host genome represents an important event in cervical carcinogenesis (Durst et al., 1985; Wentzensen et al., 2004). Viral genome is present in cervical cancer cells in either episomal or integrated or both forms. Integration event generally occurs downstream of the early genes E6 and E7, often in the E1 or E2 region (Fig. 3.1) resulting in disruption of the E2 gene.

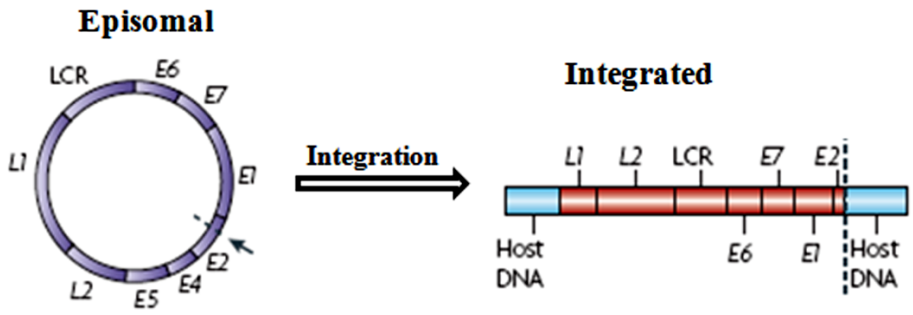

Fig. 3.1: Integration of the HPV genome into the host chromosomes. *Integration event occurs in the E2 region with associated loss or disruption of the same.*

In the absence of E2, transcriptional control on the viral oncogenes, E6 and E7 is lost thereby leading to immortalization and transformation of cells (Romanczuk and Howley, 1992). However, in many precancers and cancers, episomal viral genomes are also present along with the integrated form, thereby providing E2 in *trans* to regulate E6 and E7 expression (Arias-Pulido et al., 2006). In most cases, integration gives rise to fusion transcripts comprising of viral oncogenes E6, E7 and adjacent cellular sequences (Choo et al., 1987; Couturier et al., 1991; Di Luca et al., 1986; Klaes et al., 1999). These host genomic sequences might take over the control machinery of the integrated viral genomes using its own promoter, thereby resulting in higher levels of E6 and E7 expression even in presence of E2 (von Knebel Doeberitz et al., 1991).

Studies carried out in vitro have demonstrated that the viral-cellular fusion transcripts are more stable and impart the cells with a selective growth advantage as

compared to the episomal counterparts (Jeon et al., 1995; Jeon and Lambert, 1995). Also, according to available studies, the event of viral integration is random involving almost all the chromosomes, and till date more than 200 virus-host integration sites have been mapped into the human genome (Wentzensen et al., 2004; Yu et al., 2005). However, there are certain hotspots e.g., fragile sites, translocation break points and transcriptionally active regions which are preferred by the virus for its integration (Koopman et al., 1999; Thorland et al., 2003; Wentzensen et al., 2004; Ziegert et al., 2003). On integration within or near a gene, the virus can bring about a change in its expression. This phenomenon of insertional mutagenesis might activate cellular oncogenes or inactivate cellular tumour suppressive genes, eventually leading to alterations in cellular growth and proliferation. Besides, viral integration can render both viral coding genes as well as the cellular genes susceptible to epigenetic changes which could regulate their expression.

Another important factor that may play considerable role in carcinogenesis apart from HPV genotype and viral integration is the viral load. The relationship between viral load and the disease is quiet complex. While most of the evidence indicates that a high viral load is associated with persistent infection and progression to high-grade lesions (Dalstein et al., 2003; Gravitt et al., 2007; Hernandez-Hernandez et al., 2003), there are others that report either no association, or a higher viral load in women with low-grade squamous intraepithelial lesion (LSIL) than in those with high-grade squamous intraepithelial lesion (de Boer et al., 2007; Kim et al., 2009). Moreover, while considering the relationship between viral load and disease prognosis, the physical status of the virus also becomes important. This information, when combined together could provide a complete picture about the natural history of HPV infections and their relationship to disease.

3.2 MATERIALS and METHODS

MATERIALS

Chemical/ Reagent	Company
InsTA PCR Cloning Kit	*Fermentas, Lithuania, EU*
KOH, KCl, NaOH, Glacial acetic acid	*SD Fine-Chem. Ltd., Mumbai, India*
β- Mercaptoethanol, Formamide , DEPC, Ampicillin, X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase), IPTG (Isopropyl-β-D-Thiogalactopyranoside), PIPES, $CaCl_2$, DMSO	*Sigma, St. Louis, MO, USA*
RNAeasy Mini kit	*Qiagen, Hilden, Germany*
Formaldehyde, N-N-Dimethyl formamide	*MERCK*, Mumbai, India
MOPS	*USB* corp., Cleveland, OH, USA
Tryptone, Yeast Extract	*HiMedia Laboratories Ltd., Mumbai, India*
MgSO4, Glucose, Potassium Acetate	*Qualigens fine chemicals, Mumbai, India*
SuperscriptTM first strand synthesis for RT-PCR	*Invitrogen, Carlsbad, CA, USA*
DNA free kit	*Ambion, Austin, TX, US*
2X Power SYBR Green, 96 well plate, Optical adhesive cover	*ABI, Foster City, CA, USA*
GFX PCR DNA and Gel Band Purification Kit	*GE Healthcare, Buckinghamshire, UK*
$MnCl_2$	*Sisco Research Laboratories Pvt., Ltd., Mumbai, India*
$MgCl_2$	*Thomas Baker Chemical Ltd., Mumbai, India*
Sterile disposable 90mm petri plates	*Axygen ,* Union city CA, USA
0.2 µ membrane filter	*Millipore, USA*

Clinical Sample

The study was performed on 86 pretreatment cervicaltumour biopsies, predominantly from FIGO stage IIB and IIIB with HPV16 and/or HPV18 infection and a good clinical follow-up (Appendix-I).

METHODS

3.2.1 RNA extraction from cervical cancer biopsies

RNA was isolated from the biopsies using RNAeasy Mini kit, following manufacturer's instructions. RNA was finally eluted in 30 µl nuclease free water and its quantity and

quality assessed by measuring concentration and OD260/OD280; OD260/OD230 ratio on NanoDrop UV-Vis spectrophotometer. The integrity of RNA was ascertained by reducing agarose gel electrophoresis.

Note: Buffer RPE is supplied as a concentrate, so before using 4 volumes of ethanol (96 -100%) was added to it to obtain a working solution.

3.2.2 Formaldehyde agarose gel electrophoresis of RNA

Formaldehyde agarose gel was used for checking the integrity of RNA. Unlike DNA, RNA has a high degree of secondary structure, making it necessary to use a denaturing gel. Formaldehyde in the gel disrupts secondary RNA structure enabling the RNA molecules to be separated by charge migration.

A 1.2% reducing agarose gel was prepared by melting 0.36 gm agarose in 25.5 ml DEPC water. To this 3 ml, 10x MOPS, 1.5 ml formaldehyde was added and the gel was allowed to set at RT. 1 µl of RNA was mixed with RNA loading buffer containing 10X MOPS, Formaldehyde, Formamide and DEPC water and heated at 65°C for 7 min. This was kept in ice for 5 min and after adding 2 µl of loading dye and 0.2 µl EtBr, loaded on the gel and run at a constant voltage of 50V for 1.5 h. RNA bands were visualized and captured on a Gel documentation system.

Note: 10X MOPs was prepared by mixing 0.2M MOPS pH 7, 50mM Sodium acetate and 5mM EDTA pH 8.0. The solution was prepared in DEPC water, filtered through 0.45µ Millipore filter and stored in brown coloured bottle.

3.2.3 DNase treatment of RNA

In order to get rid of any DNA contamination which is the pre-requisite for RT-PCR, the isolated RNA samples were treated with DNase using DNA-free kit. Total RNA with ~200 ng/µl concentration was taken in thin walled 0.6 ml microfuge tube. Depending upon the amount of RNA being treated with the enzyme, per 10 µl reaction, 1 µl 10x DNase Buffer and 1U γ DNase I enzyme was added and the reaction was incubated at

37°C for 30 min. To stop the enzyme activity 1 µl DNase inactivator slurry was added and the reaction was incubated for 2 min at RT with intermittent mixing so that inactivator can interact with total enzyme in the reaction mixture. The contents were centrifuged at 10,000 rpm/ 1.5 min /RT and RNA was transferred to fresh tube.

3.2.4 Amplification of Papillomavirus Oncogene Transcripts (APOT) assay

Site of viral integration was identified in a subset of cervical tumour samples (n=86) positive for HPV16, HPV18 or both and with a follow up data of minimum 1 years, using APOT assay, a RT-PCR based assay that allows the discrimination of HPV mRNAs derived from integrated and episomal viral genomes. In this method, a modified 3'-RACE PCR using upstream HPV E7-specific and downstream oligo dT adaptor primers are applied to amplify HPV E7-specific transcripts derived from both integrated and episomal viral genome. The integrated form is then distinguished from the episomal counterpart on the basis of difference in size between the two by agarose gel electrophoresis. The primers used in the APOT are given in Table 3.1.

Table 3.1: List of primers used in APOT assay

Name	Sequence (5'-3')
p1-HPV16	F- CGGACAGAGCCCATTACAAT
p1-HPV18	F- TAGAAAGCTCAGCAGACGACC
p3	F- GACTCGAGTCGACATCG
p2-HPV16	F- CCTTTTGTTGCAAGTGTGACTCTACG
p1-HPV18	F- ACGACCTTCGAGCATTCCAGCAG
Frohman primer: (dT).17-p3	F- GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT

3.2.4.1 Preparation of cDNA from total RNA: The first step of APOT assay was synthesis of cDNA from the RNA using Superscript II Reverse Transcriptase enzyme and the Frohman primer [(dT)17-p3]. Following steps were followed: 0.5-1 µg of DNA free RNA was taken in 0.6 ml microfuge tube followed by addition of 2.5 µl of 10 mM

Frohman primer [(dT)17-p3] and 1 µl 10 mM dNTP mix. The total volume was made 10 µl with nuclease free water and the tubes were incubated at 65°C for 5 min followed by snap cooling on ice for 5 min. Next, a reaction mixture containing 10X buffer (2µl), 25mM $MgCl_2$ (4µl), 0.1M DTT (2µl), 40U/ µl RNase out (0.5 µl) and 200 U/ µl SS II (0.5 µl) was added to each sample. The contents were mixed gently and were incubated for 10 min at RT followed by 42°C for 1 h. The reaction was terminated by incubating at 85°C for 5 min. Finally, 1µl RNaseH (2U/ µl) was added to each tube and incubated at 37°C for 20 min, for specific degradation of the RNA in RNA:cDNA hybrids. This cDNA was first subjected to β - actin PCR, to check the integrity of the cDNA and then taken for nested PCR with E7 specific primers.

3.2.4.2 PCR using oligo dT / Adaptor primers and HPV E7 specific primers: The first strand cDNA was subjected to two sets of PCR. PCR reactions were carried out in 50 µl reaction volumes containing about 200 ng of cDNA template. In the first reaction, amplification was done using HPV E7–specific primer (p1-16 specific for HPV16 and p1-18 specific for HPV18) as forward primers and linker p3 as the reverse primer. Next, 7 µl of the PCR product was used as template for the second PCR using forward primers p2–16 specific for HPV16 or p2–18 specific for HPV18 and (dT)17-p3 as reverse primer. All the amplifications were carried out in a reaction volume of 50 µL containing 2.5 mM MgCl2, 200 mM dNTPs, 25 mM of each primer and 1 unit of Taq polymerase.

Table 3.2: Thermocycler program for APOT analysis

Step	1st PCR: Temperature (ºC)/ Duration (min: sec)	2nd PCR: Temperature (ºC)/ Duration (min: sec)
Initial denaturation	94/ 2.00	94/ 2.00
Denaturation	94/ 0.30	94/ 0.30
Annealing	58/0.30	66/0.30
Elongation	72/ 4.00	72/ 4.00
35 cycles step 2-4		
Final elongation	72/ 20.00	72/ 20.00

3.2.4.3 DNA fragment isolation from agarose gel: The final PCR products were electrophoresed on a 1.2% agarose gel, as described earlier pg. 67. Viral-cellular fusion transcripts (Offsize bands) would deviate from the characteristic size of the major viral transcript derived from episomal viral DNA (~1050 bp for HPV16 and ~1000 bp for HPV18) and be differentiated. These were excised from the gel and DNA isolated using GFX PCR DNA and Gel Band Purification Kit. For this, the gel slice containing the required band (seen under UV) was cut from the gel using a clean scalpel and placed into an autoclaved pre-weighed 1.5 ml microcentrifuge tube. The weight of the gel was estimated and for each 10 mg of gel slice, 10 µl capture buffer (contained in the kit) was added, followed by through mixing and incubation at 60ºC for 30 min or until the agarose was completely dissolved. This mixture was transferred onto an assembled GFX Microspin column and collection tube, incubated at RT for 1 min. and spun at 13,000 rpm /30 s/ RT. The flow through was discarded and the column was washed twice with 500 µl Wash buffer type 1. Finally, a blank spin was given to get rid of any 'carryover' wash buffer and the DNA eluted from the column with 30 µl Elution buffer.

3.2.4.4 Preparation of ultra-competent E.coli cells: For better cloning efficiency of DNA fragments, *E.coli* strain DH5α cells was made ultra-competent. Host cells were streaked on a freshly made LB agar plate from the glycerol stock and incubated at 37ºC/ON. Next day, a single colony was inoculated in 5 ml SOB broth (2% Bactotryptone, 0.5%Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM MgSO4), and incubated overnight on shaker incubator at 37ºC. This served as the starter culture from which 2.5 ml was inoculated in 250 ml SOB in a 1 L flask and incubated at 18ºC while shaking (at 250 rpm) till OD_{600} reached ~0.6. Culture was immediately transferred to pre chilled centrifuge bottles and centrifuged at 4000 rpm/10 min/ 4ºC in a swing out rotor; cell pellet was resuspended in 80 ml transformation buffer and kept on

ice for 10 min followed by centrifugation at 4000 rpm/10 min/ 4ºC. Transformation buffer was made by adding 10 mM PIPES, 15 mM $CaCl_2$, 250 mM KCl, to 100 ml D/W, followed by pH adjustment to 6.7 with 5N KOH and addition of 55 mM $MnCl_2$. The solution was finally filter sterilized through 0.2 µ membrane filter. Following this, the cell pellet was again resuspended in 20 ml transformation buffer to get a homogeneous suspension, incubated on ice for 10 min and DMSO was added drop-wise to final concentration of 7%. Suspension was mixed, made to aliquots, and 200 µl aliquots were snap frozen in liquid nitrogen and stored at -80^0C.

3.2.4.5 Cloning of viral cellular fusion transcripts: Direct cloning of a PCR amplified fragment or 'TA cloning' takes the advantage of the terminal transferase activity of thermostable polymerases lacking proofreading activities, e.g. Taq DNA polymerase, adding a single 3'-A overhang to both the ends of the PCR products. PCR products generated using these polymerases favour direct cloning into a linearized vector with single 3'-dT overhangs. The ligation mix contained DNA (300-600 ng depending on the size of the fragment of interest), 5X buffer (4 µl), T4 DNA ligase (6U), pTZ57R/T vector (50 ng) and D/W to make the volume up to 20 µl. Reaction mixture was incubated at 22ºC ON and then used for transformation in E. coli DH5α cells.

3.2.4.6 Transformation of recombinants: The process of introducing foreign DNA into bacterial cells is referred to as transformation. The cells are made susceptible (competent) to uptake DNA molecules by treatment with a solution of $CaCl_2$ and then briefly warmed to generate pores in the bacterial cell wall for very short period of time, triggering the uptake of surrounding DNA molecules in solution.

200 µl of ultra-competent cells were thawed on ice and ligation mix from the above step was gently added, mixed by gentle pipetting and incubated on ice for 30 min. Heat shock treatment was given for 55 s in a 42°C water bath and the transformation

mixture was immediately snap chilled by transferring on the ice for 5 min. 400 µl of SOC broth was gently added to the transformation mixture under the sterile conditions followed by incubation at 37°C for 1 hr in a shaker incubator at 150 rpm. Finally, the transformation mixture was spread on a LB agar plate containing ampicillin and pre-treated with X- gal/ IPTG followed by ON incubation at 37°C. Positive colonies were selected based on blue-white selection. The principle behind X- gal/IPTG selection is that vectors with an intact MCS within the β - galactosidase gene (*lacZ*) results in accumulation of blue colour in presence of substrate X-gal, while disruption in the MCS by incorporation of DNA fragment results in the formation of white colonies. Accordingly, positive colonies were screened by selecting the white colonies.

3.2.4.7 Plasmid mini preparation by alkaline lysis: Bacterial colonies were inoculated in 2 ml LB medium containing ampicillin and cells were recovered next day by centrifugation at 14000 rpm/ 2 min/ RT. The cell pellet was suspended in 100 µl solution I (50 mM Glucose, 25 mM Tris.Cl pH 8.0, 10 mM EDTA in D/W), followed by addition of 200 µl lysis solution (solution II: 0.2 N NaOH and 1% SDS in D/W). Next, 150 µl chilled neutralization solution (solution III: 5M Potassium Acetate and Glacial acetic acid in D/W) was added, mixed by inversion and incubated on ice/10 min. The mixture was centrifuged at 14000 rpm/ 10 min/ RT; the clear solution was transferred into a fresh microfuge tube and the plasmid was precipitated using 1 ml of absolute chilled ethanol at -20ºC/20 min and centrifugation at 14000 rpm/ 20 min/ RT. Finally, the plasmid pellet was washed with 0.5 ml 70% ethanol, semi air-dried at RT and dissolved in 20 µl RNAse TE buffer.

3.2.4.8 Screening of recombinants by PCR: In order to confirm presence of the insert, PCR was performed with the plasmid DNA using p2 and (dT).17-p3 primers as described in section 2.4.3.

3.2.4.9 Sequencing of viral cellular fusion transcripts: The positive clones or in some cases, the PCR product itself were sequenced on DNA Sequencer (3100 Avant Genetic analyzer, Applied Biosystems, Foster City, CA, USA). The chromosomal integration loci were determined using National Centre for Biotechnology Information (BLAST) and the University of California, Santa Cruz (UCSC) hg19 (Feb. 2009) (BLAT) human genome assemblies. Further, the integration sites were checked for the presence of fragile sites and any genes of known identity using NCBI fragile site map viewer and the UCSC Blat tool respectively.

3.2.5 Validation of the recurrent integration sites

Some of the recurrent integration sites were checked at the genomic level by genomic DNA PCR with HPV E7 primers as the forward primer and primers specific to a given chromosomal region as the reverse one. The PCR reaction was carried out as described earlier on pg. 85-86.

3.2.6 Estimation of viral load by qRT-PCR

Quantitative real time polymerase chain reaction (qRT-PCR), a technique based on PCR is used to amplify and simultaneously quantify a targeted DNA/RNA molecule. The quantity can be either an absolute number of copies or a relative amount when normalized to the input or additional normalizing genes.

Copy number of two high-risk HPV types HPV16 and HPV18 was assessed in all the 86 samples where the physical status of the virus was studied by SYBR green based qRT-PCR. The primers used for the study are given in Table 3.3.

Table 3.3: List of primers used for HPV copy number analysis

Name	Sequence (5'-3')	PCR product size (bp)
HPV16 E7	F- TCCAGCTGGACAAGCAGAAC	100
	R- CACAACCGAAGCGTAGAGTC	
HPV18 E7	F- AACATTTACCAGCCCGACGA	100
	R- TCGTCTGCTGAGCTTTCTAC	
TBP	F- GGACTGACCCCACAGCCTATTCAG	80
	R- TGTTGTTGCTGCTGCTGCCTTTG	

Copy number was determined by amplification of a 100 bp sequence located in the E7 gene of HPV16 or HPV18. A DNA fragment in the *TBP* gene, a single copy gene that maps to 6q27 was used as a reference. A reaction mixture containing 20 ng of genomic DNA, 2.5 μM of each forward and reverse primers and 5μl of 2xPower SYBR Green was made in 10μl volume in a 384 well plate. This was sealed tightly with the optical adhesive covers and the reaction was carried out in the ABI Prism 7700 Sequence Detection System *(ABI*, Foster City, CA, USA). All the PCR reactions were done in duplicate and a negative control with no template was also set. Thermocycler conditions for the reaction are given in Table 3.4.

Table 3.4: Thermocycler program for copy number analysis

Step	Temperature (ºC)	Duration (min:sec)
Initial denaturation	95	15:00
Denaturation	95	0:15
Annealing	60	0:30
Elongation	72	1:00
45 cycle step 2-4		
Final elongation	71	4:00

The read out was obtained in the form of C_t values which is defined as the number of cycles at which the fluorescence exceeds the threshold. The average C_t value was calculated for each of the samples for both E7 and TBP gene, and the ΔC_t value was

obtained for each by substracting average C_t value of the reference, TBP from that of E7. HPV copy number was finally estimated by the formula $\mathbf{2^{-\Delta Ct}*2}$ as described by Peter et al (Peter et al., 2006)**.**

3.2.7 Association with clinical data

The data obtained was compared with the clinical outcome of the patients. Kaplan-Meier analysis (SPSS 15.0) was done to determine the association of the viral state (episomal/integrated) with recurrence of the disease. Disease free survival was considered from start of radiation therapy to the time when recurrence occurred or till last follow-up (median follow-up for 86 cases was 44 months). Statistical significance was evaluated using the log-rank test (SPSS 15.0). Further, where viral integration was observed at a particular chromosomal locus in ≥4 cases, Kaplan-Meier analysis was performed to check whether the site of integration in the host genome had any bearing on the disease prognosis. The copy number of the two HR-HPV types- HPV16 and HPV18 was compared with the clinical follow-up of all the 86 patients in terms of recurrence of the disease as well as the physical status of the virus.

3.3 RESULTS

3.3.1 Determination of physical state of the virus by APOT assay

APOT assay was performed on 86 samples positive for HPV16 and/or HPV18 and with adequate clinical follow up as well as good quality RNA. The assay uses an oligo(dT)17 primer coupled to a linker sequence [(dT)17-p3] for the reverse transcription followed by nested PCR, using E7 specific forward primers in combination with primers p3 and (dT)17-p3, respectively, This results in amplification of both episome- and integrate-derived HPV oncogene transcripts . Integrate-derived transcripts were differentiated

from episome-derived transcripts (1050 bp for HPV16 and 1000 bp for HPV18) by gel electrophoresis (Fig. 3.2). Integration was further confirmed by cloning and sequencing of the PCR fragments.

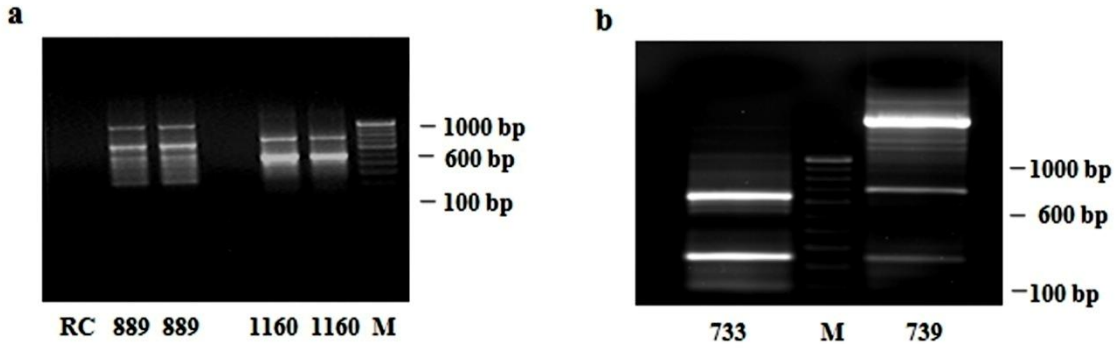

Fig. 3.2: APOT from the cervical cancer biopsies*. RT-PCR products were separated on 1.2% agarose gels. The amplicons of about 1000-bp length represent the episomal transcript, while that with a different size might be derived from the integrated form. The 250-bp product, observed in almost all the samples resulted from misannealing of the (dT)17-p3 primer to an adenosine-rich sequence stretch within the HPV16 E1 gene during RT, as confirmed by sequencing.*

3.3.2 Association of physical state of the virus with clinical outcome

APOT assay for all the 86 cases revealed the viral genome to be integrated in 79% cases (n =68), whereas in 21% (n =18) only episomal transcripts could be identified. In 12 cases both integrated as well as episomal form of HPV were detected. The physical state of the virus (episomal/ integrated) was associated with the individual patient's disease outcome. Survival data with a median follow up of 44 months, revealed that 16 out of 18 patients with only episomal form of HPV (16 and/or 18), had disease free survival as compared to those with integrated form of the virus, indicating a good clinical outcome (p= 0.067, representing a borderline significance) (Figure 3.3).

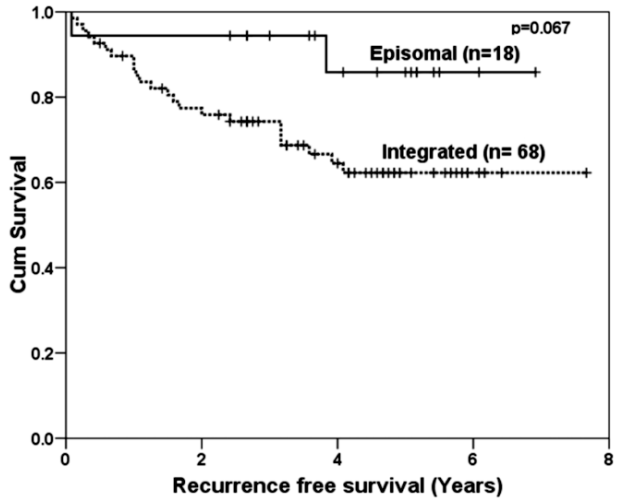

Fig. 3.3: Kaplan-Meier analysis for episomal vs. Integrated viral genome. *Kaplan-Meier survival curve for patients with episomal form of virus (n = 18) vs. integrated form (n = 68) is depicted. Most of the patients with episomal form (16 out of 18) had a disease free survival as compared to patients with integrated form, indicating a good clinical outcome, although with a borderline significance (p = 0.06).*

3.3.3 Identification of viral integration sites in the genome

In order to understand whether the integration event is random or there is some preference for certain sites within the chromosomes, the sequencing data for 68 cases derived from APOT assay were investigated by BLAST and/or BLAT. The site of integration could be predicted with a high score in 48 cases, for the remaining 20 cases the score was low. Here a low score implies that even though in these 20 cases, the viral genome was integrated, the exact site could not be identified with high confidence by the existing browsers (BLAST / BLAT). Although the sites of integration were found to be distributed throughout the genome there were recurrent sites e.g., at the chromosomal loci 1p (n =7), 3q (n= 8), 13q (n= 4), 6q (n =4), 11q (n= 4) and 20q (n= 4) (Fig. 3.4). One sample showed HPV integration at two chromosomal loci simultaneously.

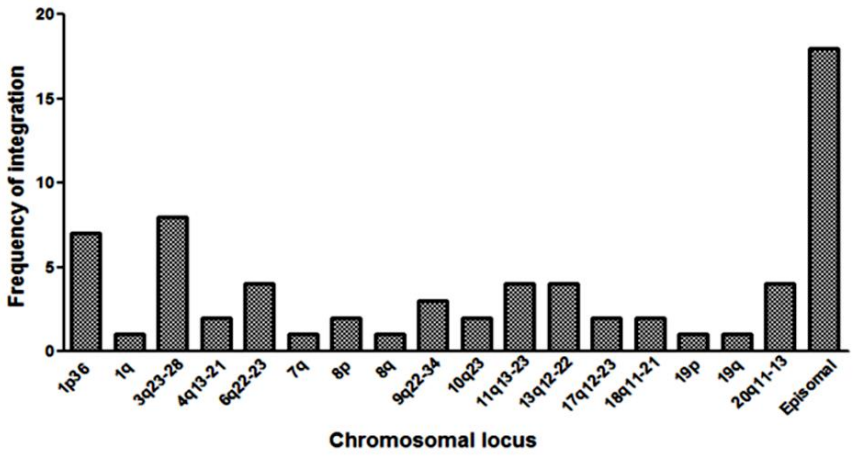

Fig. 3.4: Frequency of HPV integration at different chromosomal loci. *Site of integration as determined by APOT assay in 48 cases positive for HPV16, HPV18 or both and with high prediction score using BLAST/BLAT. Integration event was found to be more common in 1p and 3q chromosomal loci. Each bar represents different chromosomal locus.*

3.3.4 Validation of the recurrent integration sites

For most of the recurrent integration sites including 1p36.23, 3q28, 3q23, 6q22.31, 6q23.3, 8p11.21 11q13.1 and 13q22.1, PCR amplification using genomic DNA was done. Amplification was observed for all but 3 cases, confirming recurrent integrations with APOT. For some cases, multiple amplicons were observed (Fig. 3.5). The failure of amplification with genomic DNA in 3 cases with integration into 1p36.33 locus and presence of multiple amplicons might be attributed to the difference in scenario at the transcript and that at the genomic level. Effects such as splicing occur frequently in the DNA, contributing to complexities. Further, the multiple amplicons might be an indication of multiple viral integrations into the genome. However, only one or at the most two of these sites were transcriptionally active and was identified by APOT. Nonetheless, since information given by APOT analysis was based on sequencing and gave the true picture wherever sequencing coverage was good, we have reported all the sites predicted by APOT with high score.

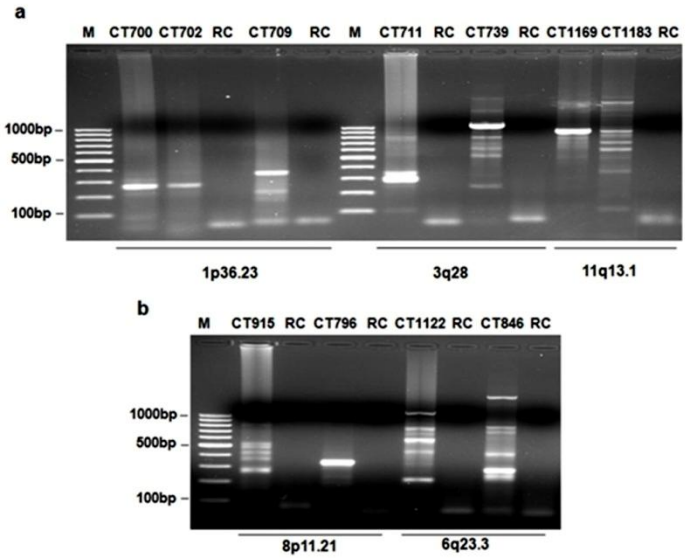

Fig. 3.5: Genomic DNA PCR of the recurrent integration sites. *Representative gel images (a, b) showing HPV integration at the genomic level. Recurrent integrations at chromosomal loci 1p36.23, 3q28, 6q23.3, 8p11.21 and 11q13.1 is depicted.*

3.3.5 Association of site of viral integration with clinical outcome

The site of viral integration could be identified with high confidence in 48 cases by BLAST and/or BLAT. Comparison with clinical information of the patients revealed that those with integration at chromosomal loci 1p (7/7), 6q (4/4) and 11q (4/4) were disease free, while most of the patients with integration at the chromosomal loci 3q (5/8), 13q (4/4) and 20q (2/4) showed recurrence of the disease in the form of either loco-regional or distant metastasis (Fig. 3.6).

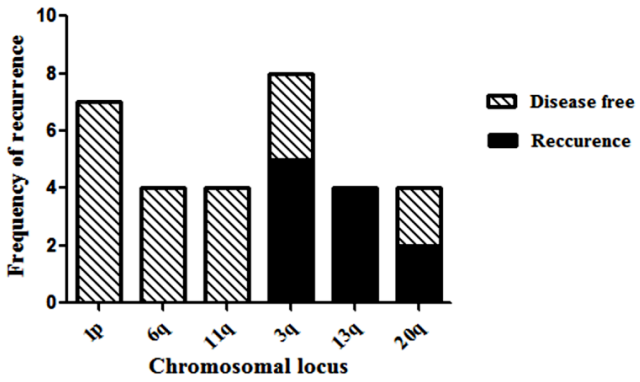

Fig. 3.6: Association of site of viral integration with clinical outcome. *Recurrent viral integration in ≥4 cases is shown. Patients with viral integration at 3q, 8q, 13q and 20q showed poor prognosis in terms of recurrence of the disease, while those with integration at 1p, 6q and 11q loci were disease free.*

3.3.6 Features associated with HPV integration

Since available reports suggest that the virus generally prefers fragile sites and transcriptionally active regions in the human genome for integration (Thorland et al., 2003; Wentzensen et al., 2004; Ziegert et al., 2003), the 48 cases where site of viral integration could be determined with high confidence were analyzed further for such sites using NCBI Fragile site Map Viewer and UCSC BLAT. With NCBI Fragile site Map Viewer it was observed that 60% of integrations (29/48) were located in or close to a common or rare fragile site (Table 3.5). Using the UCSC BLAT tool 58% of the sequences (28/48) were observed to be either within or nearby protein coding genes. These genes belonged to various categories ranging from oncogenes, transcription factors and tumour suppressor genes (Table 3.5).

Table 3.5: Summary of HPV integration

Sample ID	Histology	HPV type	Integration locus	Gene	Fragile site involved
CT 712	SCC	16	1p36.12	ALPQTL2	FRA1A(1p36) com
CT 700	SCC	16	1p36.23	-	FRA1A(1p36) com
CT 702	SCC	16	1p36.23	-	FRA1A(1p36) com
CT 709	SCC	16	1p36.23	-	FRA1A(1p36) com
CT 1138	SCC	16	1p36.33	-	FRA1A(1p36) com
CT 809	SCC	16	1p36.33	-	FRA1A(1p36) com
CT 866	SCC	16	1p36.33	-	FRA1A(1p36) com
CT 864	SCC	16	1q42.13	ABCB10	FRA1H(1q42) com
CT 755	SCC	16	3q23	SLC25A36	-
CT 785	SCC	16	3q23	SLC25A36	-
			20q11.21	COX4I2	-
		18	-	-	-
CT 1210	SCC	16	3q26.2 and Episomal	MECOM	-
CT 723	SCC	16	3q26.31	-	FRA3C(3q27) com
CT 706	SCC	16	3q26.33	SOX2	FRA3C(3q27) com
CT 892	SCC	16	3q26.33	-	FRA3C(3q27) com
CT 711	SCC	16	3q28 and Episomal	LEPREL1	FRA3C(3q27) com
CT 739	SCC	16	3q28	TP63	FRA3C(3q27) com
CT 999	SCC	16	4q13.3	IL8(~3kb)	-
CT 1114	SCC	16	4q21.22	HNRPDL	-
CT 893	SCC	16	6q22.31	-	-
CT 1117	AC	16	6q22.31	-	-
		18	-	-	-
CT 1122	SCC	16	6q23.3	PDE7B	-

CT 846	SCC	16	6q23.3	MAP3K	-
CT 1019	SCC	16	7q11.21	TYW1	FRA7J(7q11) com
CT 796	SCC	16	8p11.21	-	-
CT 915	SCC	16	8p11.21	-	-
		18	-	-	-
CT 1160	SCC	16	8q24.21	MYC/PVT	Flanked by FRA8E(8q24.1) rare; FRA8C(8q24.1) com; FRA8D(8q24.3) com
CT 714	SCC	16	9q22.32	C9orf3	FRA9D(9q22.1) com
CT 1162	SCC	16	9q22.32 and Episomal	C9orf3	FRA9D(9q22.1) com
		18	-	-	-
CT 753	SCC	16	9q34.11	SLC25A25	-
CT 821	SCC	18	10q23.33	-	FRA10A(10q23.3) rare
CT 859	SCC	16	10q23.31	PTEN	FRA10A(10q23.3) rare
CT 1169	SCC	16	11q13.1 and Episomal	DRAP1	FRA11H(11q13) com
CT 1183	SCC	16	11q13.1	DRAP1	FRA11H(11q13) com
CT 836	SCC	16	11q14.3	-	FRA11F(11q14.2) com
CT 871	AC	18	11q23.3	ARCN1	FRA11G(11q23.3) com
CT 1094	SCC	16	13q	-	-
CT 927	SCC	16	13q12.3	POMP	-
CT 896	SCC	16	13q22.1	-	Flanked by FRA13B(13q21) com
CT 976	SCC	16	13q22.1	KLF5	Flanked by FRA13B(13q21) com
CT 848	SCC	16	17q12 and Episomal	HNF1B	-
CT 914	SCC	16	17q23.1	VMP1	FRA17B(17q23.1) com
CT 1097	SCC	16	18q11.2	-	-
CT 740	SCC	18	18q21.3	-	FRA18B(18q21.3) com
		16	Episomal	-	-
CT 906	SCC	16	19p13.3	-	FRA19B(19p13) com
CT 889	SCC	16	19q13.43 and Episomal	-	FRA19A(19q13) com
CT 918	SCC	16	20q11.21 and episomal	COX4I2	-
CT 1170	SCC	16	20q11.21	COX4I2	-
		18	-	-	-
CT 839	SCC	16	20q13.13	-	-
CT 1015	SCC	16	-	-	-
CT 1082	SCC	16	-	-	-
CT 1101	SCC	16	-	-	-
CT 1107	SCC	16	-	-	-
CT 1108	SCC	16	-	-	-
CT 1135	SCC	16	-	-	-
CT 1161	SCC	16	-	-	-
CT 759	SCC	16	-	-	-

CT 787	SCC	16	-	-	-
CT 792	SCC	16	-	-	-
CT 800	SCC	16	-	-	-
CT 828	SCC	16	-	-	-
CT 868	SCC	16	-	-	-
CT 887	SCC	16	-	-	-
CT 905	SCC	18	-	-	-
CT 934	SCC	16	-	-	-
CT 935	SCC	16	-	-	-
CT 951	SCC	16	-	-	-

Key: – SCC: Squamous cell carcinoma, AC: Adenocarcinoma, Com: Common Fragile site

3.3.7 Determination of viral load by quantitative qRT-PCR

Viral copy number was determined for all the 86 cases with known viral physical status by SYBR green based relative quantitation method. Copy number could not be determined in 2 samples. The average copy number was 47.5. The copy numbers of the individual samples are given in Table 3.6.

Table 3.6: Viral load and HPV physical status in 86 cases

Sample	HPV type	Viral state	Copy no.	Survival (months)
CT 1082	16	Epi	103	36
CT 1101	16	Epi	104	55
CT 1107	16	Epi	30	66
CT 1108	16	Epi	11	65
CT 1135	16	Epi	119	43
CT 1161	16	Epi	90	49
CT 759	16	Epi	70	29
CT 787	16	Epi	10	32
CT 792	16	Epi	52	1
CT 800	16	Epi	<1	32
CT 828	16	Epi	29.9	73
CT 868	16	Epi	3	44
CT 887	16	Epi	16	83
CT 905	18	Epi	207	46
CT 934	16	Epi	3	60
CT 935	16	Epi	24	61
CT 951	16	Epi	10	62
CT 1015	16	Epi	47	62
CT 1019	16	Int	1	92
CT 1094	16	Int	112	3
CT 1097	16	Int	5	39
CT 1114	16	Int	<1	56
CT 1117	16	Int	<1	33
	18	-	3	

CT 1122	16	Int	3	68
CT 1123	16	Int	11	65
CT 1138	16	Int	<1	42
CT 1160	16	Int	260	4
CT 1170	16	Int	<1	29
	18		3	
CT 1183	16	Int	4	56
CT 1194	16	Int	<1	10
CT 1202	16	Int	ND	50
CT 1215	16	Int	115	2
CT 700	16	Int	24	48
CT 702	16	Int	3	54
CT 706	16	Int	73	73
CT 707	16	Int	219	7
CT 709	16	Int	1	32
CT 712	16	Int	2	50
CT 714	16	Int	114	6
CT 716	16	Int	12	59
CT 718	16	Int	ND	27
CT 723	16	Int	1	15
CT 733	16	Int	25	71
CT 739	16	Int	10	.
CT 744	16	Int	180	12
CT 753	16	Int	7	34
CT 755	16	Int	5	24
CT 763	16	Int	3	20
CT 785	16	Int	<1	38
	18	-	5	
CT 793	16	Int	<1	39
CT 796	16	Int	25	32
CT 809	16	Int	13	77
CT 819	16	Int	4	3
CT 821	18	Int	236	55
CT 825	16	Int	19	74
CT 836	16	Int	112	57
CT 837	16	Int	8	5
CT 839	16	Int	43	29
CT 846	16	Int	98	58
CT 859	16	Int	1	51
CT 864	16	Int	4	69
CT 866	16	Int	<1	59
CT 871	18	Int	1	70
CT 892	16	Int	3	19
CT 893	16	Int	1	41
CT 896	16	Int	47	49
CT 906	16	Int	50	44
CT 914	16	Int	<1	12
CT 915	16	Int	<1	8
	18	-	Negligible	
CT 922	16	Int	<1	43
CT 927	16	Int	442	18
CT 940	16	Int	76	61

CT 976	16	Int	4	47
CT 999	16	Int	<1	50
CT 1162	16	Int and Epi	<1	17
	18	-	<1	
CT 1169	16	Int and Epi	4	58
CT 1210	16	Int and Epi	16	53
CT 711	16	Int and Epi	20	67
CT 740	18	Int	<1	13
	16	Epi	1	
CT 752	18	Int and Epi	5	71
	16		Negligible	
CT 777	16	Int and Epi	182	32
CT 848	16	Int and Epi	12	50
CT 889	16	Int and Epi	23	20
CT 912	16	Int and Epi	77	38
CT 918	16	Int and Epi	10	31
CT 959	16	Int and Epi	12	56

Key: – Epi: Episomal, Int: Integrated, ND: Not determined

3.3.8 Physical state of the virus, viral load and clinical outcome

Estimation of viral load in samples with known integration status could provide useful information about natural history of HPV infections and disease prognosis. We therefore, tried to understand whether there was any relationship between these 3 parameters. Comparison with survival of the patients as well as physical status revealed that cases where the virus was in the episomal form, the survival had no association with copy number. But those with integrated form of the virus showed a trend towards reduced survival when the copy number was high as compared to when it was low (Fig. 3.7). Therefore, we believe that the copy number of the virus combined with the physical state might serve as a good prognostic marker for the disease.

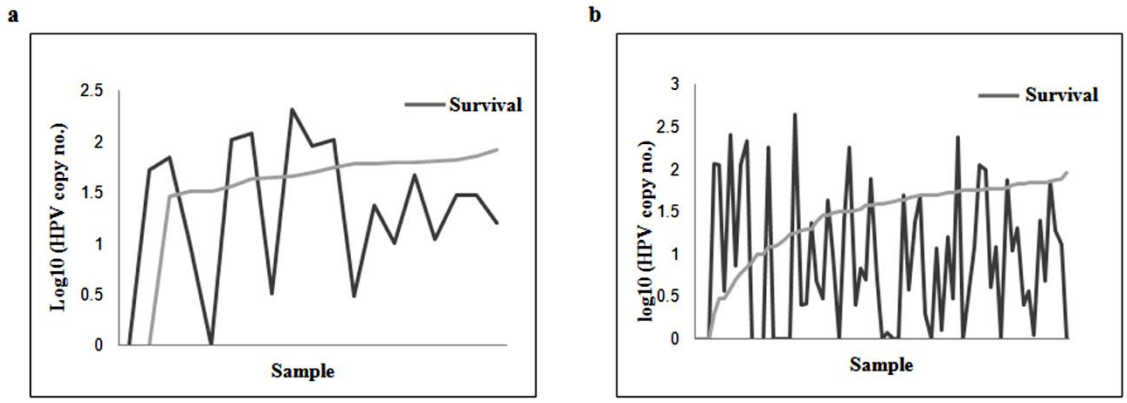

Fig. 3.7: Association of viral load with patient survival and HPV physical status*. When the virus was in the episomal form alone, the survival had no association with copy number. However, cases with only integrated form of the virus showed reduced survival when the copy number was high, as compared to when it was low.*

3.4 DISCUSSION

Integration of the HPV genome is a distinctive step in cancer of the cervix and is believed to be a characteristic of progression of the disease from precancerous lesions to invasive carcinoma (Durst et al., 1985; Kalantari et al., 2001; Klaes et al., 1999; Pett and Coleman, 2007; Wentzensen et al., 2004). However, studies have confirmed presence of only episomal form of the virus in advanced cervical squamous cell carcinomas, thereby establishing that integration might not be absolutely mandatory for the progression of the disease (Gray et al., 2010; Vinokurova et al., 2008). Integration generally occurs downstream of the early genes E6 and E7, often in the E1 or E2 region. The E2 gene gets transcriptionally inactivated once the virus gets integrated due to disruption of its open reading frame. This releases the negative regulation of E2 on the viral oncogenes - E6 and E7, in turn resulting in elevated expression of these oncogenes (Baker et al., 1987; Romanczuk and Howley, 1992). However, a recent study reports no correlation between the expression of viral oncogenes and the physical state of the virus (Hafner et al., 2008). Nevertheless, due to their role in the process of HPV induced carcinogenesis, elucidation of viral integration event becomes important.

Several strategies have been used to study viral integration including Ligation-mediated PCR (Luft et al., 2001), Restriction site-PCR (Thorland et al., 2000) and APOT assay (Klaes et al., 1999). We chose APOT not only to limit our study to integration sites with a transcriptionally active viral genome, but also because APOT assay allows detection of integrated viral genome in clinical lesions even in the presence of a large excess of unintegrated episomal form of viral genomes (Klaes et al., 1999; Vinokurova et al., 2008).

Out of 86 cases of advanced cervical lesions, viral integration was observed in 68 cases, whereas in the remaining 18 episomal form of HPV was detected. Presence of only episomal form in advanced disease stage could be an indication that viral integration might not be solely responsible for the progression of the disease. The frequency of viral integration into the host genome in cervical carcinomas has been reported to be as high as 100% in HPV18 positive tumours (Corden et al., 1999) and up to 80% in HPV16 positive tumours (Melsheimer et al., 2004). In our study we found four HPV18 positive samples where the virus was integrated and one HPV18 positive sample where the virus was episomal. The incidence of integration in HPV16 positive samples was higher (64/68). Compared to a similar study by Nambaru *et al*. in Indian population, the frequency of integration in case of HPV16 was higher in our study (56.7% vs. 79%), although the integration frequency for HPV18 was comparable (80%) (Nambaru et al., 2009). This could be attributed to the different stages of the disease in the two studies.

Comparison of the physical state of the virus (episomal/integrated) with the clinical outcome after radical radiotherapy revealed that patients with episomal form of the virus had increased disease free survival compared to those with integrated form. This observation is supported by various reports which state that the integration event is

associated with a decreased disease free survival (Kalantari et al., 1998; Vernon et al., 1997). However, there are contrasting reports as well, according to which physical state of the virus does not correlate with disease free survival (Holm et al., 2008; Nambaru et al., 2009). This requires further study with increased sample size to come to a conclusion.

Although integration sites were distributed throughout the genome in different samples, there was a preference for certain chromosomal loci such as 1p, 3q, 6q, 11q, 13q, 6q and 20q. Our finding that integration event was observed more than once into 3q28 and 13q22 chromosomal loci is in agreement with recent reports according to which regions in the genome such as 3q28, 4q13.3, 8q24.21 and 13q22.1 are more frequently affected by integration than other parts (Kraus et al., 2008; Schmitz et al., 2012). The chromosomal locus 3q28 involving members of p53 family of transcription factors was also identified by Nambaru *et al*., in their study which was the first of its kind in Indian population (Nambaru et al., 2009). In addition to these, repeated integration was observed into chromosomal loci 1p36.23, 1p36.33, 3q26.33, 3q28, 6q22.31, 6q23.3, 8p11.21, 9q22.32, 13q22.1 and 20q11.21, which may be regarded as 'hotspots' for HPV integration (Fig. 3.8). Most of these loci were also checked for integration at the genomic level to rule out any technical bias. However, integration into 4q and 8q loci was observed in only single cases in our study, which may be attributed to the difference in the ethnicity between the study groups.

The 3q, 13q and 20q loci besides being preferential target for HPV integration have been reported to be sites for genomic instability associated with cervical cancer. Gain of 3q and 20q and loss of 13q has been reported in various stages of the disease (Rao et al., 2004; Scotto et al., 2008a; Wilting et al., 2008). Also more recent reports show that a significant association exists between genomic rearrangement and HPV

integration (Peter et al., 2010). It would therefore be interesting to study whether the preferential integration of the virus into these loci has a role to play in inducing genomic instability. Comparison of the integration site with the clinical outcome after radical radiotherapy revealed that patients with viral integration at chromosomal loci 3q, 13q and 20q had poor prognosis, while those with integration at 1p, 6q and 11q loci showed increased disease free survival. Though the numbers in each group are small considering the fact that the integration event was more frequent into these loci, we speculate site of viral integration into the host genome might have important bearing on disease prognosis.

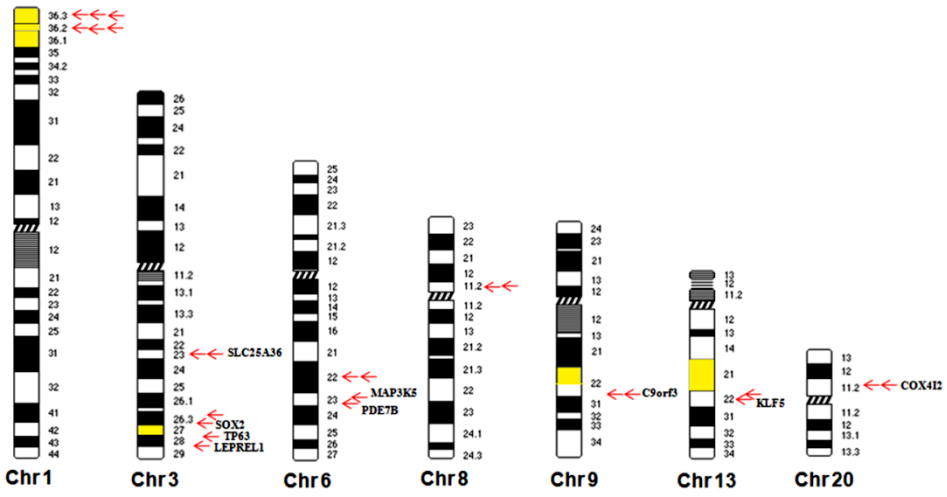

Fig. 3.8: Chromosomal hotspots for HPV integration. *All the chromosomal loci with more than one integration event are depicted. Each red arrow represents a single case and indicates the particular cytoband where integration was observed; genes, wherever involved in the process of HPV integration, are also shown; yellow regions represent fragile sites at or near the respective integration locus.*

Most of the integrations (28/48) were found to be located within or near certain genes, including candidates like myc, transcription factors such as TP63, MECOM, etc. This observation is supported by previous report by Wentzensen et al wherein they have shown involvement of tumour related genes (myc and TP63) in HPV integration process (Wentzensen et al., 2002). Studies in our lab have shown that some of the genes within

which integration was observed, such as ABCB10, SLC25A36, IL8, COX4I2, HNF1B, myc, demonstrated increased expression, implying *cis* regulatory effects being exerted by the viral genome.

Integration was also observed to be more frequent near or within fragile sites (29/ 48 cases). These are specific regions in the chromosomes that nonrandomly undergo break in response to certain stress, making them susceptible targets for foreign DNA integration. Viral integration near or within such sites have frequently been reported (Dall et al., 2008; Kraus et al., 2008; Matovina et al., 2009; Thorland et al., 2003; Thorland et al., 2000; Wentzensen et al., 2002). Besides, coding regions and fragile sites, in a few cases site of integration was associated with one or more micro RNAs such as miR-1204, miR-1205, miR-1207, miR-1302-5, miR-645, miR-7-3, miR-220b, miR-23b, miR- 27b, miR- 24-1, miR-1302-8, and miR-1825. The association of sites of HPV integration with micro RNAs have been shown by Nambaru *et al*., where they identified 53 miRNAs including a miR cluster which were associated with HPV integration (Nambaru et al., 2009) Similar report has also been recently by Schmitz et al (Schmitz et al., 2012). The role of miRNAs in the regulation of important cellular processes such as development, proliferation, differentiation and apoptosis have been well established; also they are often reported to be deregulated in cancer cells (Calin and Croce, 2006; Iorio and Croce, 2009). Whether HPV could bring about changes in the genome through integration within or near a miR would be interesting to study.

Although physical status of the virus might act as a prognostic marker for cancer of the cervix, information of the viral load could complement it and provide useful insights into the natural history of HPV infections and their relationship to disease. SYBR green based relative quantitation was done for estimation of viral load of HPV16 and HPV18 in the 86 samples with known viral physical status (Peter et al., 2010). This

method was used mainly due to its reduced cost and simplicity as compared to Taqman chemistry. The mean copy number for the cases with only integrated viral genome was 43. However, from the APOT results it is apparent that only one (or at the maximum 2) of these sites were transcriptionally active.

The association between viral load, physical status and disease prognosis was quiet complex. It was observed that where the virus was in the pure episomal form, survival of the patients had no association with viral load. But those with integrated form of the virus showed reduced survival when the copy number was high as compared to when it was low. These results hint at the fact that copy number of the virus combined with the physical state could present a much clear scenario of the disease.

3.5 CONCLUSION

We have successfully identified the physical status of HPV and the viral load in 86 cases with advanced cervical carcinoma, as well as located the exact site of integration into the genome in 48 cases. Viral integration was more frequent in chromosomal loci 1p, 3q, 6q, 11q, 13q and 20q. Also, regions within or near fragile sites and known genes were identified as integration ‘hotspots’.

Association with the clinical outcome revealed that both physical status and site of integration of the virus could have important bearing on disease prognosis. It was also observed that estimation of the viral copy number could further complement the information and together with physical status of the virus provide useful insights into the pathophysiology of HPV infection and its relationship to disease prognosis.

Chapter– 4

Identification of genomic alterations in cervical cancer biopsies by exome sequencing

4.1 INTRODUCTION

Athough HPV has been identified as the prime etiological factor for cervical cancer, several genetic alterations also play an important role in the genesis of the disease. A wide array of alterations including both germline and somatic mutations (Cui et al., 2009; Hu et al., 2012; Madsen et al., 2008; Storey et al., 1998; Wang et al., 2012; Wingo et al., 2009; Wong et al., 1995; Yee et al., 2000), indels (Maliekal et al., 2003; Nakashima et al., 1999) and chromosomal aberrations (Narayan and Murty, 2010) have been reported to have a significant role in either directly causing cervical carcinoma or predisposing an individual to disease development. However, these studies were mostly based on candidate gene approach and arrays; and till date there are no reports describing the genomic landscape of the disease.

With the advent of 'next generation' sequencing (NGS) technologies, the focus in cancer genomics has shifted from single-gene approach and arrays to comprehensive genome-wide approaches which are capable of providing voluminous information. NGS methods such as whole exome sequencing can provide increased sequence coverage of a particular region of interest at high throughput and lower cost, compared to whole genome sequencing (Meyerson et al., 2010). Also, despite constituting only about 1% of the genome (37.6 Mb), exons are reported to harbour most variations and hence sequencing them can yield significant information (Botstein and Risch, 2003). Exome sequencing has been quiet successfully employed in identifying novel mutations in diseases characterized by marked genetic heterogeneity like cancer, including rare forms of the disease (Agrawal et al., 2011; Byun et al., 2010; Harbour et al., 2010). Therefore, making use of this technique, we hope to obtain large amounts of useful information on the genetic variations in cervical cancers which could help in understanding the fundamental rules of the disease.

4.2 MATERIALS and METHODS

In order to study the genetic variations in cervical cancer, initially it was proposed to use SNP arrays, however, with the availability of NGS technology it was decided to carry out whole exome sequencing as it is known to provide much more information about the genetic make-up of an individual including the novel ones, which could be missed in the SNP arrays.

MATERIALS

Source of chemicals and reagents

Reagent	Company
DNA Mini Kit	*Qiagen, Hilden, Germany*
RNase A	*Sigma, St. Louis, MO, USA*
Paired-End Sample Preparation kit	*Illumina, San Diego, CA, USA*
SureSelect Human All Exon kit	*Agilent Technologies, Santa Clara, CA, USA*

Clinical sample

The patient population was the same as described earlier on pg. 65. From this population, whole-exome sequencing of 11 tumours (all stage IIIB and >70% tumour) and 8 matched controls (DNA from blood or buffy coat from same individual) was done in two separate sets; this included:

Dataset-I: 6 tumour and 3 matched blood DNA

Dataset-II: 5 tumour and 5 matched blood DNA

The quality of runs from the two sets was not comparable as for the dataset-I, the technique was not properly standardized and the library preparation was sub-optimal. However, the second set of runs was of a much better quality. Hence, although the methodology followed was exactly the same for the two sets, they were analyzed separately and therefore, have been discussed under separate headings.

4.2.1 Isolation of genomic DNA

Genomic DNA was isolated using DNA Mini kit from tumour and blood according to the manufacturer's instructions, and treated with RNaseA to remove any RNA contamination that might interfere with subsequent library preparation and capture.

Dataset-I

4.2.2 Exome capture and sequencing

Exome capture and sequencing was carried out at Genotypic Technology (P) Ltd, Bangalore, on Illumina Genome Analyzer IIx (GAIIx) platform. Following steps were involved in the sequencing process:

- **Paired End library preparation -** This was done using Paired-End Sample Preparation kit, according to the manufacturer's instruction. Briefly, the genomic DNA was sheared to less than 800 bp fragments and adapters ligated after end-repair and 3'-dA overhang generation. The ligated products were purified through gel extraction and finally adapter- ligated library was amplified.
- **Hybridization -** Hybridization of the DNA library to biotinylated RNA library was carried out using SureSelect Human All Exon kit that targets 37 Mb sequence from exons, as per manufacturer's instruction. This was followed by selection of DNA-RNA hybrids by streptavidin coated magnetic beads and digestion of RNA.
- **Post-Hyb processing -** The captured library was PCR amplified, purified and finally sequenced on Illumina GAIIX sequencing platform (Illumina, San Diego, CA, USA) by multiplexed paired end sequencing using 2 pools of samples.

4.2.3 Analysis for identification of single nucleotide variations and indels

The raw reads obtained as short fragments (~54 bp) were first filtered by the software SeqQC (www.genotypic.co.in/SeqQC.html). The filtered reads were then aligned against the *Homo sapiens* genome (Hg19 UCSC Build) by the program BWA (Li and Durbin, 2010). The aligned SAM-BAM file was used for variant calling using Samtools (Li et al., 2009). The derived single nucleotide variations (SNVs) and insertions and Deletions (InDels) were next compared with dbSNP and '1000 Genome' to identify the novel ones. In order to further shortlist the potential somatic variations (mutations), data from only the paired samples was considered. The 3 unpaired samples were kept as validation sets. KEGG pathway and GO analysis was done for the entire novel, somatic and non-synonymous variations to shortlist important candidate genes harboring the variations. These selected variations were further subjected to prediction tools such as SIFT (Kumar et al., 2009) and Polyphen 2 (Adzhubei et al., 2010) for identifying whether a given amino acid substitution affects protein function. In addition, all variants were screened for mutations reported in OMIM, HGMD and COSMIC databases.

4.2.4 Validation of the data by Sanger sequencing

A few of the selected variations, following all the analysis, including SNVs as well as short indels, were validated by PCR amplification followed by Sanger sequencing. Primers were designed flanking the variation - ~250 bp upstream and 250 bp downstream (Table 4.1) and standard PCR conditions were followed. The annealing temperature ranged between 55- 65˚C.

Table 4.1: Primers used for validation of dataset-I

Name	Sequence (5'-3')	PCR product size (bp)
RNASEL	F - ACTCTAGGCCTTTCCTCTCTGCAA	403
	R - TGAAGACGTTCTGTGAGGGCAGC	
PTPRJ	F - CTCAGGCTCTGCTTGATTTCCC	268
	R - CGTGACGTGCCATCTGCGTA	
ZFP64	F - AATAACAGGCAGGACTCACCCGT	307
	R - TTTCCTGCTGCCATCTGTGTGTGT	
SREBF1	F - AGCTGCCTGGGGAGCTGGTAT	258
	R - CCTGGTGGCATTGCCCACACATA	
FBN1	F - TTGGGTAGGCATGTCCAGCCTGT	230
	R - ATGTGAGAGGCTTTGTTGACTGGAC	
BANF1	F - GGGGTGGCTGCGTGTACCTA	387
	R - AGGCGTCGCACCACTCTCGAA	

4.2.5 Validation of the data by Customized SNP Array:

Apart from validation by Sanger sequencing, the exome sequencing data was also validated by customized SNP array. Accordingly, 700 known (reported in dbSNP version131) and 3178 novel nonsynonymous variations across all the samples (except B_1123 which was totally exhausted) were selected for validation on an 8x60K customized Agilent oligonucleotide microarray. Validation of the few known variations was done as a check for the quality of the exome sequencing data. For each variation, 8 probes with Tm between 50° to 70° C and probe lengths 23, 27 or 31 bp depending on Tm, covering all 4 bases (A, T, G, C; 4 sense and 4 antisense orientation) were designed. The GC percentage was calculated based on Tm range and all probes were designed with the SNP base located in the centre of the probe. The microarray slides were scanned using the Agilent microarray scanner G2505C (Agilent Technologies, Santa Clara, CA, USA) at 3 micron resolution followed by analysis using Feature Extraction software Version (Agilent 10.5.1). Variations were detected based on comparison of signal intensities (both sense and antisense direction) with reference base. Finally the results were compared with those from the exome sequencing data.

4.2.6 Analysis for detection of Copy Number Variation (CNV) and Loss of Heterozygosity (LOH)

Identification of LOH and CNV was done from the exome sequencing data of two paired samples, where the coverage between blood and tumour was comparable using 'ExomeCNV' package as described by Sathirapongsasuti *et al.* (Sathirapongsasuti et al., 2011). The steps involved in the analysis can be summarized in Fig. 4.1. This part of the work was done at Philips Research Asia, Bangalore.

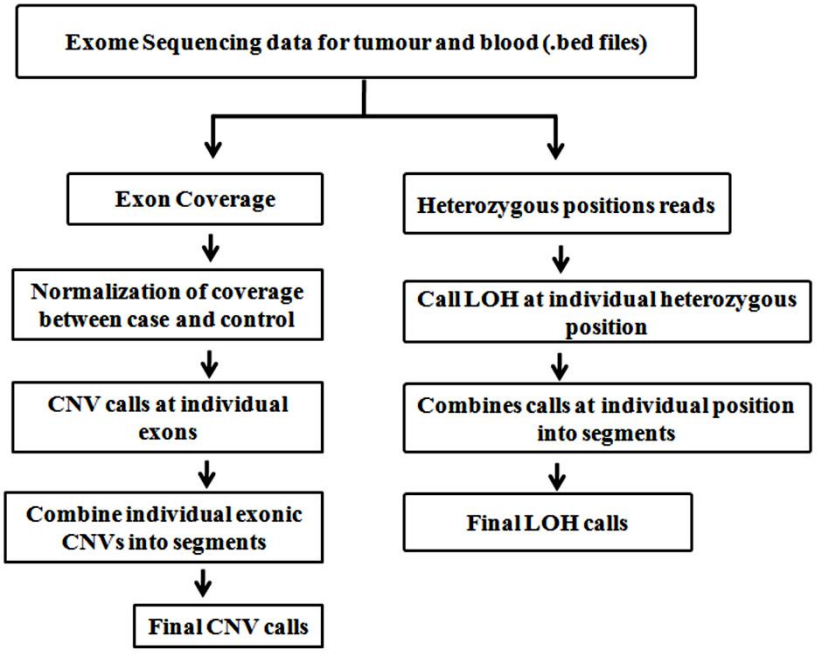

Fig. 4.1: Outline of the methodology used for identification of CNV and LOH. *Steps involved in CNV and LOH detection by ExomeCNV package is shown.*

4.2.7 Validation of Copy Number Variation by qRT-PCR

As a proof of principle, one of the predicted CNVs in 17q12 for the paired sample T_785 was validated by SYBR green based qRT-PCR on ABI Prism 7700 Sequence Detection System *(*ABI, Foster City, CA, USA) as described on pg. 91. DNA from the

matched blood was used as control and the single copy gene TBP as reference for relative quatitation. Copy number was finally evaluated by using the comparative ΔΔCt cycle threshold method.

Dataset-II

Paired end library preparation, exome capture and sequencing was done exactly the same way as for dataset-I. The average read length for the dataset-II was ~72 bp. Analysis of the sequencing data, including alignment, variant calling, filtering for novel, somatic and nonsynonymous single nucleotide variations, KEGG and GO analysis was performed in an identical manner as done for dataset-I. However, the analysis for the indels from the dataset-II is still ongoing.

Some of the SNVs in the top candidate genes were validated by PCR followed by Sanger sequencing as done for dataset-I. Criterion for designing the primers and PCR conditions were the same as for dataset-I (Table 4.2).

Table 4.2: Primers used for validation of dataset-II

Name	Sequence (5'-3')	PCR product size (bp)
FGF7	F - CCAACTTTGCTCTACAGATCATGC	292
	R - TGCTAAGGTTCATGAGCAGTAC	
MAP3K3	F - TGCTGACAGAGAAACCACCGTGG	311
	R - CTCCATGGTCCACAGCCTTGCC	
DIDO1	F - CCCTTTATCCCTTGGTCTTCGC	312
	R - GCTTGTCACTCACGTCCCAT	
RASA1	F - GGCATATTACATAAGCTTTGGAAT	281
	R - CTGCCAAAGAGACTATCATGAAC	
SOS2	F - GTGTATTTGGAGAGGTTGGTGC	264
	R - AGGTAGCTCTAAGTCACAGAAAC	

4.3 RESULTS

Exome sequencing of 11 cervical cancer biopsies with locally advanced disease and 8 matched controls (blood) was done on Illumina GAIIx platform in two sets. The quality of runs from dataset-II was more superior to those obtained in the dataset-I,

both in terms of coverage and read depth. Therefore, the two data sets were analyzed separately; to avoid loss of significant amount of information. The methodology followed was exactly identical for the two sets. Accordingly, the results for the two sets are discussed separately under dataset-I and dataset-II in results.

Results from Dataset-I

Samples taken in this dataset were DNA from 3 paired (tissue and blood from the same individual) and 3 unpaired samples.

4.3.1 Exome capture, sequencing and analysis

After filtering, reads were aligned against the *Homo sapiens* genome (Hg19 UCSC Build) by the program BWA. The alignment statistics are given in the Table 4.1.

Table 4.3 Alignment statistics – Dataset-I (Genotypic Technology P. Ltd.)

Sample*	**B_755**	**T_755**	**B_1123**	**T_1123**	**B_785**	**T_785**	**T_837**	**T_887**	**T_940**
Total Reads	28639119	53462144	67772740	22512450	20036640	38045654	20929914	23173283	24802172
% Reads aligned to genome	98.62	99.10	97.87	98.36	97.37	99.21	99.26	99.17	99.11
Reads aligned to targets	17808671	12684711	11078107	7831065	7401569	8499890	8276237	9973663	10387685
% Reads aligned to targets	62.18	23.73	16.35	34.79	36.94	22.34	39.54	43.04	41.88
Reference Sequence Length	37364139	37364139	37364139	37364139	37364139	37364139	37364139	37364139	37364139
% Covered with at least 5X read depth	88.52	70.77	55.33	70.22	61.68	57.74	70.74	72.44	80.75
% Covered with at least 10X read depth	75.34	51.11	43.12	49.48	43.31	35.41	50.64	54.03	63.03
% Covered with at least 15X read depth	61.79	36.99	35.29	34.89	31.19	22.46	36.27	40.84	47.98
% Covered with at least 20X read depth	**49.58**	**27.00**	**29.68**	**24.92**	**22.98**	**14.82**	**26.25**	**31.46**	**36.37**

* B (blood) / T (tumour) followed by the sample number

After variant calling using Samtools, about 14400 unique sequence variants including single nucleotide variations and indels across all 9 samples were detected.

Of all these variations, a total of 7407 SNVs, 94 small deletions and 71 small insertions were not reported in dbSNP131 and hence were considered putative novel sequence variants.

The distribution of the SNVs and small indels across the samples is summarized in Fig. 4.2.

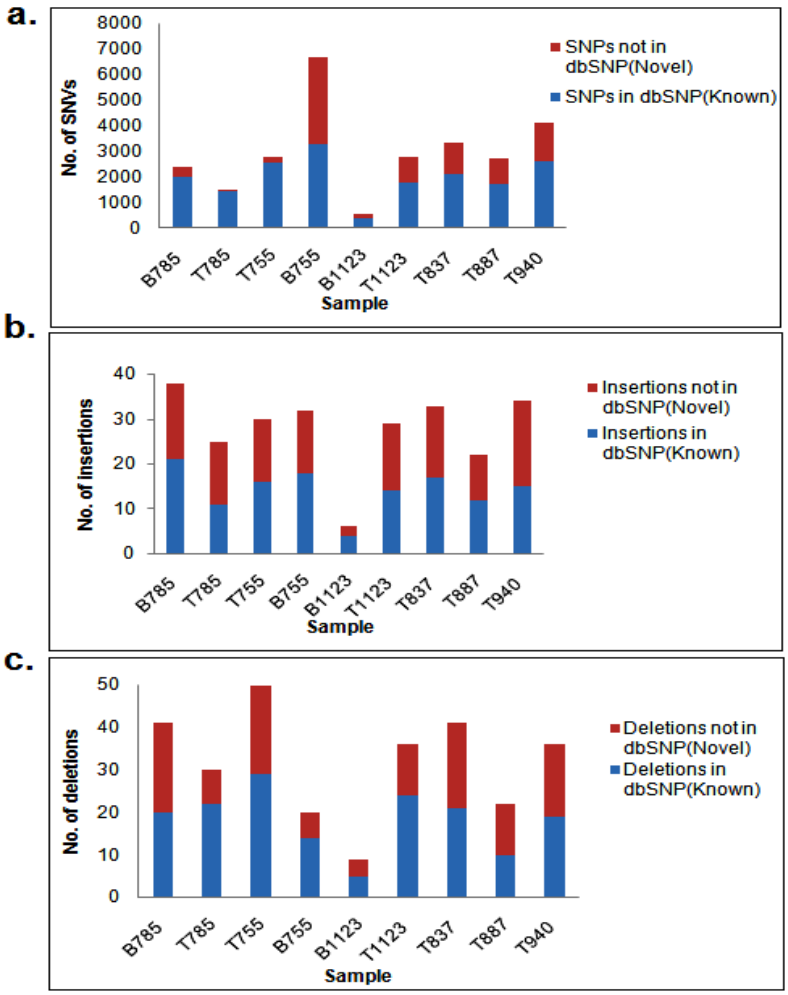

Fig. 4.2: Distribution of the SNVs, insertions and deletions from Dataset-1. *Number of known and novel SNVs (a), Insertions (b) and Deletions (c) as obtained after analysis of the exome sequencing data is shown. Each bar represents a single sample. 'Known' variations are those that are reported in dbSNP.*

The graphs clearly show that the sequencing data was not ideal since the distribution of the SNVs across the samples was not uniform. Further, filtering for

the somatic and nonsynonymous variations yielded 874 potential somatic SNVs, 30 small deletions and 22 small insertions.

Following KEGG and GO analysis, for narrowing down the variations with important/defined functions, the number reduced to 383 SNVs, 12 small deletions and 5 small insertions (Fig. 4.3). KEGG analysis identified genes such as NTRK1, PIK3CB, STAT5A, CSF2RB TAOK2, NRG3, CCNB3, NFAT5 involved in important signalling pathways as well as various members of the integrin family involved in cellular processes like ECM-receptor interaction and focal adhesion.

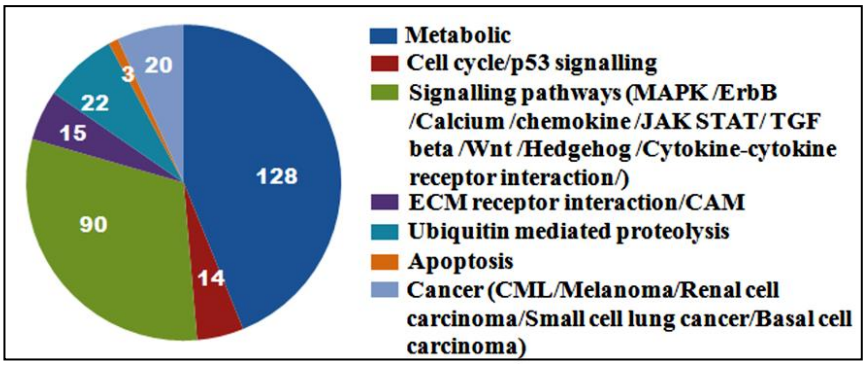

Fig. 4.3: KEGG pathway analysis. *KEGG pathway analysis was done for the genes harbouring novel, somatic and nonsynonymous variations. The important pathways are shown in the pie-chart. Metabolic pathway genes were the most common ones, followed by the ones involved in signal transduction.*

The steps involved in short listing the variations are summarized in Fig. 4.4. The 383 putative novel, somatic and nonsynonymous variations were further subjected to prediction analysis using SIFT and Polyphen-2 to predict the functional significance of a particular alteration. These prediction tools are based on sequence homology and physicochemical properties. SIFT analysis predicted 164 variations to have a potential deleterious effect on the protein function while with Polyphen 2, 118 variations were predicted to be damaging. Both the prediction tools predicted 71 common variations to be deleterious (Fig. 4.5).

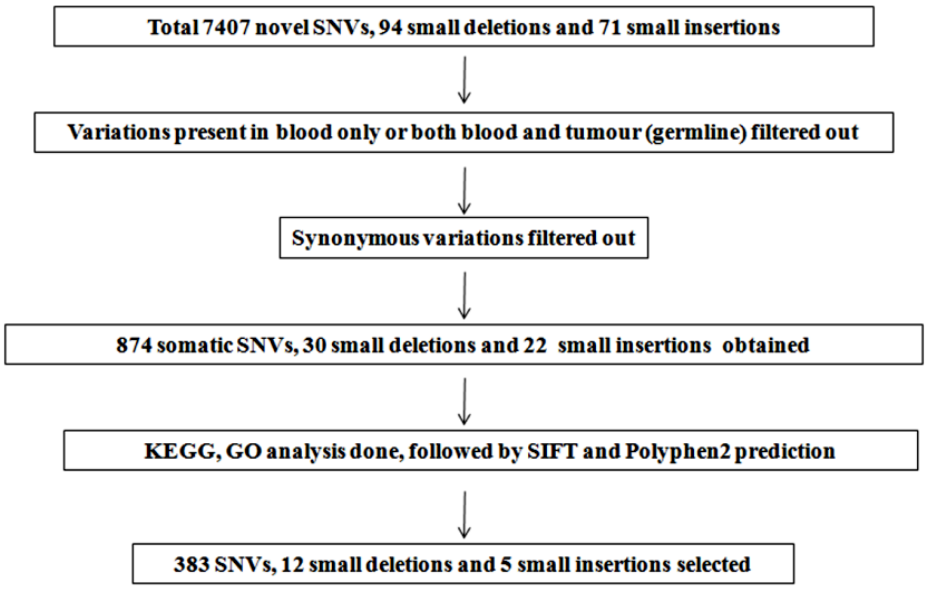

Fig. 4.4: Flow chart for short listing the SNVs and indels. *The methodology employed for short-listing the variations is shown.*

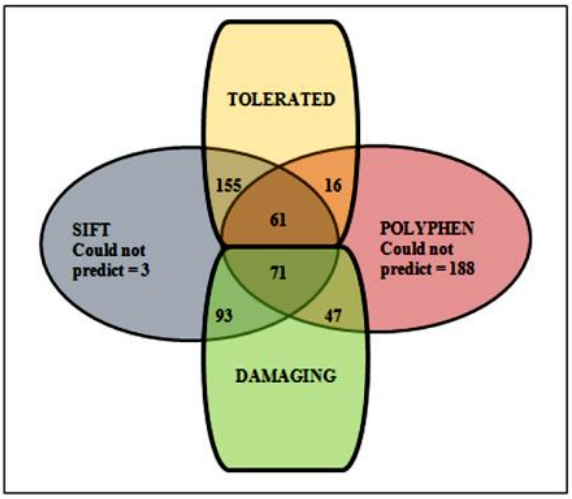

Fig. 4.5: SIFT and Polyphen prediction of selected variations from dataset-I. *SIFT and Polyphen analysis was done for the 395 putative novel, somatic and nonsynonymous variations with important function as predicted by KEGG and GO analysis. 73 variations were predicted to be damaging by both the tools*.

4.3.2 Screening for identical variations from the OMIM, HGMD and/or COSMIC databases

All the variations were screened for identical matches from the OMIM, HGMD and COSMIC databases. Variations in genes such as KCJN12 (reported in ovarian

carcinoma) and CDC27 (reported in lung carcinoma) from the COSMIC were found to be identical. Moreover there were other identical matches from the OMIM and HGMD databases as well as from published literature as shown in Fig. 4.6.

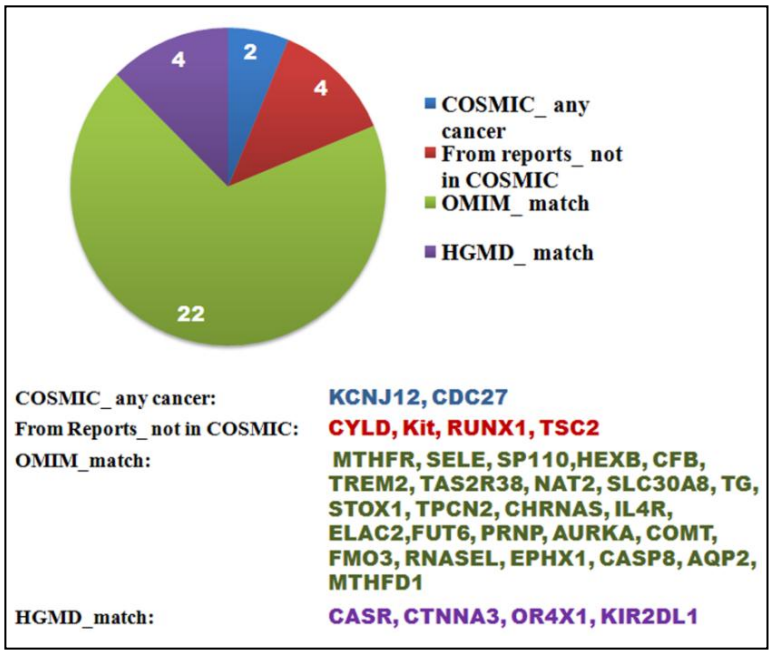

Fig. 4.6: Identical variations reported from other databases*. All the variations were checked for exact match in databases such as COSMIC, OMIM, HGMD and also in published reports.*

4.3.3 Validation of the data by Sanger sequencing

Validation of these somatic variations was done by Sanger sequencing. However, most of the variations that were selected for validation could not be validated by this method. This may be due to the fact that Sanger sequencing can accurately detect 1:2 or 1:3 representations of heterozygotes and since almost all of these variations were heterozygous, there is a chance of missing out the mutant allele. Another possibility was that the quality of the runs which was sub-optimal.

The putative somatic variations (Table 4.4) in genes such as RNASEL, PTPRJ, ZFP64, NFAT5 FBN1 SREBF1 and BANF1 were validated by Sanger sequencing. However these variations were detected in the blood as well, indicating that they were

not true somatic as given by the exome sequencing data. Nonetheless, they could represent novel germline mutations contributing to cervical cancer genesis and/or predisposition.

Table 4.4: Novel variations from dataset-I validated by Sanger sequencing

Gene	No. of paired samples with the variation	Somatic/ Germline	Nature of the Variation	Amino acid change	Identical variation from other databases
RNASEL	1	Germline	Missense point	R462Q	OMIM
PTPRJ	1	Germline	Missense point	R1222H	-
ZFP64	1	Germline	Nonsense point	Y210X	-
NFAT5	1	Germline	Missense point	Y1188C	-
FBN1	1	Germline	Deletion	del AGG	-
SREBF1	1	Germline	Deletion	del AGG	-
BANF1	1	Germline	Deletion	del CACTGAG	-

Representative gel images and electropherograms are shown in Fig. 4.7 and 4.8.

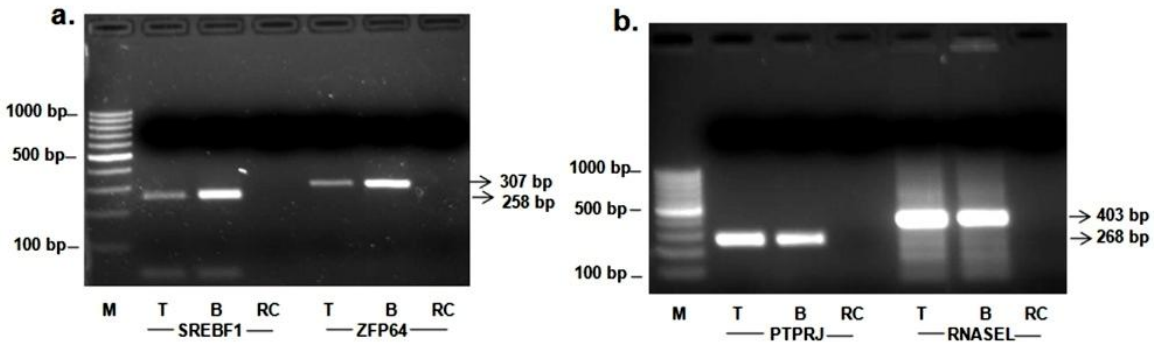

Fig. 4.7: Validation of exome sequencing data from dataset-I by PCR. *Representative gel images for validation of dataset-I are depicted. PCR was carried out with blood and tissue DNA using specific primers. The amplified PCR products were purified and subjected to Sanger sequencing for identification of specific mutation. T= Tissue, B= Blood, RC=reagent control, M= molecular weight marker*

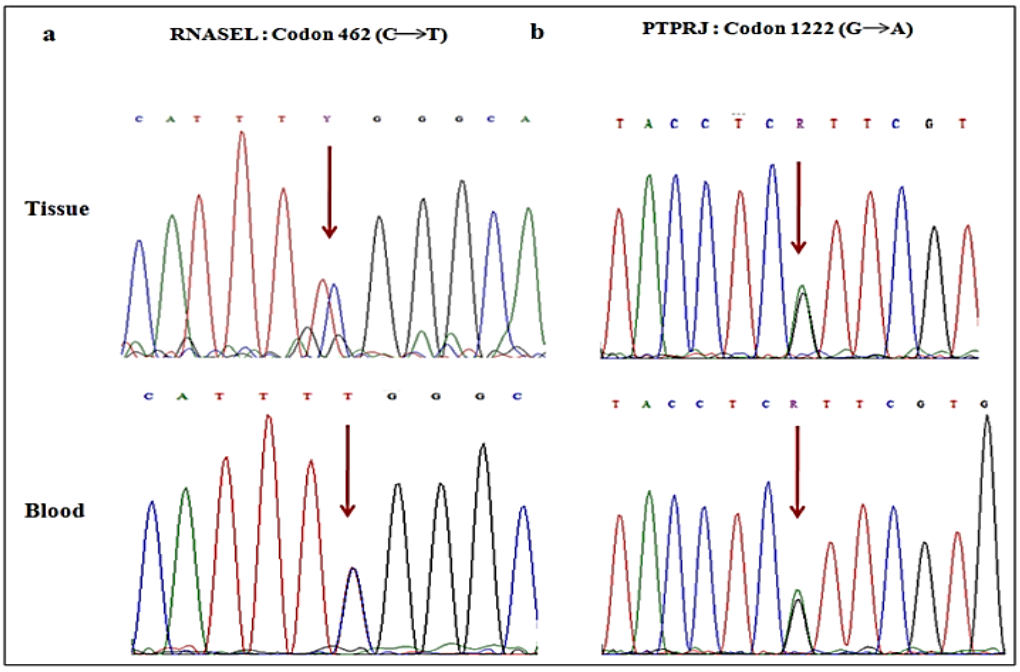

Fig. 4.8: Representative electropherograms from Dataset-I obtained by Sanger sequencing. *SNVs in RNASEL and PTPRJ in different paired have been shown. Presence of both wild type and mutant allele confirms the variations to be heterozygous. However, these were present in the corresponding blood as well, indicating them to be germline and not true somatic as was obtained in the exome sequencing data.*

4.3.3 Validation of the data by Customized SNP Array

Since the total number of variations obtained after analysis was quiet large, and there was a chance that Sanger sequencing was not able to detect the heterozygotes adequately, a customized SNP array was performed wherein a large number of potential variations were validated in a single experiment for all the samples except the sample B_1123 that was exhausted and was not available for validation. About 3178 novel nonsynonymous (somatic as well as germline) and 700 known SNVs were selected for validation on a 8x60K customized Agilent oligonucleotide microarray. From this 1455/3178 novel nonsynonymous and 192/700 known single nucleotide variations were validated. Under stringent conditions, with a very high cut-off value being kept for a SNV to be called as positive, the overall concordance rate between the exome sequencing and SNP array data was found to be 33% (Fig.4.9).

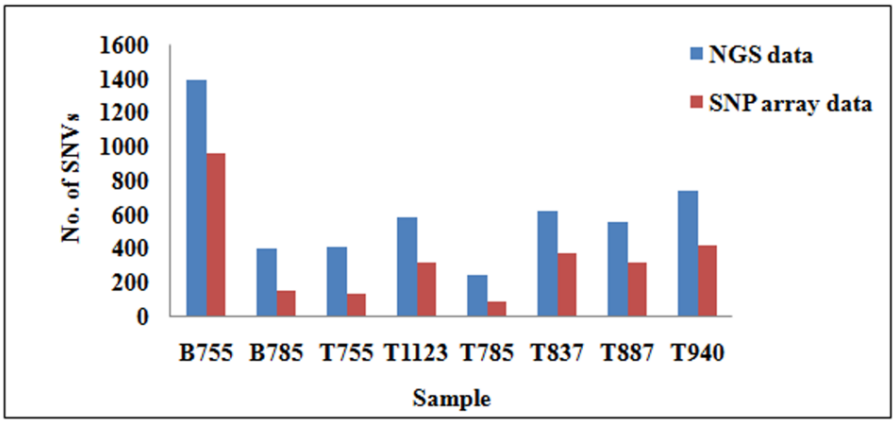

Fig. 4.9: Graph showing concordance between whole exome sequencing and microarray data. *Paired- end sequencing was done on Illumina GAIIx platform and a set of 3178 novel nonsynonymous and 700 known SNVs were selected for validation on a 8X60K customized Agilent microarray. The fraction of the exome sequencing data that was validated by DNA microarray is depicted. The sample B1123 could not be taken for microarray validation as it was totally exhausted.*

However, non-confirmation in the array did not mean that a particular variation call was incorrect since a very high threshold was used as cut off while analyzing the array data. Although, most of the SNVs that were validated were germline in nature, variations in genes such as OR51F2 (Y75C) and YWHAH (Q192K) were not present in the corresponding blood, and hence might be considered as somatic in nature.

4.3.4 CNV and LOH identification:

CNV and LOH analysis was done from the exome sequencing data of the paired samples using 'ExomeCNV' package. One of the requirements for this analysis is that the coverage between blood and tumour should be comparable. Two of the three samples (CT 785 and CT 1123) met this criterion and hence CNV and LOH analysis was done for these two. Some of the major copy number alterations that were predicted by the 'ExomeCNV' package were in chromosomal regions 1q, 3p and 3q, 8q, 11q, 17q

and Xq. A number of these aberrations such as those in 3q, 8q, 17q12 and Xq28 have already been reported.

The detection for LOH was not very accurate. However, some of the chromosomal regions that showed putative loss of heterozygosity included 2q, 3p, 5 and 11q (Fig.4.10). Nonetheless, the CNV and LOH results are only predictions and require validation by more robust methods.

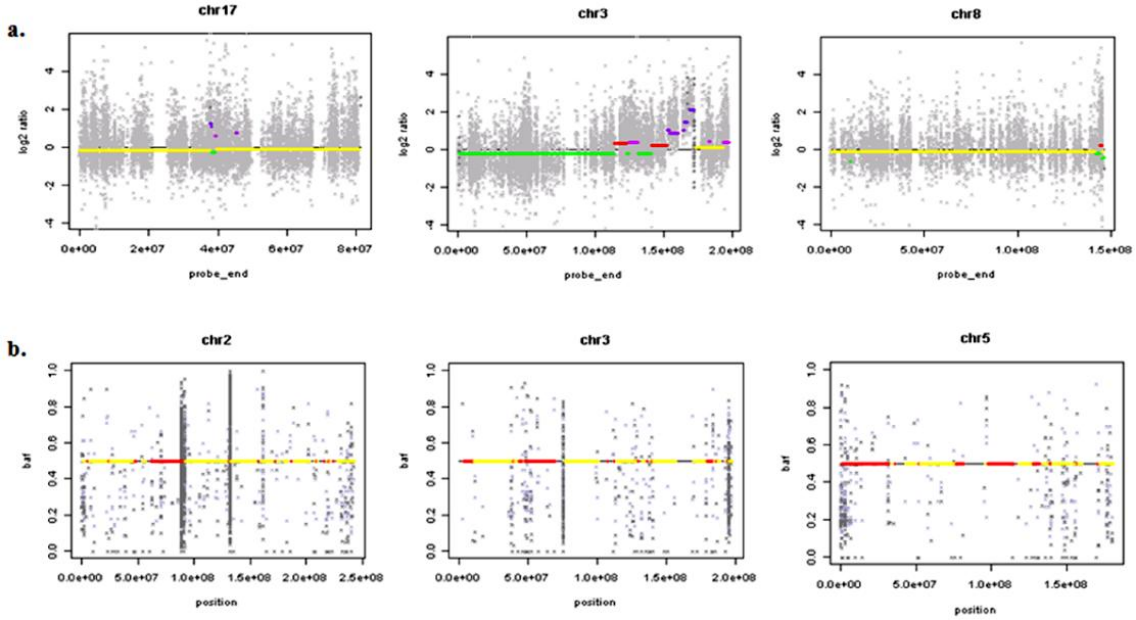

Fig. 4.10: CNV and LOH results. *CNV and LOH were identified from the exome sequencing data using 'ExomeCNV' package. Copy number alteration in chromosome 17, 3 and 8 are shown. Copy Number states in ascending order are represented by green, yellow, red, purple and blue (a). LOH in chromosome 2, 3 and 5 are depicted yellow represents no LOH and red LOH (b) (Philips Research Asia).*

Validation of Copy Number Variation by qRT-PCR

The predicted CNV in chromosomal loci 17q12 in T_785 was validated by qRT-PCR by relative method of quantitation, using TBP gene as reference and B_785 as control. $2^{-\Delta\Delta Ct}$ method of quantitation identified a 4- fold gain in copy number in the 17q12 region in the tumour as compared to blood, thereby corroborating the exome sequencing data.

Results from Dataset-II

The second set of sequencing was done much later (due to the high cost of sequencing and paucity of fund), in order to increase the sample size as well as to add on to the dataset-I. DNA from 5 cervical biopsies and the corresponding blood was used. Very preliminary data analysis and validation has been reported in this thesis.

4.3.4 Exome capture, sequencing and analysis

Exome capture, sequencing and analysis was performed exactly the same way as described for dataset-1. The alignment statistics is given in Table 4.5.

Table 4.5: Alignment Statistics - Dataset-II (Genotypic Technology P. Ltd.)

Sample*	**B_937**	**T_937**	**B_938**	**T_938**	**B_1099**	**B_1103**	**T_1099**	**T_1103**	**B_999**	**T_999**
Total Reads	34414310	41341708	37132906	38398592	50817255	44909761	30074211	37798969	31324870	38795066
% Reads Aligned to Genome	99.90%	99.91%	99.87%	99.88%	99.9194	99.9192	99.9418	99.9157	99.75%	99.86%
Reads Aligned to target	22146897	24768309	24090941	24352766	30583166	25516753	17856586	21783586	21452913	26305123
% Reads Aligned to target	64.42%	59.97%	64.96%	63.49%	60.23%	56.86%	59.41%	57.68%	68.66%	67.90%
Ref. Sequence Length	37364139	37364139	37364139	37364139	37364139	37364139	37364139	37364139	37364139	37364139
Total Ref. covered	36063930	35978018	36232212	36129283	36392722	35836909	35967452	35905764	36397905	36469016
% Covered with at least 5X Read Depth	88.68	88.25	90.57	89.70	91.1608	86.9035	86.1269	86.9935	91.20	92.30
% Covered with at least 10X Read Depth	80.44	79.99	83.28	82.21	83.89	77.74	74.82	77.05	83.66	85.88
% covered with at least 15X Read Depth	73.43	73.20	76.80	75.76	77.187	70.1399	64.621	68.1469	76.61	79.77
% Covered with at least 20X Read Depth	**67.15**	**67.37**	**70.79**	**69.89**	**70.9424**	**63.4769**	**55.1114**	**59.8639**	**69.92**	**74.01**

* B (blood) / T (tumour) followed by the sample number

After variant calling with Samtools, about 47051 variations were obtained, of which 4050 SNVs, were found to be novel (Fig. 4.11).

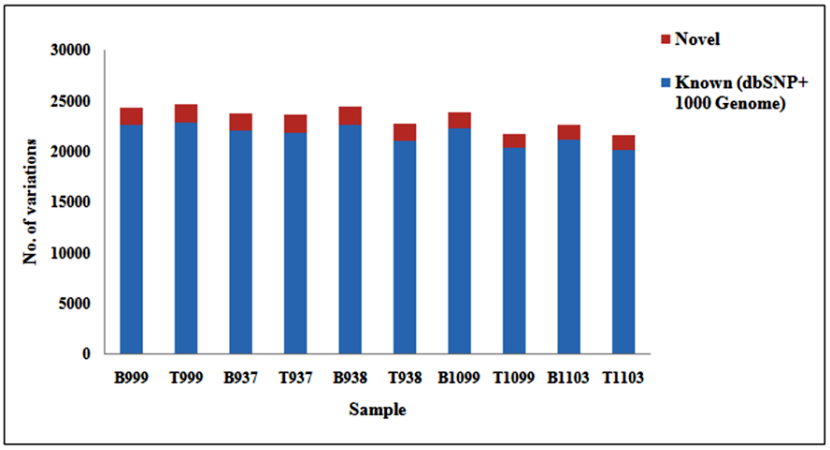

Fig. 4.11: Distribution of known and novel SNVs from Dataset-2. *Each bar represents a single sample, and the number of known and novel variations reported by the whole exome sequencing data for the 2nd set of samples is shown.*

From the graph it is apparent that distribution of novel and known variations in this set of samples is far more uniform as compared to that of the dataset-I (Fig. 4.1 a). More than 95% variations were in dbSNP and/or 1000 genome. Out of the 4050 novel SNVs, 300 were found to be somatic and nonsynonymous. KEGG and GO analysis identified 149 variations with important functions related to several cellular processes. Of these, 11 variations were of 'stop gain' nature. SIFT and Polyphen-2 analysis identified 96 variations to be damaging, with 51 being predicted as damaging by both the tools (Fig.4.12).

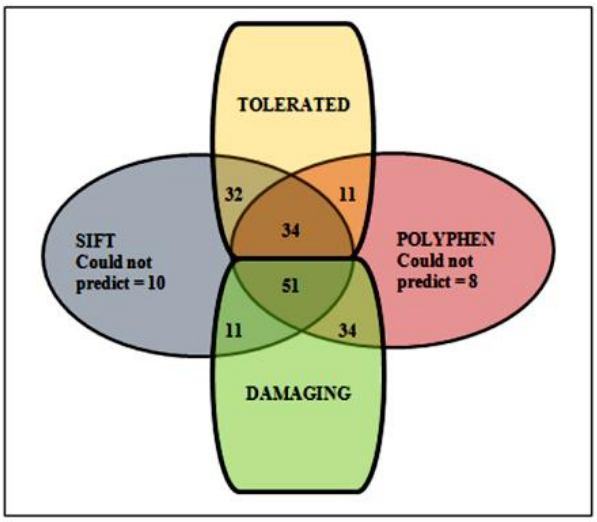

Fig. 4.12: SIFT and Polyphen prediction of selected variations from dataset-II. *SIFT and Polyphen analysis was done for the 149 putative novel, somatic and nonsynonymous variations with important function as predicted by KEGG and GO analysis. 51 variations were predicted to be damaging and 34 tolerant by both the tools.*

4.3.5 Validation of the data by Sanger sequencing

A panel of 10 variations in 9 candidate genes were selected for validation by Sanger sequencing. Of this, variations in important candidates such as FGF7, MAP3K3, RASA1, DIDO1 and SOS2 were validated as somatic variations (Table 4.4).

Table 4.6: Variations from dataset- II validated by Sanger sequencing

Gene	No. of samples with the variation	Somatic/ Germline	Nature of the Variation	Amino acid change	Domain	Identical variation from COSMIC
FGF7	1	Somatic	Missense point	Y76C	FGF	-
MAP3K3	1	Somatic	Missense point	E645K	ST Kinase	-
DIDO1	1	Somatic Somatic	Missense point Missense point	R303Q E301Q	PHD	COSMIC -
RASA1	1	Somatic	Stop Gain	W490X	PH	-
SOS2	1	Somatic	Missense point	S1049L	-	-

Representative gel images and electropherograms are shown (Fig. 4.13, Fig. 4.14)

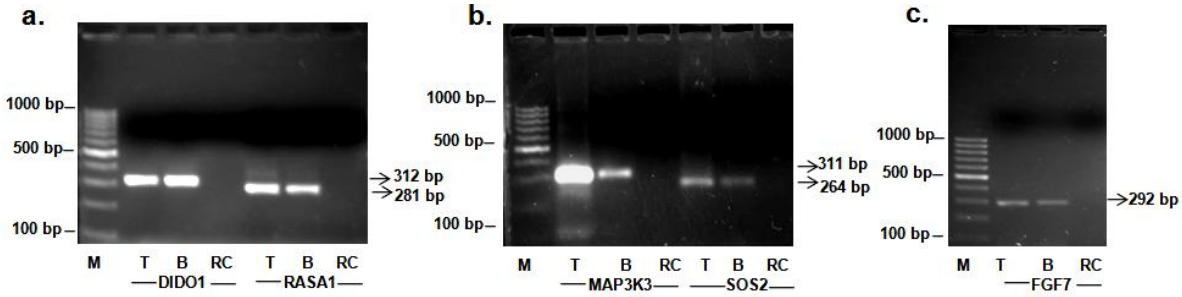

Fig. 4.7: Validation of exome sequencing data from dataset-II by PCR. *Representative gel images for validation of dataset-II are depicted. PCR was carried out with blood and tissue DNA using specific primers. The amplified PCR products were purified and subjected to Sanger sequencing for identification of specific mutation. T= Tissue, B= Blood, RC=reagent control, M= molecular weight marker*

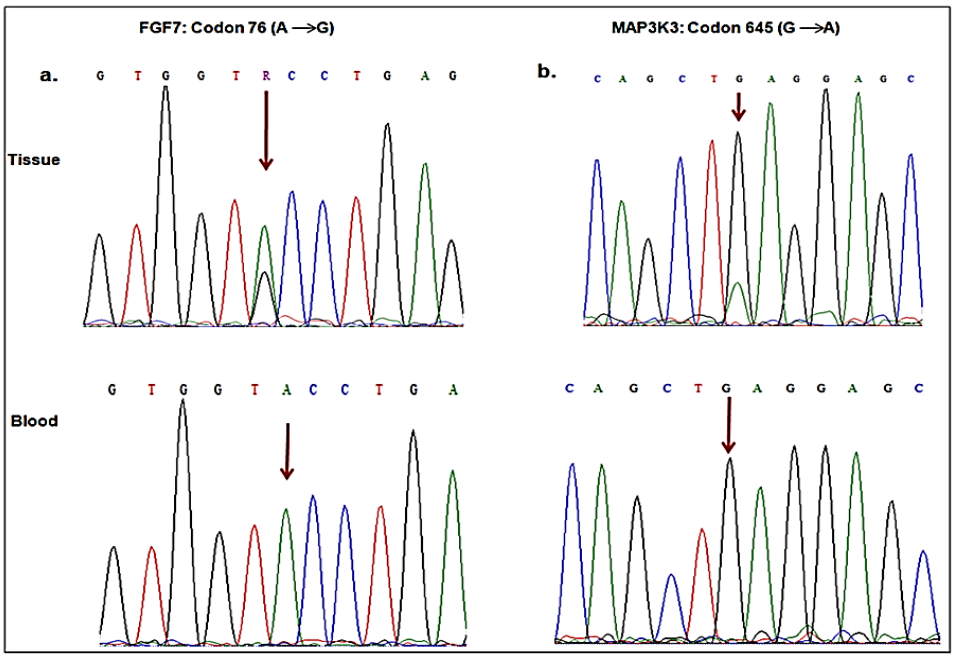

Fig. 4.11: Electropherograms for somatic mutations from Dataset-II. *Missense mutation in FGF7 and MAP3K3 identified from dataset-2 was confirmed by Sanger sequencing. The respective variations were present in the tumour but absent in the corresponding blood samples, indicating it to be somatic.*

All these candidate genes belong to important signalling pathways and are known to be involved in a number of cancers. Validation of the other important candidates from this set is ongoing in the lab.

4.4 DISCUSSION

Although HPV is the major causative factor for development of cervical cancer, the role of several genetic components in the disease has long been recognized. While allelic variation of HLA class II antigen (Hildesheim et al., 1998; Madeleine et al., 2002), and p53 codon 72 polymorphism (Storey et al., 1998) increases the risk of developing the disease, mutations in Ras oncogene, PI3KA, STK11, are reported to directly cause cervical carcinoma (Cui et al., 2009; Wingo et al., 2009; Wong et al., 1995). Apart from this, a number of chromosomal aberrations have also been associated with the disease (Narayan and Murty, 2010).

Whole exome sequencing of 11 tumour and 8 matched blood sample from cervical cancer patients with locally advanced disease (FIGO Stage IIIB), was done in two sets, in order to investigate the genomic landscape of the disease. The technique of whole-exome sequencing has proved to be an effective tool for identification of a wide variety of novel variations especially in the field of Mendelian disorders (Bamshad et al., 2011; Gilissen et al., 2011; Ng et al., 2010; Thorland et al., 2000). Over the last few years whole exome sequencing has been quiet successful in identifying novel mutations in diseases characterized by marked genetic heterogeneity like cancer including rare forms (Byun et al., 2010; Harbour et al., 2010; Yoshida et al., 2011). However, to the best of our knowledge, this is the first study where it has been applied for identifying genetic alterations (mainly single nucleotide variations) in cervical cancer in Indian women.

A significant step for short-listing the single nucleotide variations was based on prediction of the effect of the variations that might range from negligible to severe. Several programs have been developed in the past decade which can predict functional significance of nonsynonymous variants. We have used two such tools – SIFT and

Polyphen-2 to verify the effect of a particular alteration. SIFT is based on sequence homology and physicochemical properties (Kumar et al., 2009); and Polyphen is based on number of features comprising the sequence, phylogenetic and structural information characterizing the substitution (Adzhubei et al., 2010). Jaffe *et al* compared 4 programmes – SIFT, MAPP, Polyphen-2 and VarioWatch and concluded that a conservative approach such as SIFT should be used to minimize false positives and a less conservative approach such as Polyphen could be used to capture all possible deleterious variants although. (Jaffe et al., 2012). The overall agreement between SIFT and Polyphen-2 was found to be 64.2%. For dataset-I, SIFT and Polyphen-2 concurrence was only 35%, while for dataset-II, it increased to 56%.

From dataset-I, single nucleotide variations in genes such as RNASEL (R462Q), PTPRJ (R1222H), ZFP64 (Y210X), FBN1 (del AGG), SREBF1 (del AGG) and BANF1 (del CACTGAG) could be successfully validated by Sanger sequencing. These variations, although eventually found to be germline, might represent germline mutations playing a role in genetic predisposition towards developing cervical cancer. For instance, RNASEL (R462Q) variation has been shown to be associated with prostate cancer susceptibility, and germline mutation with that of cervical cancer (Madsen et al., 2008). From dataset-II, point mutation of FGF7 (Y76C) and MAP3K3 (E645K), DIDO1 (R303Q and E301Q), RASA1 (W490X) and SOS2 (S1049L) were identified and validated by Sanger sequencing as true somatic. All these genes are known to have implications in several cancers. Also, an identical mutation of DIDO1 (R303Q) has been reported in breast carcinoma (Stephens et al., 2012). However, to the best of our knowledge there are no reports of these two mutations in cervical carcinoma. Further functional studies of these mutations might provide a complete new insight into the disease and add on significant information for understanding the mutational landscape

of cervical carcinoma.. Apart from this, there were several variations including both known and novel that were common between the two sets of runs. However, before arriving at any conclusion regarding the role of these mutations in cervical carcinogenesis, their presence should be checked in blood from healthy controls, so as to confirm the true somatic nature of these variations. The process of further validating some of these variations as well as those obtained exclusively from dataset-2 of sequencing runs are still ongoing in the lab.

Although, Array Comparative Genomic Hybridization (CGH) and SNP genotyping arrays are beyond doubt the standard methods for CNV and LOH detection, with rapid growth in genomic technologies over the past few years, using next generation sequencing data for the same, are becoming popular. However, as opposed to whole genome sequencing, the detection of CNV and LOH from whole exome sequencing data is quiet challenging, and till date there are very few studies in this context. This is mainly due to the fact that almost all the algorithms for detecting CNV and LOH assumes random read distribution that fails to hold in the context of exome capture as the specificity as well as efficiency for the targeted exonic regions varies between the probes. 'ExomeCNV' package as described by Sathirapongsasuti et al., which uses depth-of-coverage and B-allele frequencies from mapped short sequence reads is perhaps the only tool that talks about prediction of copy number variation and loss of heterozygosity from exome sequencing data (Sathirapongsasuti et al., 2011). However, there are still issues regarding its resolution; also the Exome CNV does not perform well when coverage levels between the case and control differ significantly. Nevertheless, using 'Exome CNV' for the two paired samples from the dataset-1 (where the coverage between blood and tumour was comparable), copy number alteration in chromosomal regions 1q, 3q, 8q, 11q, 17q and Xq and putative loss of heterozygosity in

2q, 3p, and 11q was predicted. Although the detection for LOH was not so precise, some of our CNV results such as those in 3q, 8q, 17q12, Xq28, corroborated with published reports (Narayan and Murty, 2010). Also, chromosomal gain in 17q12 was validated by qRT-PCR, justifying the prediction by 'ExomeCNV' package.

Apart from identification of a number of germline as well as somatic mutations that might play an important role in cervical carcinogenesis, a major highlight of this part of our study is that it provided an insight into the genome of Indian individuals by identifying a number of novel variations which could enrich the existing databases. Besides, the study marked the beginning of understanding the genomic landscape of cervical carcinogenesis.

4.5 CONCLUSION

Overall, the exome sequencing data from 8 paired cervical tumour biopsies (data from 3 unpaired samples was not considered for shortlisting somatic variations) in 2 sets identified novel germline mutations / polymorphisms in PTPRJ, ZFP64, SREBF1, FBN1 and BANF1, as well as somatic mutations in genes such as FGF7, MAP3K3, RASA1, DIDO1 and SOS2. These variations might either represent predisposing alterations or actually 'drive' disease development. Besides, copy number variation in 1q, 3q, 8q, 11q, 17q and Xq and putative loss of heterozygosity in 2q, 3p, and 11q was predicted from the exome data by 'ExomeCNV' package. A large number of recurrent, novel genetic variations have also been identified although they need to be validated by Sanger sequencing (Appendix II). The study is still ongoing in the laboratory.

Chapter– 5

General Discussion

&

Study Perspective

Cancer of the cervix is estimated to affect approximately 530000 women each year worldwide with >88% of the burden being felt in the developing countries. In India it is the most common cancer as well as the major cause of cancer deaths among women (GLOBOCAN, 2008b; IARC, 2009). The disease is usually preceded by a well-defined and long pre-malignant phase, characterized by a series of microscopic events progressing from cellular atypia to various grades of dysplasia or cervical intraepithelial neoplasia (CIN) and eventually invasive carcinoma (Sellors and Sankaranarayanan, 2003) . Persistent infection with high risk Human Papillomavirus (HR-HPV) is almost the universal etiological factor associated with the disease. The virus infects the dividing cells of the epithelial basal layer, continues to divide and proliferate in the upper epithelial layers, ultimately driving the cells towards malignant progression (Doorbar, 2006). The viral oncogenes - E6 and E7, are responsible for bringing about this transformation via interfering mainly with the cell cycle regulators p53 and Rb respectively. Several viral co-factors such as infecting HPV type, viral integration status and viral load determine the fate of virus- host interaction and play a key role in disease progression and predicting patient prognosis. Although, HPV infection plays a major role, it might not be sufficient for disease development (Cui et al., 2009). The evidence for this comes from the observation that not all HPV positive women develop the disease; also the window period between acquiring an infection and showing the symptoms is very long. All these hint towards a possible involvement of genetic factors in genesis of cervical cancer. Over the years, studies have reported role of SNPs, several point mutations, insertions, deletions and chromosomal abnormalities in cervical carcinogenesis.

Through this study we have attempted to address all the major aspects involved in the pathogenesis of cervical cancer such as HPV infection, including the role of various

viral cofactors such as genotype, physical state of the virus, site of viral integration and viral load; as well as, genetic alterations that might either contribute directly or predispose an individual to disease development. Such approaches would help in better understanding the fundamental rules of cervical carcinogenesis and be a step forward towards identifying biomarkers and newer treatment modalities for management and cure of the disease.

Role of HPV in cervical carcinoma

Reports from across the globe indicate a prevalence of HPV ranging from 73–99% in cervical carcinoma, with the highest incidences being reported from the Indian subcontinent (Basu et al., 2009; Bhatla et al., 2008; de Sanjose et al., 2010; Franceschi et al., 2003; Grace Nirmala and Narendhirakannan, 2012; Li et al., 2011). Of all the HPV types infection by HPV16 is most common, generally followed by HPV18, 31, 33 and 45 (Basu et al., 2011; de Sanjose et al., 2010; Grace Nirmala and Narendhirakannan, 2012; Li et al., 2011; Pillai et al., 2010). In the present study, the incidence of HPV infection (95%) as well as the prevalence of the two HR-HPV types – HPV 16 and/or 18 (72%) in the 270 pre-treatment cervical cancer biopsies from Indian women, with locally advanced stage of the disease and undergoing radiation therapy was in agreement with the published reports. The study, therefore, might add to the existing reports on HPV detection and genotyping from the Indian subcontinent and contribute to the projected prophylactic vaccine trials in the country.

Reports on prognostic value of HPV genotypes indicate a strong correlation between infecting HPV types and response of patients to radiation therapy (Wang et al., 2010). In the present study, clinical outcome of the patients infected with the two most common HR-HPV types HPV16 and/or 18 revealed that HPV16 positive cases showed a

trend towards better disease free survival following radiation therapy, as compared to HPV18 or dual infection by HPV16 and 18. However, this data was not statistically significant, probably owing to the small number of cases in the HPV18 positive group.

Apart from particular HPV genotypes, the physical status of the virus plays a major role in HPV induced cervical carcinogenesis. Although viral integration has long been associated with progression of the disease (Durst et al., 1985; Kalantari et al., 2001; Klaes et al., 1999), some recent reports have confirmed presence of episomal form alone in advanced cervical squamous cell carcinomas, thereby establishing that integration might not be absolutely mandatory for the process of carcinogenesis (Gray et al., 2010; Vinokurova et al., 2008). In the present study, only episomal form of HPV was detected in 18 out of 86 cases of advanced cervical carcinoma where the physical status of the virus was studied by APOT assay, further affirming these recent reports.

Reports regarding the significance of viral physical status in disease prognosis have been quiet contradictory. While studies by Kalantari et al and Vernon et al. have demonstrated a decrease in disease free survival with viral integration (Kalantari et al., 1998; Vernon et al., 1997), according to some recent studies physical state of the virus bear no correlation with disease prognosis (Holm et al., 2008; Nambaru et al., 2009). In our study we found a significant association between viral physical status with disease outcome, the episomal form being associated with an increased disease free survival as compared to the integrated one.

Contrary to most of the earlier reports that considered integration to be a random event (Klaes et al., 1999; Vinokurova et al., 2008), our study identified certain chromosomal regions to be frequently affected by integration. Two recent reports have identified 3q28, 4q13.3, 8q24.21 and 13q22.1 as preferred sites for viral integration (Kraus et al., 2008; Schmitz et al., 2012) which support our findings. Our observation

that regions within or near known genes and chromosomal fragile sites might represent 'hotspots' for viral integration was in agreement with the published reports (Kraus et al., 2008; Matovina et al., 2009; Wentzensen et al., 2004). Comparison of the site of viral integration with clinical outcome revealed that not only the physical status of the virus, but also the site of integration is important in predicting the disease prognosis, since patients with integration in the chromosomal loci 3q, 13q and 20q showed an increased recurrence of the disease as compared to others. To the best of our knowledge this is the first report where the site of viral integration into the chromosomal loci has been shown to be an important factor in disease prognosis.

In order to have a clearer picture about disease prognosis, the study of viral integration was further complemented with estimation of viral load in the same 86 samples. Although several studies have identified viral load as an important marker for disease progression (Dalstein et al., 2003; Gravitt et al., 2007; Hernandez-Hernandez et al., 2003), none of these studies have taken into account the physical state of the virus. In the present study, association of viral load with survival of the patients as well as physical status revealed that, copy number of the virus combined with the physical state might serve as a good prognostic marker for the disease.

Role of Genetic Alterations in cervical carcinoma

Although mutations in important candidate genes such as Ras, PI3KA, STK11 (Cui et al., 2009; Madsen et al., 2008; Wingo et al., 2009; Wong et al., 1995) as well as several chromosomal aberrations (Narayan and Murty, 2010) have been associated with development of cervical carcinoma, till date there are no reports describing the genomic landscape of the disease. Therefore, in order to have a better understanding of the fundamental rules of the disease, whole exome sequencing of

DNA from 11 pre-treatment tumour biopsies (all of stage IIIB and >70% tumour) and 8 matched blood was done in 2 sets (dataset- I and II). Although initially SNP array was the method of choice for high throughput SNP analysis, with the availability of the far more developed and informative NGS technology at hand, we switched over to whole exome sequencing for the identification of the genetic alterations in cervical cancer. The sequencing data provided significant information about the novel variations in the Indian genome, the representation of which is inadequate in dbSNP. Therefore, our study could have immense contribution towards enriching the dbSNP with respect to the coding regions, in the Indian population.

Some of the germline variations in important candidates such as RNASEL, PTPRJ, ZFP64, FBN1 were identified from dataset-I and confirmed by Sanger sequencing, while somatic mutations in FGF7, MAP3K3, DIDO1, SOS2 and RASA1 were identified and confirmed from dataset-II. These might have significant implication in the context of carcinoma of the cervix. The germline variations could represent important predisposing mutations, while the somatic ones might directly be responsible for disease development. Studies have reported germline mutations of RNASEL including R462Q to be associated with prostate as well as cervical cancer susceptibility, (Barbisan et al., 2011; Casey et al., 2002; Madsen et al., 2008). Also, several missense mutations of FGF7, MAP3K3, RASA1 and SOS2 have been reported in the COSMIC database (http://www.sanger.ac.uk/genetics/CGP/cosmic/) for cancers of the breast, lung, kidney, ovary, prostrate, colon, etc. Besides, a recent study has identified exact identical mutation in DIDO1 (R303Q) in breast carcinoma (Stephens et al., 2012). To the best of our knowledge this is the first report stating the association of

mutations of these important proteins in cervical cancer. However, whether these mutations are actually 'driver' or 'passenger', would be really interesting and challenging to study.

The exome sequencing data was also used successfully to some extent to predict copy number variation and LOH that might be significant in cervical carcinogenesis. Overall, our study is the first demonstration of the application of whole exome sequencing in understanding the genomics of cervical carcinoma with special reference to the Indian population.

The overall work presented in the thesis is shown in Fig. 5.1.

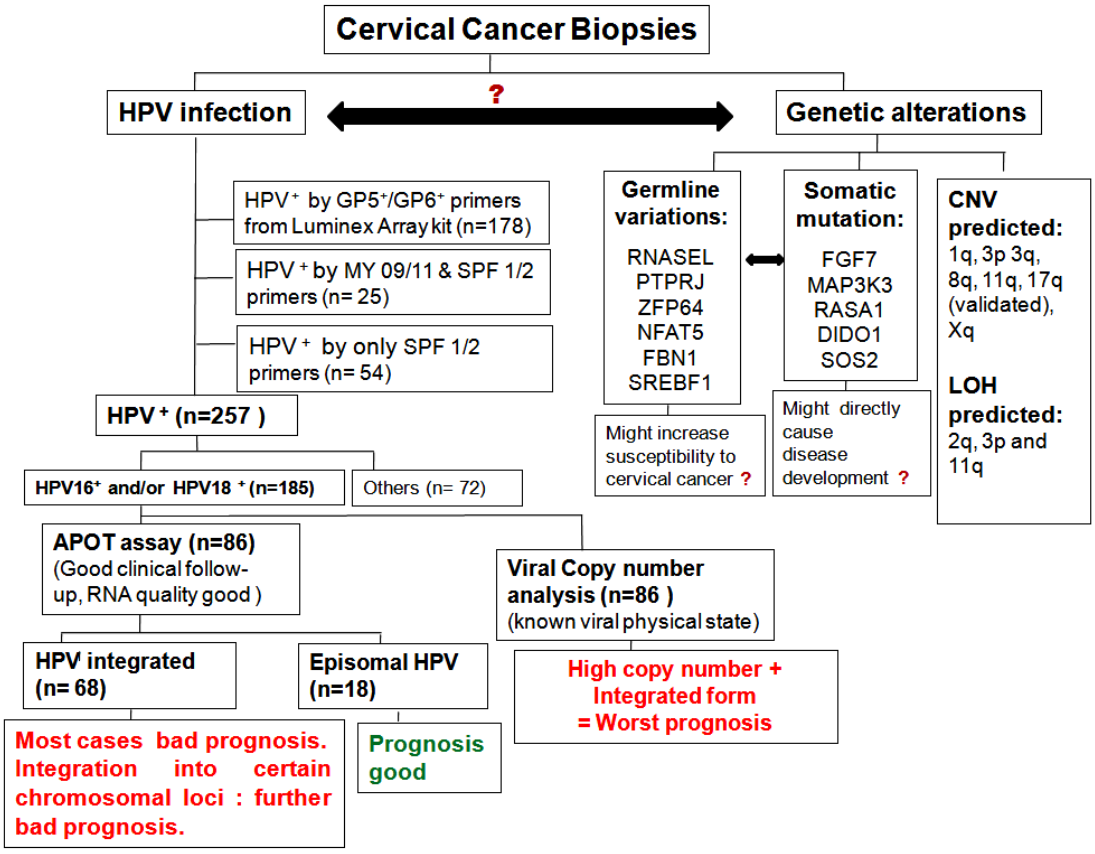

Fig. 5.1 Overall summary of the study

Chapter– 6

Summary & Conclusion

Summary

The study focuses on two main factors associated with the pathogenesis of cervical cancer – HPV infection and genetic alterations. The study was carried out on pretreatment cervical cancer biopsies from a cohort of Indian women with locally advanced stage of the disease. The major highlights are as follows:

- Incidence of HPV in 270 cervical cancer biopsies as determined by 3 primer sets – $GP5^{+}/6^{+}$, MY09/11 and SPF1/2, was as high as 95% (257/270).
- Genotyping of the samples positive for 24 HPV types using $GP5^{+}/6^{+}$ primers (n=178), by Luminex bead array, showed most samples (168/178) to be positive for one or more HR-HPV types indicating a high association of cervical cancer with HR-HPV infection. Of the 24 HPV genotypes included in the HPV Genotyping Kit, infection by HPV 11, 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 73 were identified. HPV16 and/or HPV18 infection was most common - 114 samples being positive for HPV16 alone and 6 samples for HPV18 alone. This was followed by dual infection by HPV16/18 (n=12) and 16/45 (n=9).
- Genotyping of the remaining 92 samples (which were negative using GP primers), showed 49 samples to be positive for HPV16 and none for HPV18. Therefore, the overall prevalence of HPV16 and/or HPV18 in this cohort was 72%, whereas overall HPV positivity was 95%.
- Patients positive for HPV16 alone showed a trend for better survival following radiotherapy as compared to HPV18 alone and dual infection by HPV16/18.
- Physical state of the virus as well site of viral integration was determined in a subset of cases (n=86) positive for HPV16 and/or HPV18 by APOT assay. Integration was observed in 79% cases (n =68), whereas in 21% (n =18) only episomal transcripts were identified. In 12 cases both integrated and episomal form of HPV were detected.

- Comparison with clinical outcome revealed 16 out of 18 patients with only episomal form of HPV (16 and/or 18), to have disease free survival as compared to those with integrated form (p= 0.067).
- The site of viral integration could be predicted with a high score in 48 cases and was found to be more frequent at chromosomal loci 1p (n =7), 3q (n= 8), 13q (n= 4), 6q (n =4), 11q (n= 4) and 20q (n= 4). Only one sample showed HPV integration at two chromosomal loci simultaneously.
- Comparison with clinical outcome revealed that patients with integration at loci 1p (7/7), 6q (4/4) and 11q (4/4) were disease free, while most of them with integration at loci 3q (5/8), 13q (4/4) and 20q (2/4) showed recurrence of the disease in the form of either loco-regional or distant metastasis.
- With NCBI Fragile site Map Viewer and UCSC Blat tool, most of the integration sites were found to be within or close to a fragile site (29/48) and/or protein coding genes (28/48).
- Estimation of viral load in the same 86 samples by SYBR green based relative qRT-PCR, identified a wide range of viral titre across the samples, with the average being 47.5.
- High viral load combined with viral integration showed poor prognosis; thus viral load and state of the virus can serve as a good prognostic marker of the disease.
- Whole exome sequencing study of 11 cervical tumour and 8 matched control (blood from the same individual), in two sets (Dataset-I and II) identified a number of novel, nonsynonymous variations, that were validated by Sanger sequencing and/or customized Agilent SNP array. These might represent SNPs characteristic to the Indian genome and could add on to the existing dbSNP.

- While germline variation of RNASEL, PTPRJ, ZFP64, FBN1, SREB1 and BANF1, identified from this data and validated by Sanger sequencing, might represent predisposing mutations to cervical cancer, missense mutations in FGF7, SOS2, DIDO1 and MAP3K3 as well as nonsense mutation in RASA1, might act as potential disease causing mutations.
- Copy number alterations in chromosomal regions 1q, 3p, 3q, 8q, 11q, 17q and Xq were predicted.
- Putative loss of heterozygosity in the chromosomal regions 2q, 3p and 11q was also predicted by 'ExomeCNV' package from the whole exome sequencing data.

Conclusion

The significance of the study lies in the fact that we have addressed all the major factors associated with pathogenesis of cervical cancer and assessed the role of each. The findings related to several viral cofactors – incidence, genotype, integration and viral load gives a detailed overview of the importance of each individually and also in relation to one another, in the context of cervical carcinogenesis. Through this study, we have showed for the first time, association between site of viral integration in the host genome and disease prognosis. Also, the combined effect of viral load and physical status on disease prognosis was demonstrated for the first time in the Indian cohort. The whole exome sequencing data from cervical cancer patients reported here, besides being a source of information on novel variations in the Indian genome, would mark the beginning of understanding the genomic landscape of cervical cancer.

Bibliography

A, E.L., Hamoudi, R.A., Kocjan, G., and Du, M.Q. (2004). Cervical intraepithelial neoplasia: prognosis by combined LOH analysis of multiple loci. Gynecol Oncol *94*, 671-679.

Abba, M.C., Mouron, S.A., Gomez, M.A., Dulout, F.N., and Golijow, C.D. (2003). Association of human papillomavirus viral load with HPV16 and high-grade intraepithelial lesion. Int J Gynecol Cancer *13*, 154-158.

Abban, C.Y., and Meneses, P.I. (2010). Usage of heparan sulfate, integrins, and FAK in HPV16 infection. Virology *403*, 1-16.

Acevedo, C.M., Henriquez, M., Emmert-Buck, M.R., and Chuaqui, R.F. (2002). Loss of heterozygosity on chromosome arms 3p and 6q in microdissected adenocarcinomas of the uterine cervix and adenocarcinoma in situ. Cancer *94*, 793-802.

Adler, K., Erickson, T., and Bobrow, M. (1997). High sensitivity detection of HPV-16 in SiHa and CaSki cells utilizing FISH enhanced by TSA. Histochem Cell Biol *108*, 321-324.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.

Agrawal, N., Frederick, M.J., Pickering, C.R., Bettegowda, C., Chang, K., Li, R.J., Fakhry, C., Xie, T.X., Zhang, J., Wang, J*., et al.* (2011). Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. Science *333*, 1154-1157.

Andersson, S., Hellstrom, A.C., Ren, Z.P., and Wilander, E. (2006). The carcinogenic role of oncogenic HPV and p53 gene mutation in cervical adenocarcinomas. Med Oncol *23*, 113-119.

Arias-Pulido, H., Peyton, C.L., Joste, N.E., Vargas, H., and Wheeler, C.M. (2006). Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. J Clin Microbiol *44*, 1755-1762.

Baay, M.F., Quint, W.G., Koudstaal, J., Hollema, H., Duk, J.M., Burger, M.P., Stolz, E., and Herbrink, P. (1996). Comprehensive study of several general and type-specific primer pairs for detection of human papillomavirus DNA by PCR in paraffin-embedded cervical carcinomas. J Clin Microbiol *34*, 745-747.

Baker, C.C., Phelps, W.C., Lindgren, V., Braun, M.J., Gonda, M.A., and Howley, P.M. (1987). Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. J Virol *61*, 962-971.

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet *12*, 745-755.

Barbisan, G., Contreras, A., Perez, L.O., Difranza, L., and Golijow, C.D. (2011). The effect of TP53 codon 72 and RNASEL codon 462 polymorphisms on the development of cervical cancer in Argentine women. Cancer Genet *204*, 270-277.

Basu, P., Chandna, P., Bamezai, R.N., Siddiqi, M., Saranath, D., Lear, A., and Ratnam, S. (2011). MassARRAY spectrometry is more sensitive than PreTect HPV-Proofer and consensus PCR for type-specific detection of high-risk oncogenic human papillomavirus genotypes in cervical cancer. J Clin Microbiol *49*, 3537-3544.

Basu, P., Roychowdhury, S., Bafna, U.D., Chaudhury, S., Kothari, S., Sekhon, R., Saranath, D., Biswas, S., Gronn, P., Silva, I*., et al.* (2009). Human papillomavirus genotype distribution in cervical cancer in India: results from a multi-center study. Asian Pac J Cancer Prev *10*, 27-34.

Berkhout, R.J., Tieben, L.M., Smits, H.L., Bavinck, J.N., Vermeer, B.J., and ter Schegget, J. (1995). Nested PCR approach for detection and typing of epidermodysplasia verruciformis-associated

human papillomavirus types in cutaneous cancers from renal transplant recipients. J Clin Microbiol *33*, 690-695.

Bernard, H.U., Burk, R.D., Chen, Z., van Doorslaer, K., Hausen, H., and de Villiers, E.M. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology *401*, 70-79.

Bethwaite, P.B., Koreth, J., Herrington, C.S., and McGee, J.O. (1995). Loss of heterozygosity occurs at the D11S29 locus on chromosome 11q23 in invasive cervical carcinoma. Br J Cancer *71*, 814-818.

Bhatla, N., Dar, L., Patro, A.R., Kriplani, A., Gulati, A., Verma, K., Broor, S., Shah, K.V., and Gravitt, P.E. (2006). Human papillomavirus type distribution in cervical cancer in Delhi, India. Int J Gynecol Pathol *25*, 398-402.

Bhatla, N., Lal, N., Bao, Y.P., Ng, T., and Qiao, Y.L. (2008). A meta-analysis of human papillomavirus type-distribution in women from South Asia: implications for vaccination. Vaccine *26*, 2811-2817.

Bhattacharya, N., Singh, R.K., Mondal, S., Roy, A., Mondal, R., Roychowdhury, S., and Panda, C.K. (2004). Analysis of molecular alterations in chromosome 8 associated with the development of uterine cervical carcinoma of Indian patients. Gynecol Oncol *95*, 352-362.

Bosch, F.X., and Munoz, N. (2002). The viral etiology of cervical cancer. Virus Res *89*, 183-190.

Botstein, D., and Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet *33 Suppl*, 228-237.

Boyer, S.N., Wazer, D.E., and Band, V. (1996). E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. Cancer Res *56*, 4620-4624.

Buckley, C.H., Butler, E.B., and Fox, H. (1982). Cervical intraepithelial neoplasia. J Clin Pathol *35*, 1-13.

Busby-Earle, R.M., Steel, C.M., Williams, A.R., Cohen, B., and Bird, C.C. (1994). p53 mutations in cervical carcinogenesis--low frequency and lack of correlation with human papillomavirus status. Br J Cancer *69*, 732-737.

Byun, M., Abhyankar, A., Lelarge, V., Plancoulaine, S., Palanduz, A., Telhan, L., Boisson, B., Picard, C., Dewell, S., Zhao, C.*, et al.* (2010). Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. J Exp Med *207*, 2307-2312.

Calin, G.A., and Croce, C.M. (2006). MicroRNAs and chromosomal abnormalities in cancer cells. Oncogene *25*, 6202-6210.

Casey, G., Neville, P.J., Plummer, S.J., Xiang, Y., Krumroy, L.M., Klein, E.A., Catalona, W.J., Nupponen, N., Carpten, J.D., Trent, J.M.*, et al.* (2002). RNASEL Arg462Gln variant is implicated in up to 13% of prostate cancer cases. Nat Genet *32*, 581-583.

Chan, S.Y., Delius, H., Halpern, A.L., and Bernard, H.U. (1995). Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. J Virol *69*, 3074-3083.

Chao, A., Wang, T.H., Lee, Y.S., Hsueh, S., Chao, A.S., Chang, T.C., Kung, W.H., Huang, S.L., Chao, F.Y., Wei, M.L.*, et al.* (2006). Molecular characterization of adenocarcinoma and squamous carcinoma of the uterine cervix using microarray analysis of gene expression. Int J Cancer *119*, 91-98.

Chatterjee, R., Mandal, B., and Bandyopadhyay, S. (2003). Detection of HPV DNA in cervical carcinomas by PCR and hybrid capture assay. Indian J Pathol Microbiol *46*, 596-599.

Choi, C.H., Lee, K.M., Choi, J.J., Kim, T.J., Kim, W.Y., Lee, J.W., Lee, S.J., Lee, J.H., Bae, D.S., and Kim, B.G. (2007). Hypermethylation and loss of heterozygosity of tumor suppressor genes on chromosome 3p in cervical cancer. Cancer Lett *255*, 26-33.

Choo, K.B., Pan, C.C., and Han, S.H. (1987). Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. Virology *161*, 259-261.

Chuaqui, R., Silva, M., and Emmert-Buck, M. (2001). Allelic deletion mapping on chromosome 6q and X chromosome inactivation clonality patterns in cervical intraepithelial neoplasia and invasive carcinoma. Gynecol Oncol *80*, 364-371.

Ciapponi, A., Bardach, A., Glujovsky, D., Gibbons, L., and Picconi, M.A. (2011). Type-specific HPV prevalence in cervical cancer and high-grade lesions in Latin America and the Caribbean: systematic review and meta-analysis. PLoS One *6*, e25493.

Clifford, G.M., Rana, R.K., Franceschi, S., Smith, J.S., Gough, G., and Pimenta, J.M. (2005). Human papillomavirus genotype distribution in low-grade cervical lesions: comparison by geographic region and with cervical cancer. Cancer Epidemiol Biomarkers Prev *14*, 1157-1164.

Clifford, G.M., Smith, J.S., Plummer, M., Munoz, N., and Franceschi, S. (2003). Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. Br J Cancer *88*, 63-73.

Contag, S.A., Gostout, B.S., Clayton, A.C., Dixon, M.H., McGovern, R.M., and Calhoun, E.S. (2004). Comparison of gene expression in squamous cell carcinoma and adenocarcinoma of the uterine cervix. Gynecol Oncol *95*, 610-617.

Corden, S.A., Sant-Cassia, L.J., Easton, A.J., and Morris, A.G. (1999). The integration of HPV-18 DNA in cervical carcinoma. Mol Pathol *52*, 275-282.

Couturier, J., Sastre-Garau, X., Schneider-Maunoury, S., Labib, A., and Orth, G. (1991). Integration of papillomavirus DNA near myc genes in genital carcinomas and its consequences for proto-oncogene expression. J Virol *65*, 4534-4538.

Crook, T., Wrede, D., Tidy, J.A., Mason, W.P., Evans, D.J., and Vousden, K.H. (1992). Clonal p53 mutation in primary cervical cancer: association with human-papillomavirus-negative tumours. Lancet *339*, 1070-1073.

Crum, C.P., Abbott, D.W., and Quade, B.J. (2003). Cervical cancer screening: from the Papanicolaou smear to the vaccine era. J Clin Oncol *21*, 224s-230s.

Cui, B., Zheng, B., Zhang, X., Stendahl, U., Andersson, S., and Wallin, K.L. (2009). Mutation of PIK3CA: possible risk factor for cervical carcinogenesis in older women. Int J Oncol *34*, 409-416.

Cullen, A.P., Reid, R., Campion, M., and Lorincz, A.T. (1991). Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. J Virol *65*, 606-612.

Culp, T.D., Budgeon, L.R., Marinkovich, M.P., Meneguzzi, G., and Christensen, N.D. (2006). Keratinocyte-secreted laminin 5 can function as a transient receptor for human papillomaviruses by binding virions and transferring them to adjacent cells. J Virol *80*, 8940-8950.

Dall, K.L., Scarpini, C.G., Roberts, I., Winder, D.M., Stanley, M.A., Muralidhar, B., Herdman, M.T., Pett, M.R., and Coleman, N. (2008). Characterization of naturally occurring HPV16

integration sites isolated from cervical keratinocytes under noncompetitive conditions. Cancer Res *68*, 8249-8259.

Dalstein, V., Riethmuller, D., Pretet, J.L., Le Bail Carval, K., Sautiere, J.L., Carbillet, J.P., Kantelip, B., Schaal, J.P., and Mougin, C. (2003). Persistence and load of high-risk HPV are predictors for development of high-grade cervical lesions: a longitudinal French cohort study. Int J Cancer *106*, 396-403.

Das, P., Thomas, A., Mahantshetty, U., Shrivastava, S.K., Deodhar, K., and Mulherkar, R. (2012). HPV Genotyping and Site of Viral Integration in Cervical Cancers in Indian Women. PLoS One *7*, e41012.

Day, P.M., Baker, C.C., Lowy, D.R., and Schiller, J.T. (2004). Establishment of papillomavirus infection is enhanced by promyelocytic leukemia protein (PML) expression. Proc Natl Acad Sci U S A *101*, 14252-14257.

de Boer, M.A., Jordanova, E.S., Kenter, G.G., Peters, A.A., Corver, W.E., Trimbos, J.B., and Fleuren, G.J. (2007). High human papillomavirus oncogene mRNA expression and not viral DNA load is associated with poor prognosis in cervical cancer patients. Clin Cancer Res *13*, 132-138.

de Sanjose, S., Quint, W.G., Alemany, L., Geraets, D.T., Klaustermeier, J.E., Lloveras, B., Tous, S., Felix, A., Bravo, L.E., Shin, H.R.*, et al.* (2010). Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol *11*, 1048-1056.

de Villiers, E.M., Gissmann, L., and zur Hausen, H. (1981). Molecular cloning of viral DNA from human genital warts. J Virol *40*, 932-935.

Deodhar, K., Gheit, T., Vaccarella, S., Romao, C.C., Tenet, V., Nene, B.M., Jayant, K., Kelkar, R., Malvi, S.G., Sylla, B.S.*, et al.* (2012). Prevalence of human papillomavirus types in cervical lesions from women in rural Western India. J Med Virol *84*, 1054-1060.

Di Luca, D., Pilotti, S., Stefanon, B., Rotola, A., Monini, P., Tognon, M., De Palo, G., Rilke, F., and Cassai, E. (1986). Human papillomavirus type 16 DNA in genital tumours: a pathological and molecular analysis. J Gen Virol *67 (Pt 3)*, 583-589.

Dikshit, R., Gupta, P.C., Ramasundarahettige, C., Gajalakshmi, V., Aleksandrowicz, L., Badwe, R., Kumar, R., Roy, S., Suraweera, W., Bray, F.*, et al.* (2012). Cancer mortality in India: a nationally representative survey. Lancet *379*, 1807-1816.

Dong, S.M., Kim, H.S., Rha, S.H., and Sidransky, D. (2001). Promoter hypermethylation of multiple genes in carcinoma of the uterine cervix. Clin Cancer Res *7*, 1982-1986.

Doorbar, J. (2006). Molecular biology of human papillomavirus infection and cervical cancer. Clin Sci (Lond) *110*, 525-541.

Dowhanick, J.J., McBride, A.A., and Howley, P.M. (1995). Suppression of cellular proliferation by the papillomavirus E2 protein. J Virol *69*, 7791-7799.

Duensing, S., and Munger, K. (2001). Centrosome abnormalities, genomic instability and carcinogenic progression. Biochim Biophys Acta *1471*, M81-88.

Duk, J.M., de Bruijn, H.W., Groenier, K.H., Hollema, H., ten Hoor, K.A., Krans, M., and Aalders, J.G. (1990). Cancer of the uterine cervix: sensitivity and specificity of serum squamous cell carcinoma antigen determinations. Gynecol Oncol *39*, 186-194.

Durst, M., Kleinheinz, A., Hotz, M., and Gissmann, L. (1985). The physical state of human papillomavirus type 16 DNA in benign and malignant genital tumours. J Gen Virol *66 (Pt 7)*, 1515-1522.

Evans, M.F., and Cooper, K. (2004). Human papillomavirus integration: detection by in situ hybridization and potential clinical application. J Pathol *202*, 1-4.

Farghaly, S.A. (1992). Tumor markers in gynecologic cancer. Gynecol Obstet Invest *34*, 65-72.

Ferber, M.J., Montoya, D.P., Yu, C., Aderca, I., McGee, A., Thorland, E.C., Nagorney, D.M., Gostout, B.S., Burgart, L.J., Boix, L*., et al.* (2003a). Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. Oncogene *22*, 3813-3820.

Ferber, M.J., Thorland, E.C., Brink, A.A., Rapp, A.K., Phillips, L.A., McGovern, R., Gostout, B.S., Cheung, T.H., Chung, T.K., Fu, W.Y*., et al.* (2003b). Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. Oncogene *22*, 7233-7242.

Flores, E.R., and Lambert, P.F. (1997). Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle. J Virol *71*, 7167-7179.

Forslund, O., Antonsson, A., Nordin, P., Stenquist, B., and Hansson, B.G. (1999). A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. J Gen Virol *80 (Pt 9)*, 2437-2443.

Franceschi, S., Rajkumar, T., Vaccarella, S., Gajalakshmi, V., Sharmila, A., Snijders, P.J., Munoz, N., Meijer, C.J., and Herrero, R. (2003). Human papillomavirus and risk factors for cervical cancer in Chennai, India: a case-control study. Int J Cancer *107*, 127-133.

Francis, D.A., Schmid, S.I., and Howley, P.M. (2000). Repression of the integrated papillomavirus E6/E7 promoter is required for growth suppression of cervical cancer cells. J Virol *74*, 2679-2686.

Gilissen, C., Hoischen, A., Brunner, H.G., and Veltman, J.A. (2011). Unlocking Mendelian disease using exome sequencing. Genome Biol *12*, 228.

Gissmann, L., and zur Hausen, H. (1980). Partial characterization of viral DNA from human genital warts (Condylomata acuminata). Int J Cancer *25*, 605-609.

GLOBOCAN (2008a). Cancer Incidence, Mortality and Prevalence Worldwide in 2008. .

GLOBOCAN (2008b). International Agency for Research on Cancer, World Health

Organisation. Available at http://www-dep.iarc.fr.

Goncalves, A., Fabbro, M., Lhomme, C., Gladieff, L., Extra, J.M., Floquet, A., Chaigneau, L., Carrasco, A.T., and Viens, P. (2008). A phase II trial to evaluate gefitinib as second- or third-line treatment in patients with recurring locoregionally advanced or metastatic cervical cancer. Gynecol Oncol *108*, 42-46.

Grace Nirmala, J., and Narendhirakannan, R.T. (2011). Detection and genotyping of high-risk HPV and evaluation of anti-oxidant status in cervical carcinoma patients in Tamil Nadu State, India--a case control study. Asian Pac J Cancer Prev *12*, 2689-2695.

Grace Nirmala, J., and Narendhirakannan, R.T. (2012). Detection and genotyping of high-risk HPV and evaluation of anti-oxidant status in cervical carcinoma patients in Tamil Nadu State, India--a case control study. Asian Pac J Cancer Prev *12*, 2689-2695.

Gravitt, P.E., Burk, R.D., Lorincz, A., Herrero, R., Hildesheim, A., Sherman, M.E., Bratti, M.C., Rodriguez, A.C., Helzlsouer, K.J., and Schiffman, M. (2003). A comparison between real-time polymerase chain reaction and hybrid capture 2 for human papillomavirus DNA quantitation. Cancer Epidemiol Biomarkers Prev *12*, 477-484.

Gravitt, P.E., Kovacic, M.B., Herrero, R., Schiffman, M., Bratti, C., Hildesheim, A., Morales, J., Alfaro, M., Sherman, M.E., Wacholder, S*., et al.* (2007). High load for most high risk human

papillomavirus genotypes is associated with prevalent cervical cancer precursors but only HPV16 load predicts the development of incident disease. Int J Cancer *121*, 2787-2793.

Gravitt, P.E., Peyton, C.L., Alessi, T.Q., Wheeler, C.M., Coutlee, F., Hildesheim, A., Schiffman, M.H., Scott, D.R., and Apple, R.J. (2000). Improved amplification of genital human papillomaviruses. J Clin Microbiol *38*, 357-361.

Gray, E., Pett, M.R., Ward, D., Winder, D.M., Stanley, M.A., Roberts, I., Scarpini, C.G., and Coleman, N. (2010). In vitro progression of human papillomavirus 16 episome-associated cervical neoplasia displays fundamental similarities to integrant-associated carcinogenesis. Cancer Res *70*, 4081-4091.

Grendys, E.C., Jr., Barnes, W.A., Weitzel, J., Sparkowski, J., and Schlegel, R. (1997). Identification of H, K, and N-ras point mutations in stage IB cervical carcinoma. Gynecol Oncol *65*, 343-347.

Hafner, N., Driesch, C., Gajda, M., Jansen, L., Kirchmayr, R., Runnebaum, I.B., and Durst, M. (2008). Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts. Oncogene *27*, 1610-1617.

Harbour, J.W., Onken, M.D., Roberson, E.D., Duan, S., Cao, L., Worley, L.A., Council, M.L., Matatall, K.A., Helms, C., and Bowcock, A.M. (2010). Frequent mutation of BAP1 in metastasizing uveal melanomas. Science *330*, 1410-1413.

Harima, Y., Sawada, S., Miyazaki, Y., Kin, K., Ishihara, H., Imamura, M., Sougawa, M., Shikata, N., and Ohnishi, T. (2003). Expression of Ku80 in cervical cancer correlates with response to radiotherapy and survival. Am J Clin Oncol *26*, e80-85.

Harima, Y., Sawada, S., Nagata, K., Sougawa, M., Ostapenko, V., and Ohnishi, T. (2001). Mutation of the PTEN gene in advanced cervical cancer correlated with tumor progression and poor outcome after radiotherapy. Int J Oncol *18*, 493-497.

Hart, K.W., Williams, O.M., Thelwell, N., Fiander, A.N., Brown, T., Borysiewicz, L.K., and Gelder, C.M. (2001). Novel method for detection, typing, and quantification of human papillomaviruses in clinical samples. J Clin Microbiol *39*, 3204-3212.

Hernandez-Hernandez, D.M., Ornelas-Bernal, L., Guido-Jimenez, M., Apresa-Garcia, T., Alvarado-Cabrero, I., Salcedo-Vargas, M., Mohar-Betancourt, A., and Garcia-Carranca, A. (2003). Association between high-risk human papillomavirus DNA load and precursor lesions of cervical cancer in Mexican women. Gynecol Oncol *90*, 310-317.

Herrington, C.S., Worsham, M., Southern, S.A., Mackowiak, P., and Wolman, S.R. (2001). Loss of sequences on the short arm of chromosome 17 is a late event in squamous carcinoma of the cervix. Mol Pathol *54*, 160-164.

Heselmeyer, K., Schrock, E., du Manoir, S., Blegen, H., Shah, K., Steinbeck, R., Auer, G., and Ried, T. (1996). Gain of chromosome 3q defines the transition from severe dysplasia to invasive carcinoma of the uterine cervix. Proc Natl Acad Sci U S A *93*, 479-484.

Hidalgo, A., Baudis, M., Petersen, I., Arreola, H., Pina, P., Vazquez-Ortiz, G., Hernandez, D., Gonzalez, J., Lazos, M., Lopez, R.*, et al.* (2005). Microarray comparative genomic hybridization detection of chromosomal imbalances in uterine cervix carcinoma. BMC Cancer *5*, 77.

Hildesheim, A., Schiffman, M., Scott, D.R., Marti, D., Kissner, T., Sherman, M.E., Glass, A.G., Manos, M.M., Lorincz, A.T., Kurman, R.J.*, et al.* (1998). Human leukocyte antigen class I/II alleles and development of human papillomavirus-related cervical neoplasia: results from a case-control study conducted in the United States. Cancer Epidemiol Biomarkers Prev *7*, 1035-1041.

Ho, C.M., Yang, S.S., Chien, T.Y., Huang, S.H., Jeng, C.J., and Chang, S.F. (2005). Detection and quantitation of human papillomavirus type 16, 18 and 52 DNA in the peripheral blood of cervical cancer patients. Gynecol Oncol *99*, 615-621.

Ho, G.Y., Bierman, R., Beardsley, L., Chang, C.J., and Burk, R.D. (1998). Natural history of cervicovaginal papillomavirus infection in young women. N Engl J Med *338*, 423-428.

Holm, R., Kraus, I., Skomedal, H., Langerod, A., Kristensen, G.B., and Lyng, H. (2008). Human papillomavirus DNA and e6/e7 mRNA status in relation to survival of patients treated for cervical squamous cell carcinoma. Open Virol J *2*, 74-81.

Hu, M., Jian, L., Zhang, L., Zheng, J., You, Y., Deng, J., Li, H., and Zhou, Y. (2012). Functional polymorphism in the EpCAM gene is associated with occurrence and advanced disease status of cervical cancer in Chinese population. Mol Biol Rep *39*, 7303-7309.

Huang, G., Xiao, X., Huang, Y., and Huang, R. (1996). [Carcinogenic mechanisms of multiple genes in cervical carcinoma]. Hua Xi Yi Ke Da Xue Xue Bao *27*, 5-9.

Huang, L., Zheng, M., Zhou, Q.M., Zhang, M.Y., Jia, W.H., Yun, J.P., and Wang, H.Y. (2011). Identification of a gene-expression signature for predicting lymph node metastasis in patients with early stage cervical carcinoma. Cancer *117*, 3363-3373.

Huang, L., Zheng, M., Zhou, Q.M., Zhang, M.Y., Yu, Y.H., Yun, J.P., and Wang, H.Y. (2012). Identification of a 7-gene signature that predicts relapse and survival for early stage patients with cervical carcinoma. Med Oncol.

Hwang, E.S., Riese, D.J., 2nd, Settleman, J., Nilson, L.A., Honig, J., Flynn, S., and DiMaio, D. (1993). Inhibition of cervical carcinoma cell line proliferation by the introduction of a bovine papillomavirus regulatory gene. J Virol *67*, 3720-3729.

IARC (2009). "Cervical Cancer", http://screening.iarc.fr/cervicalindex.php,.

Ikenberg, H., Matthay, K., Schmitt, B., Bauknecht, T., Kiechle-Schwarz, M., Goppinger, A., and Pfleiderer, A. (1995). p53 mutation and MDM2 amplification are rare even in human papillomavirus-negative cervical carcinomas. Cancer *76*, 57-66.

Insinga, R.P., Liaw, K.L., Johnson, L.G., and Madeleine, M.M. (2008). A systematic review of the prevalence and attribution of human papillomavirus types among cervical, vaginal, and vulvar precancers and cancers in the United States. Cancer Epidemiol Biomarkers Prev *17*, 1611-1622.

Iorio, M.V., and Croce, C.M. (2009). MicroRNAs in cancer: small molecules with a huge impact. J Clin Oncol *27*, 5848-5856.

Iwakawa, M., Ohno, T., Imadome, K., Nakawatari, M., Ishikawa, K., Sakai, M., Katoh, S., Ishikawa, H., Tsujii, H., and Imai, T. (2007). The radiation-induced cell-death signaling pathway is activated by concurrent use of cisplatin in sequential biopsy specimens from patients with cervical cancer. Cancer Biol Ther *6*, 905-911.

Jacobs, M.V., Snijders, P.J., van den Brule, A.J., Helmerhorst, T.J., Meijer, C.J., and Walboomers, J.M. (1997). A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings. J Clin Microbiol *35*, 791-795.

Jaffe, A., Wojcik, G., Chu, A., Golozar, A., Maroo, A., Duggal, P., and Klein, A.P. (2012). Identification of functional genetic variation in exome sequence analysis. BMC Proc *5 Suppl 9*, S13.

Jenkins, A., Kristiansen, B.E., Ask, E., Oskarsen, B., Kristiansen, E., Lindqvist, B., Trope, C., and Kjorstad, K. (1991). Detection of genital papillomavirus types by polymerase chain reaction using common primers. APMIS *99*, 667-673.

Jeon, S., Allen-Hoffmann, B.L., and Lambert, P.F. (1995). Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. J Virol *69*, 2989-2997.

Jeon, S., and Lambert, P.F. (1995). Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. Proc Natl Acad Sci U S A *92*, 1654-1658.

Josefsson, A.M., Magnusson, P.K., Ylitalo, N., Sorensen, P., Qwarforth-Tubbin, P., Andersen, P.K., Melbye, M., Adami, H.O., and Gyllensten, U.B. (2000). Viral load of human papilloma virus 16 as a determinant for development of cervical carcinoma in situ: a nested case-control study. Lancet *355*, 2189-2193.

Joyce, J.G., Tung, J.S., Przysiecki, C.T., Cook, J.C., Lehman, E.D., Sands, J.A., Jansen, K.U., and Keller, P.M. (1999). The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes. J Biol Chem *274*, 5810-5822.

Kalantari, M., Blennow, E., Hagmar, B., and Johansson, B. (2001). Physical state of HPV16 and chromosomal mapping of the integrated form in cervical carcinomas. Diagn Mol Pathol *10*, 46-54.

Kalantari, M., Karlsen, F., Kristensen, G., Holm, R., Hagmar, B., and Johansson, B. (1998). Disruption of the E1 and E2 reading frames of HPV 16 in cervical carcinoma is associated with poor prognosis. Int J Gynecol Pathol *17*, 146-153.

Kang, S., Kim, H.S., Seo, S.S., Park, S.Y., Sidransky, D., and Dong, S.M. (2007). Inverse correlation between RASSF1A hypermethylation, KRAS and BRAF mutations in cervical adenocarcinoma. Gynecol Oncol *105*, 662-666.

Karlsen, F., Kalantari, M., Jenkins, A., Pettersen, E., Kristensen, G., Holm, R., Johansson, B., and Hagmar, B. (1996). Use of multiple PCR primer sets for optimal detection of human papillomavirus. J Clin Microbiol *34*, 2095-2100.

Kersemaekers, A.M., Kenter, G.G., Hermans, J., Fleuren, G.J., and van de Vijver, M.J. (1998). Allelic loss and prognosis in carcinoma of the uterine cervix. Int J Cancer *79*, 411-417.

Kim, J.Y., Park, S., Nam, B.H., Roh, J.W., Lee, C.H., Kim, Y.H., Shin, H.J., Lee, S.K., Kong, S.Y., Seong, M.W*., et al.* (2009). Low initial human papilloma viral load implicates worse prognosis in patients with uterine cervical cancer treated with radiotherapy. J Clin Oncol *27*, 5088-5093.

Kim, K.H., and Kim, Y.S. (1995). Role of human papillomavirus and p53 tumor suppressor gene in cervical carcinogenesis. Yonsei Med J *36*, 412-425.

Kiyono, T., Foster, S.A., Koop, J.I., McDougall, J.K., Galloway, D.A., and Klingelhutz, A.J. (1998). Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. Nature *396*, 84-88.

Klaes, R., Woerner, S.M., Ridder, R., Wentzensen, N., Duerst, M., Schneider, A., Lotz, B., Melsheimer, P., and von Knebel Doeberitz, M. (1999). Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res *59*, 6132-6136.

Kleter, B., van Doorn, L.J., ter Schegget, J., Schrauwen, L., van Krimpen, K., Burger, M., ter Harmsel, B., and Quint, W. (1998). Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. Am J Pathol *153*, 1731-1739.

Klingelhutz, A.J., Foster, S.A., and McDougall, J.K. (1996). Telomerase activation by the E6 gene product of human papillomavirus type 16. Nature *380*, 79-82.

Konishi, I., Fujii, S., Nonogaki, H., Nanbu, Y., Iwai, T., and Mori, T. (1991). Immunohistochemical analysis of estrogen receptors, progesterone receptors, Ki-67 antigen, and human papillomavirus DNA in normal and neoplastic epithelium of the uterine cervix. Cancer *68*, 1340-1350.

Koopman, L.A., Szuhai, K., van Eendenburg, J.D., Bezrookove, V., Kenter, G.G., Schuuring, E., Tanke, H., and Fleuren, G.J. (1999). Recurrent integration of human papillomaviruses 16, 45, and 67 near translocation breakpoints in new cervical cancer cell lines. Cancer Res *59*, 5615-5624.

Kraus, I., Driesch, C., Vinokurova, S., Hovig, E., Schneider, A., von Knebel Doeberitz, M., and Durst, M. (2008). The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. Cancer Res *68*, 2514-2522.

Kumar, A., White, T.A., MacKenzie, A.P., Clegg, N., Lee, C., Dumpit, R.F., Coleman, I., Ng, S.B., Salipante, S.J., Rieder, M.J*., et al.* (2011). Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. Proc Natl Acad Sci U S A *108*, 17087-17092.

Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc *4*, 1073-1081.

Kurkure, A.P., and Yeole, B.B. (2006). Social inequalities in cancer with special reference to South Asian countries. Asian Pac J Cancer Prev *7*, 36-40.

Lee, D., Kim, H.Z., Jeong, K.W., Shim, Y.S., Horikawa, I., Barrett, J.C., and Choe, J. (2002). Human papillomavirus E2 down-regulates the human telomerase reverse transcriptase promoter. J Biol Chem *277*, 27748-27756.

Lee, J.H., Lee, S.K., Yang, M.H., Ahmed, M.M., Mohiuddin, M., and Lee, E.Y. (1996). Expression and mutation of H-ras in uterine cervical cancer. Gynecol Oncol *62*, 49-54.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, N., Franceschi, S., Howell-Jones, R., Snijders, P.J., and Clifford, G.M. (2011). Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. Int J Cancer *128*, 927-935.

Liu, F.S., Kohler, M.F., Marks, J.R., Bast, R.C., Jr., Boyd, J., and Berchuck, A. (1994). Mutation and overexpression of the p53 tumor suppressor gene frequently occurs in uterine and ovarian sarcomas. Obstet Gynecol *83*, 118-124.

Lockwood, W.W., Coe, B.P., Williams, A.C., MacAulay, C., and Lam, W.L. (2007). Whole genome tiling path array CGH analysis of segmental copy number alterations in cervical cancer cell lines. Int J Cancer *120*, 436-443.

Luft, F., Klaes, R., Nees, M., Durst, M., Heilmann, V., Melsheimer, P., and von Knebel Doeberitz, M. (2001). Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. Int J Cancer *92*, 9-17.

Madeleine, M.M., Brumback, B., Cushing-Haugen, K.L., Schwartz, S.M., Daling, J.R., Smith, A.G., Nelson, J.L., Porter, P., Shera, K.A., McDougall, J.K*., et al.* (2002). Human leukocyte antigen class II and cervical cancer risk: a population-based study. J Infect Dis *186*, 1565-1574.

Madsen, B.E., Ramos, E.M., Boulard, M., Duda, K., Overgaard, J., Nordsmark, M., Wiuf, C., and Hansen, L.L. (2008). Germline mutation in RNASEL predicts increased risk of head and neck, uterine cervix and breast cancer. PLoS One *3*, e2492.

Maliekal, T.T., Antony, M.L., Nair, A., Paulmurugan, R., and Karunagaran, D. (2003). Loss of expression, and mutations of Smad 2 and Smad 4 in human cervical cancer. Oncogene *22*, 4889-4897.

Manolaraki, M.M., Arvanitis, D.A., Sourvinos, G., Sifakis, S., Koumantakis, E., and Spandidos, D.A. (2002). Frequent loss of heterozygosity in chromosomal region 9pter-p13 in tumor biopsies and cytological material of uterine cervical cancer. Cancer Lett *176*, 175-181.

Manos, M.M., Ting, Y., Wright, D.K., Lewis, A.J., Broker, T.R., and Wolinsky, S.M. (1989). Use of polymerase chain reaction amplification for the detection og genital human papillomaviruses

Cancer Cells *7*, 209-214.

Matovina, M., Sabol, I., Grubisic, G., Gasperov, N.M., and Grce, M. (2009). Identification of human papillomavirus type 16 integration sites in high-grade precancerous cervical lesions. Gynecol Oncol *113*, 120-127.

Melsheimer, P., Vinokurova, S., Wentzensen, N., Bastert, G., and von Knebel Doeberitz, M. (2004). DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. Clin Cancer Res *10*, 3059-3063.

Meyerson, M., Gabriel, S., and Getz, G. (2010). Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet *11*, 685-696.

Minaguchi, T., Yoshikawa, H., Nakagawa, S., Yasugi, T., Yano, T., Iwase, H., Mizutani, K., Shiromizu, K., Ohmi, K., Watanabe, Y*., et al.* (2004). Association of PTEN mutation with HPV-negative adenocarcinoma of the uterine cervix. Cancer Lett *210*, 57-62.

Mittal, K.R., Demopoulos, R.I., and Goswami, S. (1993). Proliferating cell nuclear antigen (cyclin) expression in normal and abnormal cervical squamous epithelia. Am J Surg Pathol *17*, 117-122.

Miwa, K., Miyamoto, S., Kato, H., Imamura, T., Nishida, M., Yoshikawa, Y., Nagata, Y., and Wake, N. (1995). The role of p53 inactivation in human cervical cell carcinoma development. Br J Cancer *71*, 219-226.

Miyai, K., Furugen, Y., Matsumoto, T., Iwabuchi, K., Hirose, S., Kinoshita, K., and Fujii, H. (2004). Loss of heterozygosity analysis in uterine cervical adenocarcinoma. Gynecol Oncol *94*, 115-120.

Miyake, T., Yoshino, K., Enomoto, T., Takata, T., Ugaki, H., Kim, A., Fujiwara, K., Miyatake, T., Fujita, M., and Kimura, T. (2008). PIK3CA gene mutations and amplifications in uterine cancers, identified by methods that avoid confounding by PIK3CA pseudogene sequences. Cancer Lett *261*, 120-126.

Monk, B.J., Sill, M.W., Burger, R.A., Gray, H.J., Buekers, T.E., and Roman, L.D. (2009). Phase II trial of bevacizumab in the treatment of persistent or recurrent squamous cell carcinoma of the cervix: a gynecologic oncology group study. J Clin Oncol *27*, 1069-1074.

Munirajan, A.K., Kannan, K., Bhuvarahamurthy, V., Ishida, I., Fujinaga, K., Tsuchida, N., and Shanmugam, G. (1998). The status of human papillomavirus and tumor suppressor genes p53 and p16 in carcinomas of uterine cervix from India. Gynecol Oncol *69*, 205-209.

Munoz, N., and Bosch, F.X. (1997). Cervical cancer and human papillomavirus: epidemiological evidence and perspectives for prevention. Salud Publica Mex *39*, 274-282.

Munoz, N., Bosch, F.X., de Sanjose, S., Herrero, R., Castellsague, X., Shah, K.V., Snijders, P.J., and Meijer, C.J. (2003). Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med *348*, 518-527.

Murphy, N., Ring, M., Killalea, A.G., Uhlmann, V., O'Donovan, M., Mulcahy, F., Turner, M., McGuinness, E., Griffin, M., Martin, C*., et al.* (2003). p16INK4A as a marker for cervical dyskaryosis: CIN and cGIN in cervical biopsies and ThinPrep smears. J Clin Pathol *56*, 56-63.

Nakashima, R., Fujita, M., Enomoto, T., Haba, T., Yoshino, K., Wada, H., Kurachi, H., Sasaki, M., Wakasa, K., Inoue, M*., et al.* (1999). Alteration of p16 and p15 genes in human uterine tumours. Br J Cancer *80*, 458-467.

Nambaru, L., Meenakumari, B., Swaminathan, R., and Rajkumar, T. (2009). Prognostic significance of HPV physical status and integration sites in cervical cancer. Asian Pac J Cancer Prev *10*, 355-360.

Narayan, G., and Murty, V.V. (2010). Integrative genomic approaches in cervical cancer: implications for molecular pathogenesis. Future Oncol *6*, 1643-1652.

Narayan, G., Pulido, H.A., Koul, S., Lu, X.Y., Harris, C.P., Yeh, Y.A., Vargas, H., Posso, H., Terry, M.B., Gissmann, L*., et al.* (2003). Genetic analysis identifies putative tumor suppressor sites at 2q35-q36.1 and 2q36.3-q37.1 involved in cervical cancer progression. Oncogene *22*, 3489-3499.

Ng, S.B., Nickerson, D.A., Bamshad, M.J., and Shendure, J. (2010). Massively parallel sequencing and rare disease. Hum Mol Genet *19*, R119-124.

Nobbenhuis, M.A., Walboomers, J.M., Helmerhorst, T.J., Rozendaal, L., Remmink, A.J., Risse, E.K., van der Linden, H.C., Voorhorst, F.J., Kenemans, P., and Meijer, C.J. (1999). Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study. Lancet *354*, 20-25.

Nomenklatura., B.R. (2010). In 26th International PapillomavirusConference (Montreal, Canada).

O'Sullivan M, J., Rader, J.S., Gerhard, D.S., Li, Y., Trinkaus, K.M., Gersell, D.J., and Huettner, P.C. (2001). Loss of heterozygosity at 11q23.3 in vasculoinvasive and metastatic squamous cell carcinoma of the cervix. Hum Pathol *32*, 475-478.

Pappa, K.I., Choleza, M., Markaki, S., Giannikaki, E., Kyroudi, A., Vlachos, G., Voulgaris, Z., and Anagnou, N.P. (2006). Consistent absence of BRAF mutations in cervical and endometrial cancer despite KRAS mutation status. Gynecol Oncol *100*, 596-600.

Park, J.S., Dong, S.M., Kim, H.S., Lee, J.Y., Um, S.J., Park, I.S., Kim, S.J., and Namkoong, S.E. (1999). Detection of p16 gene alteration in cervical cancer using tissue microdissection and LOH study. Cancer Lett *136*, 101-108.

Parker, M.F., Arroyo, G.F., Geradts, J., Sabichi, A.L., Park, R.C., Taylor, R.R., and Birrer, M.J. (1997). Molecular characterization of adenocarcinoma of the cervix. Gynecol Oncol *64*, 242-251.

Peedicayil, A., Abraham, P., Sathish, N., John, S., Shah, K., Sridharan, G., and Gravitt, P. (2006). Human papillomavirus genotypes associated with cervical neoplasia in India. Int J Gynecol Cancer *16*, 1591-1595.

Peitsaro, P., Johansson, B., and Syrjanen, S. (2002). Integrated human papillomavirus type 16 is frequently found in cervical cancer precursors as demonstrated by a novel quantitative real-time PCR technique. J Clin Microbiol *40*, 886-891.

Peter, M., Rosty, C., Couturier, J., Radvanyi, F., Teshima, H., and Sastre-Garau, X. (2006). MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. Oncogene *25*, 5985-5993.

Peter, M., Stransky, N., Couturier, J., Hupe, P., Barillot, E., de Cremoux, P., Cottu, P., Radvanyi, F., and Sastre-Garau, X. (2010). Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. J Pathol *221*, 320-330.

Pett, M., and Coleman, N. (2007). Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? J Pathol *212*, 356-367.

Pett, M.R., Herdman, M.T., Palmer, R.D., Yeo, G.S., Shivji, M.K., Stanley, M.A., and Coleman, N. (2006). Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. Proc Natl Acad Sci U S A *103*, 3822-3827.

Pillai, R.M., Babu, J.M., Jissa, V.T., Lakshmi, S., Chiplunkar, S.V., Patkar, M., Tongaonkar, H., Reddy, K.B., Chakka, K.N., Siddiqui, M*., et al.* (2010). Region-wise distribution of high-risk human papillomavirus types in squamous cell carcinomas of the cervix in India. Int J Gynecol Cancer *20*, 1046-1051.

Pirami, L., Giache, V., and Becciolini, A. (1997). Analysis of HPV16, 18, 31, and 35 DNA in pre-invasive and invasive lesions of the uterine cervix. J Clin Pathol *50*, 600-604.

Poljak, M., and Kocjan, B.J. (2010). Commercially available assays for multiplex detection of alpha human papillomaviruses. Expert Rev Anti Infect Ther *8*, 1139-1162.

Pulido, H.A., Fakruddin, M.J., Chatterjee, A., Esplin, E.D., Beleno, N., Martinez, G., Posso, H., Evans, G.A., and Murty, V.V. (2000). Identification of a 6-cM minimal deletion at 11q23.1-23.2 and exclusion of PPP2R1B gene as a deletion target in cervical cancer. Cancer Res *60*, 6677-6682.

Qi, X.P., Ma, J.M., Du, Z.F., Ying, R.B., Fei, J., Jin, H.Y., Han, J.S., Wang, J.Q., Chen, X.L., Chen, C.Y*., et al.* (2011). RET germline mutations identified by exome sequencing in a Chinese multiple endocrine neoplasia type 2A/familial medullary thyroid carcinoma family. PLoS One *6*, e20353.

Rajkumar, T., Vijayalakshmi, N., Sabitha, K., Shirley, S., Selvaluxmy, G., Bose, M.V., and Nambaru, L. (2009). A 7 gene expression score predicts for radiation response in cancer cervix. BMC Cancer *9*, 365.

Rao, P.H., Arias-Pulido, H., Lu, X.Y., Harris, C.P., Vargas, H., Zhang, F.F., Narayan, G., Schneider, A., Terry, M.B., and Murty, V.V. (2004). Chromosomal amplifications, 3q gain and deletions of 2q33-q37 are the frequent genetic changes in cervical carcinoma. BMC Cancer *4*, 5.

Reuter, S., Bartelmann, M., Vogt, M., Geisen, C., Napierski, I., Kahn, T., Delius, H., Lichter, P., Weitz, S., Korn, B*., et al.* (1998). APM-1, a novel human gene, identified by aberrant co-transcription with papillomavirus oncogenes in a cervical carcinoma cell line, encodes a BTB/POZ-zinc finger protein with growth inhibitory activity. EMBO J *17*, 215-222.

Richards, R.I. (2001). Fragile and unstable chromosomes in cancer: causes and consequences. Trends Genet *17*, 339-345.

Romanczuk, H., and Howley, P.M. (1992). Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. Proc Natl Acad Sci U S A *89*, 3159-3163.

Sankaranarayanan, R., Black, R.J., Swaminathan, R., and Parkin, D.M. (1998). An overview of cancer survival in developing countries. IARC Sci Publ, 135-173.

Sankaranarayanan, R., and Wesley, R. (2003). A practical manual on visual screening for cervical neoplasia,

International Agency for Research on Cancer Technical Paper No. 41. WHO, IARCPress.

Saranath, D., Khan, Z., Tandle, A.T., Dedhia, P., Sharma, B., Contractor, R., Shrivastava, S., and Dinshaw, K. (2002). HPV16/18 prevalence in cervical lesions/cancers and p53 genotypes in cervical cancer patients from India. Gynecol Oncol *86*, 157-162.

Sathirapongsasuti, J.F., Lee, H., Horst, B.A., Brunner, G., Cochran, A.J., Binder, S., Quackenbush, J., and Nelson, S.F. (2011). Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics *27*, 2648-2654.

Sathish, N., Abraham, P., Peedicayil, A., Sridharan, G., John, S., Shaji, R.V., and Chandy, G. (2004). HPV DNA in plasma of patients with cervical carcinoma. J Clin Virol *31*, 204-209.

Scheffner, M., Huibregtse, J.M., Vierstra, R.D., and Howley, P.M. (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell *75*, 495-505.

Schiffman, M., Clifford, G., and Buonaguro, F.M. (2009). Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. Infect Agent Cancer *4*, 8.

Schmitt, M., Bravo, I.G., Snijders, P.J., Gissmann, L., Pawlita, M., and Waterboer, T. (2006). Bead-based multiplex genotyping of human papillomaviruses. J Clin Microbiol *44*, 504-512.

Schmitz, M., Driesch, C., Jansen, L., Runnebaum, I.B., and Durst, M. (2012). Non-Random Integration of the HPV Genome in Cervical Cancer. PLoS One *7*, e39632.

Scotto, L., Narayan, G., Nandula, S.V., Arias-Pulido, H., Subramaniyam, S., Schneider, A., Kaufmann, A.M., Wright, J.D., Pothuri, B., Mansukhani, M.*, et al.* (2008a). Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. Genes Chromosomes Cancer *47*, 755-765.

Scotto, L., Narayan, G., Nandula, S.V., Subramaniyam, S., Kaufmann, A.M., Wright, J.D., Pothuri, B., Mansukhani, M., Schneider, A., Arias-Pulido, H.*, et al.* (2008b). Integrative genomics analysis of chromosome 5p gain in cervical cancer reveals target over-expressed genes, including Drosha. Mol Cancer *7*, 58.

Sellors, J.W., and Sankaranarayanan, R. (2003). Colposcopy and Treatment of Cervical Intraepithelial Neoplasia. A Beginner's Manual. In.

Shafti-Keramat, S., Handisurya, A., Kriehuber, E., Meneguzzi, G., Slupetzky, K., and Kirnbauer, R. (2003). Different heparan sulfate proteoglycans serve as cellular receptors for human papillomaviruses. J Virol *77*, 13125-13135.

Sherwood, J.B., Shivapurkar, N., Lin, W.M., Ashfaq, R., Miller, D.S., Gazdar, A.F., and Muller, C.Y. (2000). Chromosome 4 deletions are frequent in invasive cervical cancer and differ between histologic variants. Gynecol Oncol *79*, 90-96.

Shinohara, A., Yokoyama, Y., Wan, X., Takahashi, Y., Mori, Y., Takami, T., Shimokawa, K., and Tamaya, T. (2001). Cytoplasmic/nuclear expression without mutation of exon 3 of the beta-catenin gene is frequent in the development of the neoplasm of the uterine cervix. Gynecol Oncol *82*, 450-455.

Singh, A., Datta, P., Jain, S.K., Bhatla, N., Dutta Gupta, S., Dey, B., and Singh, N. (2009). Human papilloma virus genotyping, variants and viral load in tumors, squamous intraepithelial lesions, and controls in a north Indian population subset. Int J Gynecol Cancer *19*, 1642-1648.

Smith, J.S., Lindsay, L., Hoots, B., Keys, J., Franceschi, S., Winer, R., and Clifford, G.M. (2007). Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. Int J Cancer *121*, 621-632.

Smits, H.L. (1996). The use of consensus PCR and direct sequence analysis for the identification of HPV. Methods Mol Biol *50*, 109-126.

Smits, H.L., Bollen, L.J., Tjong, A.H.S.P., Vonk, J., Van Der Velden, J., Ten Kate, F.J., Kaan, J.A., Mol, B.W., and Ter Schegget, J. (1995). Intermethod variation in detection of human papillomavirus DNA in cervical smears. J Clin Microbiol *33*, 2631-2636.

Soderlund-Strand, A., Carlson, J., and Dillner, J. (2009). Modified general primer PCR system for sensitive detection of multiple types of oncogenic human papillomavirus. J Clin Microbiol *47*, 541-546.

Sowjanya, A.P., Jain, M., Poli, U.R., Padma, S., Das, M., Shah, K.V., Rao, B.N., Devi, R.R., Gravitt, P.E., and Ramakrishna, G. (2005). Prevalence and distribution of high-risk human papilloma virus (HPV) types in invasive squamous cell carcinoma of the cervix and in normal women in Andhra Pradesh, India. BMC Infect Dis *5*, 116.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R.*, et al.* (2012). The landscape of cancer genes and mutational processes in breast cancer. Nature *486*, 400-404.

Stoeber, K., Swinn, R., Prevost, A.T., de Clive-Lowe, P., Halsall, I., Dilworth, S.M., Marr, J., Turner, W.H., Bullock, N., Doble, A.*, et al.* (2002). Diagnosis of genito-urinary tract cancer by detection of minichromosome maintenance 5 protein in urine sediments. J Natl Cancer Inst *94*, 1071-1079.

Storey, A., Thomas, M., Kalita, A., Harwood, C., Gardiol, D., Mantovani, F., Breuer, J., Leigh, I.M., Matlashewski, G., and Banks, L. (1998). Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. Nature *393*, 229-234.

Swan, D.C., Tucker, R.A., Tortolero-Luna, G., Mitchell, M.F., Wideroff, L., Unger, E.R., Nisenbaum, R.A., Reeves, W.C., and Icenogle, J.P. (1999). Human papillomavirus (HPV) DNA copy number is dependent on grade of cervical disease and HPV type. J Clin Microbiol *37*, 1030-1034.

Tenti, P., Pavanello, S., Padovan, L., Spinillo, A., Vesentini, N., Zappatore, R., Migliora, P., Zara, C., Ranzani, G.N., and Carnevali, L. (1998). Analysis and clinical implications of p53 gene mutations and human papillomavirus type 16 and 18 infection in primary adenocarcinoma of the uterine cervix. Am J Pathol *152*, 1057-1063.

Thorland, E.C., Myers, S.L., Gostout, B.S., and Smith, D.I. (2003). Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. Oncogene *22*, 1225-1237.

Thorland, E.C., Myers, S.L., Persing, D.H., Sarkar, G., McGovern, R.M., Gostout, B.S., and Smith, D.I. (2000). Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. Cancer Res *60*, 5916-5921.

Tieben, L.M., ter Schegget, J., Minnaar, R.P., Bouwes Bavinck, J.N., Berkhout, R.J., Vermeer, B.J., Jebbink, M.F., and Smits, H.L. (1993). Detection of cutaneous and genital HPV types in clinical samples by PCR using consensus primers. J Virol Methods *42*, 265-279.

Timmermann, B., Kerick, M., Roehr, C., Fischer, A., Isau, M., Boerno, S.T., Wunderlich, A., Barmeyer, C., Seemann, P., Koenig, J.*, et al.* (2011). Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. PLoS One *5*, e15661.

Travasso, C.M., Anand, M., Samarth, M., Deshpande, A., and Kumar-Sinha, C. (2008). Human papillomavirus genotyping by multiplex pyrosequencing in cervical cancer patients from India. J Biosci *33*, 73-80.

Tsuda, H., Takarabe, T., Okada, S., Uchida, H., Kasamatsu, T., Yamada, T., Tsunematsu, R., Ohmi, K., and Hirohashi, S. (2002). Different pattern of loss of heterozygosity among endocervical-type adenocarcinoma, endometrioid-type adenocarcinoma and adenoma malignum of the uterine cervix. Int J Cancer *98*, 713-717.

van den Brule, A.J., Meijer, C.J., Bakels, V., Kenemans, P., and Walboomers, J.M. (1990). Rapid detection of human papillomavirus in cervical scrapes by combined general primer-mediated and type-specific polymerase chain reaction. J Clin Microbiol *28*, 2739-2743.

van Duin, M., Snijders, P.J., Schrijnemakers, H.F., Voorhorst, F.J., Rozendaal, L., Nobbenhuis, M.A., van den Brule, A.J., Verheijen, R.H., Helmerhorst, T.J., and Meijer, C.J. (2002). Human papillomavirus 16 load in normal and abnormal cervical scrapes: an indicator of CIN II/III and viral clearance. Int J Cancer *98*, 590-595.

Van Ranst, M., Kaplan, J.B., and Burk, R.D. (1992). Phylogenetic classification of human papillomaviruses: correlation with clinical manifestations. J Gen Virol *73 (Pt 10)*, 2653-2660.

Van Tine, B.A., Kappes, J.C., Banerjee, N.S., Knops, J., Lai, L., Steenbergen, R.D., Meijer, C.L., Snijders, P.J., Chatis, P., Broker, T.R.*, et al.* (2004). Clonal selection for transcriptionally active viral oncogenes during progression to cancer. J Virol *78*, 11172-11186.

Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C.K., Stephens, P., Davies, H., Jones, D., Lin, M.L., Teague, J.*, et al.* (2011). Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature *469*, 539-542.

Veldman, T., Horikawa, I., Barrett, J.C., and Schlegel, R. (2001). Transcriptional activation of the telomerase hTERT gene by human papillomavirus type 16 E6 oncoprotein. J Virol *75*, 4467-4472.

Vernon, S.D., Unger, E.R., Miller, D.L., Lee, D.R., and Reeves, W.C. (1997). Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer. Int J Cancer *74*, 50-56.

Vinokurova, S., Wentzensen, N., Kraus, I., Klaes, R., Driesch, C., Melsheimer, P., Kisseljov, F., Durst, M., Schneider, A., and von Knebel Doeberitz, M. (2008). Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. Cancer Res *68*, 307-313.

Virmani, A.K., Muller, C., Rathi, A., Zoechbauer-Mueller, S., Mathis, M., and Gazdar, A.F. (2001). Aberrant methylation during cervical carcinogenesis. Clin Cancer Res *7*, 584-589.

von Knebel Doeberitz, M., Bauknecht, T., Bartsch, D., and zur Hausen, H. (1991). Influence of chromosomal integration on glucocorticoid-regulated transcription of growth-stimulating papillomavirus genes E6 and E7 in cervical carcinoma cells. Proc Natl Acad Sci U S A *88*, 1411-1415.

Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J., and Munoz, N. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol *189*, 12-19.

Wang, C.C., Lai, C.H., Huang, H.J., Chao, A., Chang, C.J., Chang, T.C., Chou, H.H., and Hong, J.H. (2010). Clinical effect of human papillomavirus genotypes in patients with cervical cancer undergoing primary radiotherapy. Int J Radiat Oncol Biol Phys *78*, 1111-1120.

Wang, N., Qian, X., Wang, S., Gao, H., Wang, L., Huo, Y., and Zhang, S. (2012). CCND1 rs9344 polymorphisms are associated with the genetic susceptibility to cervical cancer in Chinese population. Mol Carcinog *51*, 196-205.WCR (2008).

Wentzensen, N., Ridder, R., Klaes, R., Vinokurova, S., Schaefer, U., and Doeberitz, M.K. (2002). Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions. Oncogene *21*, 419-426.

Wentzensen, N., Schiffman, M., Dunn, T., Zuna, R.E., Gold, M.A., Allen, R.A., Zhang, R., Sherman, M.E., Wacholder, S., Walker, J.*, et al.* (2009). Multiple human papillomavirus genotype infections in cervical cancer progression in the study to understand cervical cancer early endpoints and determinants. Int J Cancer *125*, 2151-2158.

Wentzensen, N., Vinokurova, S., and von Knebel Doeberitz, M. (2004). Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. Cancer Res *64*, 3878-3884.

WHO (2009). Information Centre on HPV and Cervical Cancer (HPV Information Centre). Human Papillomavirus and Related Cancers in India. Available at www. who. int/ hpvcentre.

Wilson, R., Fehrmann, F., and Laimins, L.A. (2005). Role of the E1--E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. J Virol *79*, 6732-6740.

Wilting, S.M., de Wilde, J., Meijer, C.J., Berkhof, J., Yi, Y., van Wieringen, W.N., Braakhuis, B.J., Meijer, G.A., Ylstra, B., Snijders, P.J.*, et al.* (2008). Integrated genomic and transcriptional profiling identifies chromosomal loci with altered gene expression in cervical cancer. Genes Chromosomes Cancer *47*, 890-905.

Wingo, S.N., Gallardo, T.D., Akbay, E.A., Liang, M.C., Contreras, C.M., Boren, T., Shimamura, T., Miller, D.S., Sharpless, N.E., Bardeesy, N.*, et al.* (2009). Somatic LKB1 mutations promote cervical cancer progression. PLoS One *4*, e5137.

Wise-Draper, T.M., Allen, H.V., Thobe, M.N., Jones, E.E., Habash, K.B., Munger, K., and Wells, S.I. (2005). The human DEK proto-oncogene is a senescence inhibitor and an upregulated target of high-risk human papillomavirus E7. J Virol *79*, 14309-14317.

Wong, Y.F., Chung, T.K., Cheung, T.H., Lam, S.K., Xu, Y.G., and Chang, A.M. (1995). Frequent ras gene mutations in squamous cell cervical cancer. Cancer Lett *95*, 29-32.

Woodman, C.B., Collins, S.I., and Young, L.S. (2007). The natural history of cervical HPV infection: unresolved issues. Nat Rev Cancer *7*, 11-22.

Yaginuma, Y., Yamashita, T., Ishiya, T., Morizaki, A., Katoh, Y., Takahashi, T., Hayashi, H., and Ishikawa, M. (2000). Abnormal structure and expression of PTEN/MMAC1 gene in human uterine cancers. Mol Carcinog *27*, 110-116.

Yan, X.J., Xu, J., Gu, Z.H., Pan, C.M., Lu, G., Shen, Y., Shi, J.Y., Zhu, Y.M., Tang, L., Zhang, X.W.*, et al.* (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet *43*, 309-315.

Yang, Y.C., Shyong, W.Y., Chang, M.S., Chen, Y.J., Lin, C.H., Huang, Z.D., Wang, Hsu, M.T., and Chen, M.L. (2001). Frequent gain of copy number on the long arm of chromosome 3 in human cervical adenocarcinoma. Cancer Genet Cytogenet *131*, 48-53.

Yee, C.J., Lin, O., and Boyd, J. (2000). Analysis of fibroblast growth factor receptor 3 S249C mutation in cervical carcinoma. J Natl Cancer Inst *92*, 1848-1849.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M.*, et al.* (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64-69.

Yu, T., Ferber, M.J., Cheung, T.H., Chung, T.K., Wong, Y.F., and Smith, D.I. (2005). The role of viral integration in the development of cervical cancer. Cancer Genet Cytogenet *158*, 27-34.

Zhang, Z., Borecki, I., Nguyen, L., Ma, D., Smith, K., Huettner, P.C., Mutch, D.G., Herzog, T.J., Gibb, R.K., Powell, M.A*., et al.* (2007). CD83 gene polymorphisms increase susceptibility to human invasive cervical cancer. Cancer Res *67*, 11202-11208.

Ziegert, C., Wentzensen, N., Vinokurova, S., Kisseljov, F., Einenkel, J., Hoeckel, M., and von Knebel Doeberitz, M. (2003). A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques. Oncogene *22*, 3977-3984.

zur Hausen, H. (2002). Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer *2*, 342-350.

Appendix I

SI No.	Lab ID	HPV type	Age	Disease free	Status
1	CT1082	HPV 16	50	YES	Alive
2	CT1089	HPV 16	48	YES	Alive
3	CT1094	HPV 16	58	NO	Died of disease
4	CT1097	HPV 16	55	YES	Alive
5	CT1098	HPV 16	50	YES	Alive
6	CT1101	HPV 16	50	YES	Alive
7	CT1107	HPV 16	50	YES	Alive
8	CT1108	HPV 16	54	YES	Alive
9	CT1111	HPV 16	55	YES	Alive
10	CT1121	HPV 16	60	YES	Alive
11	CT1122	HPV 16	37	YES	Alive
12	CT1123	HPV 16	40	YES	Alive
13	CT1125	HPV 16	63	YES	Alive
14	CT1126	HPV 16	63	YES	Alive
15	CT1135	HPV 16	38	YES	Alive
16	CT1138	HPV 16	40	YES	Alive
17	CT1160	HPV 16	64	NO	Died of disease
18	CT1161	HPV 16	36	YES	Alive
19	CT1174	HPV 16	50	No	Died of disease
20	CT1181	HPV 16	51	YES	Alive
21	CT1183	HPV 16	58	YES	Alive
22	CT1184	HPV 16	40	NO	Died of disease
23	CT1185	HPV 16	60	YES	Alive
24	CT1193	HPV 16	35	NO	Died of disease
25	CT1209	HPV 16	40	NO	Died of disease
26	CT1210	HPV 16	48	YES	Alive
27	CT1215	HPV 16	55	NO	Died of disease
28	CT1228	HPV 16	50	YES	Alive
29	CT700	HPV 16	54	YES	Alive
30	CT702	HPV 16	60	YES	Alive
31	CT706	HPV 16	53	YES	Alive
32	CT707	HPV 16	44	NO	Died of disease
33	CT709	HPV 16	65	YES	Alive
34	CT711	HPV 16	50	YES	Alive
35	CT712	HPV 16	62	YES	Alive
36	CT714	HPV 16	65	YES	Died of unknown reason
37	CT716	HPV 16	50	NO	Died of disease
38	CT721	HPV 16	38	YES	Alive
39	CT723	HPV 16	62	NO	Died of disease
40	CT726	HPV 16	64	NO	Died of disease
41	CT733	HPV 16	62	YES	Alive
42	CT735	HPV 16	45	NO	Died of disease

43	CT739	HPV 16	70	NO	Died of disease
44	CT744	HPV 16	50	NO	Died of disease
45	CT745	HPV 16	65		Alive
46	CT753	HPV 16	40	YES	Alive
47	CT755	HPV 16	44	NO	Died of disease
48	CT759	HPV 16	57	YES	Alive
49	CT776	HPV 16	55	YES	Alive
50	CT777	HPV 16	72	YES	Alive
51	CT781	HPV 16	40	NO	Died of disease
52	CT787	HPV 16	45	YES	Alive
53	CT792	HPV 16	60	NO	Died of disease
54	CT796	HPV 16	50	YES	Alive
55	CT801	HPV 16	40	YES	Alive
56	CT805	HPV 16	52	YES	Alive
57	CT809	HPV 16	62	YES	Alive
58	CT825	HPV 16	45	YES	Alive
59	CT828	HPV 16	66	YES	Alive
60	CT831	HPV 16	50	YES	Alive
61	CT833	HPV 16	46	NO	Died of disease
62	CT836	HPV 16	58	YES	Alive
63	CT837	HPV 16	34	NO	Died of disease
64	CT839	HPV 16	61	NO	Died of disease
65	CT844	HPV 16	45	YES	Alive
66	CT846	HPV 16	33	YES	Alive
67	CT848	HPV 16	55	YES	Alive
68	CT851	HPV 16	48	YES	Alive
69	CT853	HPV 16	53	NO	Died of disease
70	CT859	HPV 16	46	YES	Alive
71	CT864	HPV 16	39	YES	Alive
72	CT866	HPV 16	35	YES	Alive
73	CT867	HPV 16	65	NO	Died of disease
74	CT868	HPV 16	55	YES	Alive
75	CT877	HPV 16	50	YES	Alive
76	CT887	HPV 16	45	YES	Alive
77	CT889	HPV 16	50	NO	Died of disease
78	CT890	HPV 16	56	YES	Alive
79	CT892	HPV 16	50	YES	Alive
80	CT893	HPV 16	55	YES	Alive
81	CT896	HPV 16	40	NO	Died of disease
82	CT906	HPV 16	57	YES	Alive
83	CT912	HPV 16	51	NO	Died of disease
84	CT914	HPV 16	55	NO	Died of disease
85	CT918	HPV 16	45	YES	Died of unkown disease
86	CT922	HPV 16	48	NO	Died of disease

87	CT934	HPV 16	44	YES	Alive
88	CT935	HPV 16	62	YES	Alive
89	CT940	HPV 16	55	YES	Alive
90	CT951	HPV 16	50	YES	Alive
91	CT958	HPV 16	44	YES	Alive
92	CT992	HPV 16	70	NO	Died of disease
93	CT999	HPV 16	49	YES	Alive
94	CT805	HPV 16	52	YES	Alive
95	CT1015	HPV 16	40	YES	Alive
96	CT1154	HPV 16	34	YES	Alive
97	CT 1199	HPV 16	46	YES	Alive
98	CT1202	HPV 16	46	YES	Alive
99	CT1212	HPV 16	54	YES	Died of unrelated cause
100	CT757	HPV 16	34	NO	Died of disease
101	CT763	HPV 16	60	NO	Died of disease
102	CT766	HPV 16	58	YES	Alive
103	CT783	HPV 16	45	YES	Alive
104	CT800	HPV 16	49	YES	Alive
105	CT843	HPV 16	35	YES	Alive
106	CT919	HPV 16	55	NO	Died of disease
107	CT927	HPV 16	40	NO	Died of disease
108	CT959	HPV 16	58	YES	Alive
109	CT976	HPV 16	35	NO	Died of disease
110	CT1073	HPV 18	48	YES	Alive
111	CT1109	HPV 18	60	NO	Died of disease
112	CT821	HPV 18	80	YES	Alive
113	CT871	HPV 18	46	YES	Alive
114	CT884	HPV 18	46	NO	Died of disease
115	CT905	HPV 18	42	NO	Died of disease
116	CT1019	HPV 16 & 18	48	YES	Alive
117	CT785	HPV 16 & 18	37	NO	Died of disease
118	CT1102	HPV 16 & 18	63	YES	Alive
119	CT1117	HPV 16 & 18	49	YES	Alive
120	CT1162	HPV 16 & 18	62	YES	Dead of other cause
121	CT1170	HPV 16 & 18	55	YES	Alive
122	CT740	HPV 16 & 18	33	NO	Died of disease
123	CT752	HPV 16 & 18	53	YES	Alive
124	CT915	HPV 16 & 18	50	NO	Died of disease
125	CT938	HPV 16 & 18	50	NO	Died of disease

APPENDIX II

Gene	No. of Samples	Mutation	Position	Zygosity	Effect	Prediction		Important pathways/ Function	Reported in COSMIC
						SIFT	POLYPHEN		
AAMP	1	V164G (cAc/cCc)	chr2:219131616	Ht	NS	damaging	probably damaging	GO:0001525~angiogenesis;GO:0010595~ positive regulation of endothelial cell migration	
AMIGO2	1	R370H (gCg/gTg)	chr12:47471677	Ht	NS	damaging	benign	GO:0006916~anti-apoptosis;GO:0043069~negative regulation of programmed cell death	
CDC42BPA	1	R1248T (tCt/tGt)	chr1:227216942	Ht	NS	damaging	probably damaging	GO:0004674~protein serine/threonine kinase activity	
CDH26	1	R288* (Cga/Tga)	chr20:58562532	Ht	Stop Gained			GO:0007156~homophilic cell adhesion	
COL15A1	1	G990* (Gga/Tga)	chr9:101816875	Ht	Stop Gained			04974~Protein digestion and absorption	
COX10	1	H418P (cAc/cCc)	chr17:14110451	Ht	NS	damaging	probably damaging	1100~Metabolic pathways	
CREB3L4	1	D1188N (Gac/Aac)	chr1:153941029	Ht	NS	damaging	probably damaging	05215~Prostate cancer	
DHDH	1	P271T (Ccc/Acc)	chr19:49447680	Ht	NS	tolerated	probably damaging	00040~Pentose and glucuronate interconversions; 00980~Metabolism of xenobiotics by cytochrome P450	
EDNRB	1	L233F (aaC/aaG)	chr13:78492280	Ht	NS	damaging	probably damaging	GO:0000122~negative regulation of transcription from RNA polymerase II promoter	
EP300	1	D1399N (Gat/Aat)	chr22:41565529	Ht	NS	damaging	probably damaging	04110~Cell cycle; 04310~Wnt signaling pathway; 04330~Notch signaling pathway; 04350~TGF-beta signaling pathway; 04630~Jak-STAT signaling pathway; 05200~Pathways in cancer;	Stomach cancer
EWSR1	1	A126P (Gcc/Scc)	chr22:29674168	Ht	NS	damaging	not predicted	GO:0045449~regulation of transcription	
FAM5B	1	A68T (Gct/Act)	chr1:177199214	Ht	NS	damaging	probably damaging	GO:0045786~negative regulation of cell cycle	

FGF22	1	R100W (Cgg/Tgg)	chr19:643318	Ht	NS	damaging	probably damaging	04010~MAPK signaling pathway; 05200~Pathways in cancer; etc.	
G6PD	1	V390M (Cac/Tac)	chrX:153761178	Ht	NS	damaging	not predicted	01100~Metabolic pathways	
HAUS3	1	V1304A (gTg/gCg)	chr4:2242247	Ht	NS	tolerated	possibly damaging	GO:0051301~cell division	
HMX2	1	N199T (aAc/aCc)	chr10:124909413	Ht	NS	damaging	probably damaging	GO:0008284~positive regulation of cell proliferation; GO:0006351~transcription, DNA-dependent	
HNMT	1	N24Y (Aac/Tac)	chr2:138722131	Ht	NS	damaging	benign	00340~Histidine metabolism	
KDM5C	1	H445Y (Gta/Ata)	chrX:53240747	Ht	NS	damaging	probably damaging	GO:0006351~transcription, GO:0034720~histone H3-K4 demethylation	
KLF9	1	H225Y (Gtg/Atg)	chr9:73002754	Ht	NS	damaging	probably damaging	GO:0006351~transcription	
LAMB2	1	T1796S/A (Tgg/Kgg)	chr3:49158670	Ht	NS	damaging	not predicted	04510~Focal adhesion; 04512~ECM-receptor interaction; 05200~Pathways in cancer;	
LATS2	1	D510H (Ctg/Gtg)	chr13:21562391	Ht	NS	tolerated	possibly damaging	GO:0051301~cell division;GO:0090090~negative regulation of canonical Wnt receptor signaling pathway;GO:0006468~protein phosphorylation;GO:0007067~mitosis	
LNPEP	1	Y237* (taT/taG)	chr5:96315533	Ht	Stop Gained			04614~Renin-angiotensin system	
MACF1	1	R2091W (Cgg/Tgg)	chr1:39798420	Ht	NS	damaging	possibly damaging	GO:0007050~cell cycle arrest	
MAPK15	1	R470Q (cGg/cAg)	chr8:144804001	Hm	NS	tolerated	probably damaging	GO:0000165~MAPK cascade;GO:0046777~protein autophosphorylation	
MYD88	1	D113H (Gac/Cac)	chr3:38180489	Ht	NS	damaging	probably damaging	04210~Apoptosis; etc	
NCOR2	1	G1289V (cCa/cAa)	chr12:124832860	Hm	NS	not predicted	probably damaging	04330~Notch signaling pathway	
NUDT2	1	A96G (gCg/gGg)	chr9:34343281	Ht	NS	tolerated	probably damaging	GO:0006917~induction of apoptosis	
OGDHL	1	F836Y/C (aAg/aRg)	chr10:50946003	Ht	NS	damaging	not predicted	01100~Metabolic pathways	
PDHB	1	L104V (Aac/Cac)	chr3:58416663	Ht	NS	damaging	possibly damaging	01100~Metabolic pathways	

PIK3R1	1	aacact→act at codon 453	chr5:67589595	Ht	Deletion				Breast cancer
PKHD1	1	P1026A (Gga/Cga)	chr6:51907678	Ht	NS	damaging	probably damaging	GO:0010824~regulation of centrosome duplication;GO:0016337~cell-cell adhesion	
POMT2	1	Q371* (Gtc/Atc)	chr14:77762512	Ht	Stop Gained			00514~O-Mannosyl glycan biosynthesis	
PROK1	1	H61P(cAc/cCc)	chr1:110996692	Ht	NS	damaging	probably damaging	GO:0000187~activation of MAPK activity;GO:0008284~positive regulation of cell proliferation;GO:0051781~positive regulation of cell division	
PRPF4B	1	R856H (cGt/cAt)	chr6:4056655	Ht	NS	damaging	probably damaging	GO:0004674~protein serine/threonine kinase activity;GO:0005524~ATP binding	
PSPH	2	R65H (gCg/gTg)	chr7:56087374	Ht	NS	damaging	probably damaging	01100~Metabolic pathways, etc.	
PTPDC1	1	E639K (Gag/Aag)	chr9:96860769	Ht	NS	damaging	probably damaging	GO:0004725~protein tyrosine phosphatase activity;GO:0008138~protein tyrosine/serine/threonine phosphatase activity	
RASGRF1	2	T661P/S (Tgc/Ygc)	chr15:79298661	Ht	NS	damaging	not predicted	04010~MAPK signaling pathway; 04510~Focal adhesion	
RELA	1	A544S(Cgg/Agg)	chr11:65421875	Ht	NS	tolerated	probably damaging	04010~MAPK signaling pathway; 05200~Pathways in cancer; etc.	
RNF111	1	Q605* (Cag/Tag)	chr15:59368279	Ht	Stop Gained			GO:0016567~protein ubiquitination;GO:0045893~positive regulation of transcription, DNA-dependent	
SIPA1L3	1	A670T (Gct/Act)	chr19:38591845	Hm	NS	damaging	benign	GO:0005096~GTPase activator activity	
TAF5	1	E383Q (Gag/Cag)	chr10:105139398	Ht	NS	tolerated	possibly damaging	03022~Basal transcription factors	
TGFBR2	1	Q220* (Cag/Tag)	chr3:30713258	Ht	Stop Gained			04010~MAPK signaling pathway; 05200~Pathways in cancer;	
UBE2O	1	N1010H (Ttg/Ctg)	chr17:74388113	Hm	NS	tolerated	probably damaging	04120~Ubiquitin mediated proteolysis	
USP17L2	1	E244K (Ctt/Ttt)	chr8:11995540	Ht	NS	damaging	possibly damaging	GO:0006915~apoptotic process;GO:0007049~cell cycle	
MTMR8	1	Q511H	chrX:63490902	Ht	NS	damaging	probably	GO:0004725~protein tyrosine phosphatase	

		(gtC/gtG)					damaging	activity	
MED23	1	V97G (cAa/cCa)	chr6:131944597	Ht	NS	damaging	probably damaging	GO:0006355~regulation of transcription, DNA-dependent; GO:0010467~gene expression	
MAST1	1	P1520L (cCc/cTc)	chr19:12985530	Ht	NS	damaging	benign	GO:0006468~protein phosphorylation;GO:0004674~protein serine/threonine kinase activity	
CDCA7L	1	Y64C (aTg/aCg)	chr7:21951347	Ht	NS	tolerated	probably damaging	GO:0006351~transcription, DNA-dependent;GO:0008284~positive regulation of cell proliferation	
HIVEP3	1	E155K (Ctc/Ttc)	chr1:42050006	Ht	NS	tolerated	probably damaging	GO:0006351~transcription, DNA-dependent; GO:0045893~positive regulation of transcription	
MYBBP1A	1	L883M (Gac/Tac)	chr17:4446453	Ht	NS	damaging	probably damaging	GO:0006351~transcription, DNA-dependent	
SELPLG	1	Q151P (gTt/gGt)	chr12:109017680	Ht	NS	tolerated	not predicted	04514~Cell adhesion molecules (CAMs)	Acute myeloid leukaemia
DCC	1	Q460E (Caa/Saa)	chr18:50683842	Ht	NS	tolerated	benign	05200~Pathways in cancer	Ovarian cancer

Ht= Heterozygous; Hm= Homozygous; NS= Non-synonymous;

OPEN ACCESS Freely available online

PLoS one

HPV Genotyping and Site of Viral Integration in Cervical Cancers in Indian Women

Poulami Das[1], Asha Thomas[1], Umesh Mahantshetty[2], Shyam K. Shrivastava[2], Kedar Deodhar[3], Rita Mulherkar[1]*

1 Mulherkar Lab, Advanced Centre for Treatment, Research and Education in Cancer (ACTREC), Tata Memorial Centre, Navi Mumbai, India, **2** Department of Radiation Oncology, Tata Memorial Hospital, Tata Memorial Centre, Mumbai, India, **3** Department of Pathology, Tata Memorial Hospital, Tata Memorial Centre, Mumbai, India

Abstract

Persistent HPV infection plays a major role in cervical cancer. This study was undertaken to identify HPV types in a cohort of Indian women with locally advanced cervical cancer as well as to determine the physical state and/or site of viral integration in the host genome. Pretreatment biopsies (n = 270) from patients were screened for HPV infection by a high throughput HPV genotyping assay based on luminex xMAP technology as well as MY09/11 PCR and SPF1/2 PCR. Overall HPV positivity was observed to be 95%, with HPV16 being most common (63%) followed by infection with HPV18. Integration status of the virus was identified using Amplification of Papillomavirus Oncogene Transcripts (APOT) assay in a subset of samples positive for HPV16 and/or HPV18 (n = 86) and with an adequate follow-up. The data was correlated with clinical outcome of the patients. Integration of the viral genome was observed in 79% of the cases and a preference for integration into the chromosomal loci 1p, 3q, 6q, 11q, 13q and 20q was seen. Clinical data revealed that the physical state of the virus (integrated or episomal) could be an important prognostic marker for cervical cancer.

Citation: Das P, Thomas A, Mahantshetty U, Shrivastava SK, Deodhar K, et al. (2012) HPV Genotyping and Site of Viral Integration in Cervical Cancers in Indian Women. PLoS ONE 7(7): e41012. doi:10.1371/journal.pone.0041012

Editor: Torbjörn Ramqvist, Karolinska Institutet, Sweden

Received February 17, 2012; **Accepted** June 15, 2012; **Published** July 16, 2012

Funding: The authors wish to express their gratitude for the financial support from Women's Cancer Initiative IRG#634. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rmulherkar@actrec.gov.in

Introduction

Cervical cancer is the third most common cancer among women worldwide and the most common cancer found in Indian women. HPV infection has been shown to play a critical, though not sufficient, role in the etiology of cervical cancers. Till date more than 200 HPV types have been reported, of which HPV16 is most common, followed generally by HPV18, HPV45, HPV31 and HPV33 [1,2]. Most of the high risk HPV (HR-HPV) infections (90%) regress spontaneously and only in about 10% cases the infection persists and progresses to high-grade cervical intraepithelial neoplasia. This generally occurs through integration of the HPV genome into the host chromosome with associated loss or disruption of E2 [3]. According to available reports, viral E2 gene has the ability to repress viral E6 and E7 oncogenes in cells harbouring integrated HPV DNA [4]. Therefore integration of the virus with loss of transcriptional control by E2 results in overexpression of E6 and E7 leading to immortalization and transformation of cells [5]. In most of the cases integration of HR-HPV genome gives rise to fusion transcripts comprising of viral oncogenes E6, E7 and adjacent cellular sequences [6,7,8,9]. In vitro studies have demonstrated that the viral-cellular fusion transcripts are more stable and impart the cells with a selective growth advantage as compared to the episomal counterparts [10,11].

Studies report that the integration event is random involving almost all the chromosomes, and accordingly several virus-host integration sites have been mapped till date [12]. However, there are certain hotspots e.g., fragile sites, translocation break points and transcriptionally active regions [3,13,14,15] which are preferred by the virus for its integration into the human genome. On integration within or near a gene, the virus can bring about a change in its expression which may eventually lead to alterations in cellular growth and proliferation. Also, viral integration can render both viral coding genes as well as the cellular genes susceptible to epigenetic changes which could regulate their expression. Hence, the integration of HPV into the human genome is considered an important event in cervical carcinogenesis.

The aim of this study was HPV genotyping and identification of site of integration of two HR-HPV types (16 and 18), along with evaluation of the prognostic value of integration site, in locally advanced cervical cancers in Indian women.

Materials and Methods

Clinical Samples

Pretreatment cervical tumor biopsies, predominantly from FIGO stage IIIB, were obtained from patients (median age, 50 years; age range, 33–80 years) undergoing radiotherapy alone or concomitant chemo-radiation at the Radiation Oncology Department, Tata Memorial Hospital, Mumbai, after obtaining IRB approval. A generic consent for basic research was obtained prior to obtaining the biopsies. However, for the current study a consent waiver was obtained from the Hospital Ethics Committee. The biopsies were obtained from histologically proven, primary

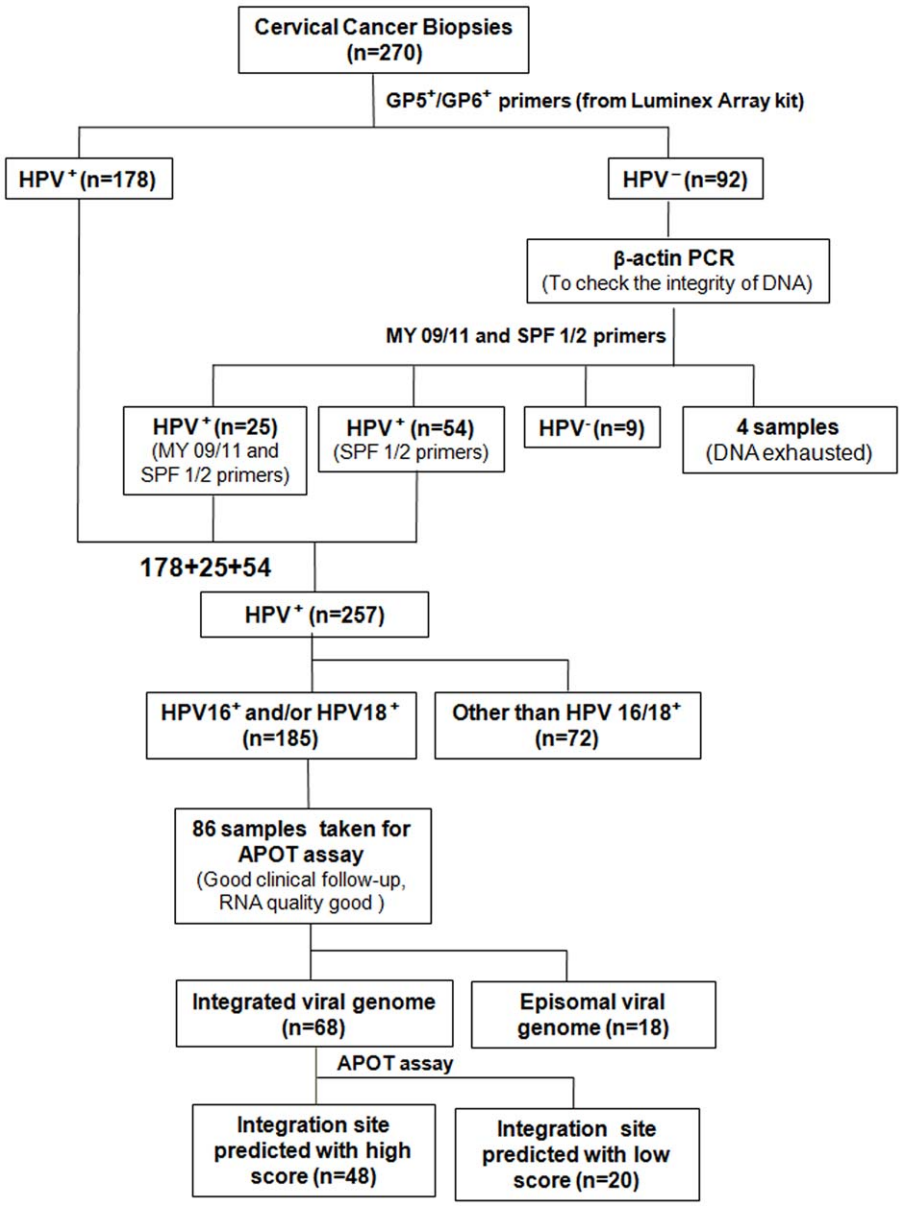

Figure 1. Flowchart depicting summary of the study. Genotyping was carried out on 270 advanced stage cervical cancer samples by high-throughput, $GP5^+/6^+$ primers based luminex array; consensus MY09/11 and SPF1/2 primers. HPV positivity was 95% (257/270). APOT assay was done on 86 $HPV16^+$ and/or $HPV18^+$ samples, with good clinical follow-up and good quality RNA. In 18 samples, only episomal form of HPV was identified, rest 68 hinted toward possible integration. Site of integration could be predicted with high score by BLAST and/or BLAT in 48 samples.
doi:10.1371/journal.pone.0041012.g001

cervical tumor, before the start of radical radiation therapy and were coded for de-identification by the physician prior to testing. The samples were collected in liquid nitrogen and stored at $-80^{\circ}C$ until further use. All the samples were assigned a laboratory code to maintain confidentiality.

Processing of tumor samples

Frozen tissues were cryocut for extraction of DNA and RNA. For DNA extraction, five 30 µm sections were collected in STE buffer (0.1 M NaCl, 0.05 M Tris pH 7.5, 1 mM EDTA, 1% SDS) containing 10 mg/ml Proteinase K (USB, Cleveland, OH, US). DNA was isolated by standard phenol-chloroform method. For isolation of total RNA, RNeasy Mini Kit (Qiagen, Hilden, Germany) was used. Ten 30 µm sections were collected in RLT buffer containing guanidine thiocyanate provided in the kit and processed following manufacturer's instructions. DNAse treatment of the RNA samples was carried out using DNA free kit (Ambion, Austin, TX, US).

HPV genotyping by high throughput luminex array

Genotyping of 24 HPV types which included 15 high-risk types (16, 18, 31, 33, 35, 39, 45, 51,52, 56, 58, 59, 68, 73, and 82), 3 putative high-risk types (26, 53, and 66), and 6 low-risk types (6, 11, 42, 43, 44, and 70) [1,2] was carried out using multiplex HPV genotyping array (Multimetrix GmbH, Heidelberg, Germany) based on luminex xMAP technology. As per the manufacturer's instructions, PCR was carried out using sets of biotinylated broad range primers in a total volume of 50 µL containing 3.5 mM

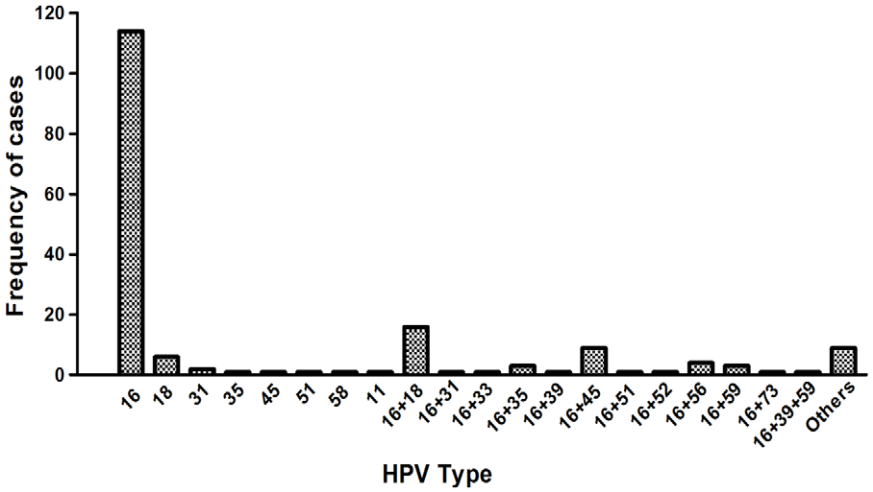

Figure 2. Frequency of 24 HPV types as detected by Luminex array. The graph depicts frequency of 24 HPV types in 178 cervical cancer biopsy samples which were found to be positive $GP5^{+}/6^{+}$ primers. HPV16 infection predominated in the samples. Each bar represents different HPV types.
doi:10.1371/journal.pone.0041012.g002

$MgCl_2$, 200 μM dNTPs, 0.75 unit of Taq DNA polymerase and 1μl primer mix. The amplification steps included an initial DNA denaturation at 94°C for 5 min, followed by 40 cycles of denaturation for 20 s at 94°C, annealing for 30 s at 38°C, and extension for 1 min 20 s at 71°C, before a final extension for 4 min at 71°C. PCR positive samples were then subjected to the luminex run. Ten microlitres of the PCR product was mixed with the luminex bead mix containing distinct bead populations coupled to 24 HPV types. After thermal denaturation, the target sequences were hybridized to bead-bound probes. The hybridized PCR products were labeled by binding to R-phycoerythrin conjugated streptavidin. The read-out was obtained in the luminex bioanalyzer (Luminex Corporation, Austin, TX, USA). HPV types were discerned according to the unique bead signature, whereas the presence of PCR products was determined by phycoerythrin fluorescence. An analytical sensitivity cut-off was calculated based on the negative control which was deducted from each of the read-out.

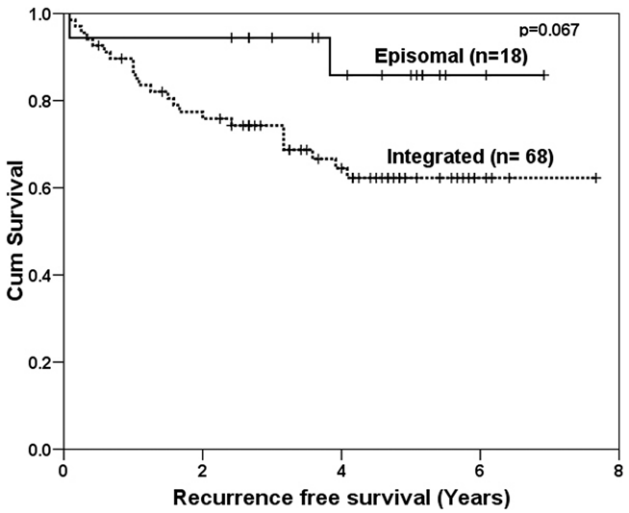

Figure 3. Kaplan-Meier analysis for episomal vs. integrated viral genome. Kaplan-Meier survival curve for patients with episomal form of virus (n = 18) vs. integrated form (n = 68) is depicted. Most of the patients with episomal form (16 out of 18) had a disease free survival as compared to patients with integrated form, indicating a good clinical outcome, although with a borderline significance (p = 0.067).
doi:10.1371/journal.pone.0041012.g003

HPV genotyping by PCR using MY09/11 and SPF1/2 primers

Since the amount of DNA available for the study was limiting, β-actin PCR was done only in those samples which failed to show amplification by $GP5^{+}/6^{+}$ primers (n = 92). These were further screened for HPV by PCR using MY09/11 L1 primers (Table S1) and SPF1/2 primers [16]. PCR was carried out in a reaction volume of 25 μl containing, 1.5 mM $MgCl_2$, 10μM of each primer, 200 μM dNTPs and 0.75 unit of Taq DNA polymerase. The samples which tested positive for HPV either by MY09/11 or SPF1/2 or both were further genotyped for the two most common HR-HPV types- HPV16 and 18 using HPV16/18 specific primers (Table S1). Since the amount of DNA was limiting, SPF 1/2 PCR could not be carried out in 4 of the 92 samples.

Association of HPV16, HPV18 and HPV16/18 infection with clinical outcome

The genotyping data for the two HR-HPV types, HPV16, HPV18 and HPV16/18 together, where adequate follow-up data was available was compared with the clinical outcome of the patients. Kaplan-Meier analysis (SPSS 15.0) was done to determine association between infection with these HR-HPV types and recurrence of disease. Disease free survival was considered from start of radiation therapy to the time when recurrence occurred or till last follow-up. Statistical significance was evaluated using the log-rank test (SPSS 15.0).

Identification of integration site by APOT assay

Samples (n = 86) positive for HPV16, HPV18 or both and with an adequate clinical follow-up (minimum 1yr or till occurrence of first event, whichever was earlier) were taken to study viral integration using Amplification of Papillomavirus Oncogene Transcripts (APOT) assay, as described by Klaes *et al* [6]. Briefly, total RNA (0.5–1 μg) was reverse transcribed using oligo(dT)17-primer coupled to a linker sequence [(dT)17–p3] [17] and 50 units

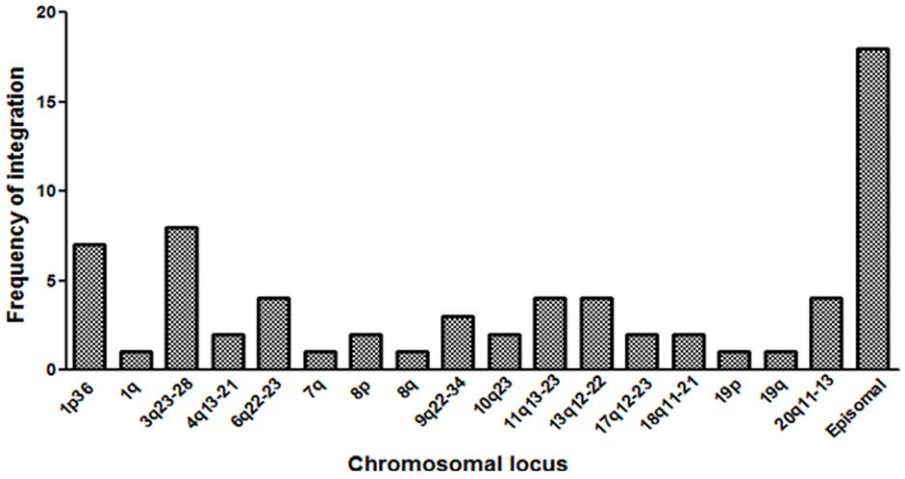

Figure 4. Frequency of HPV integration into different chromosomal loci. Site of integration as determined by APOT assay in 48 cases positive for HPV16, HPV18 or both and with high prediction score using BLAST/BLAT. Integration event was found to be more common in 1p and 3q chromosomal loci. Each bar represents different chromosomal locus.
doi:10.1371/journal.pone.0041012.g004

of Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA, USA) for 1 h at 42°C. PCR with β-actin primers was carried out to check the integrity of the cDNA. First strand cDNA was amplified using HPV E7–specific primer (p1–16 specific for HPV16 and p1–18 specific for HPV18) as forward primers and linker p3 as the reverse primer (Table S1). The PCR amplification was carried out in a reaction volume of 50 µL containing 2.5 mM $MgCl_2$, 200 µM dNTPs, 25 µM of each primer and 1 unit of Taq DNA polymerase. The reaction comprised of an initial denaturation step at 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 4 min. This was followed by a final extension at 72°C for 20 min. Next, 7 µl of the PCR product was used as template for nested PCR using forward primers p2–16 specific for HPV16 or p2–18 specific for HPV18 and (dT)17–p3 as reverse primer (Table S1). The PCR conditions were same as that of first PCR except that the annealing temperature was 66°C.

Cloning of fusion transcripts

Amplicons other than the major episomal transcripts (~1050 bp for HPV16 and ~1000 bp for HPV18) were suspected to be derived from the integrated HPV genomes. These were excised from the gel and DNA isolated using GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Buckinghamshire, UK). The isolated DNA was either sequenced directly or following cloning into pTZ57R/T vector using the InsTA PCR Cloning Kit (Fermentas, Lithuania, EU), on DNA sequencer (3100 Avant Genetic analyzer, Applied Biosystems, Foster City, CA, USA). The chromosomal integration loci were determined using National Centre for Biotechnology Information (BLAST) and the University of California, Santa Cruz (UCSC) hg19 (Feb. 2009) (BLAT) human genome assemblies. Further, the integration sites were checked for the presence of fragile sites and any genes of known identity by using NCBI fragile site map viewer and the UCSC Blat tool respectively.

Association of viral integration with clinical outcome

The data obtained was compared with the clinical outcome of the patients. Kaplan-Meier analysis (SPSS 15.0) was done to determine the association of the viral state (episomal/integrated) with recurrence of the disease. Disease free survival was considered from start of radiation therapy to the time when recurrence occurred or till last follow-up (median follow-up for 86 cases was 44 months). Statistical significance was evaluated using the log-rank test (SPSS 15.0).

Results

HPV genotyping by high throughput luminex array

Although there are a few reports on different high risk HPVs in Indian women, here we report 15 high risk HPVs, 3 intermediate risk and 6 low risk HPV types using the high throughput luminex array. The primers provided in the luminex array kit for HPV genotyping were biotinylated, broad range $GP5^+/GP6^+$ primers. Using this primer set for PCR, we obtained 178 out of 270 samples positive for HPV (Figure 1). The HPV positive samples were further subjected to hybridization to bead-bound probes by luminex array as described earlier. One hundred sixty nine samples were found to hybridize to the different HPV probes whereas 9 samples were negative. These 9 samples could have HPV infection not included in the 24 types detected by the kit. Out of 169 HPV positive samples, 168 samples were positive for one or more HR-HPV types indicating a high association of cervical cancer with HR-HPV infection. Among these, HPV16 and/or HPV18 infection were most common – 114 samples being positive for HPV16 alone, 6 samples for HPV18 alone and 16 samples for both HPV16 and HPV18 (Figure 2).

HPV genotyping by PCR using MY09/11 and SPF1/2 primers

In order to estimate the true HPV positivity in the 270 cases, the 92 cervical cancer biopsies negative for HPV by luminex array, were first subjected to PCR using β-actin primers to check the quality of DNA. All were found to be positive for β-actin. Next they were subjected to PCR using MY09/11 and SPF1/2 primers. Twenty five out of 92 samples were found to be positive for HPV by MY09/11 PCR. Since the amount of DNA was limiting, SPF 1/2 PCR could not be carried out in 4 of the 92 samples. Screening with SPF1/2 primers revealed 79 samples to be HPV positive. These 79 samples also included the 25 samples that tested HPV positive by MY09/11 PCR. The HPV positivity was therefore calculated taking into account the results from luminex

Table 1. Summary of HPV Integration Sites.

Sample ID	Histology	HPV type	Integration locus	Gene	Fragile site involved
CT 712	SCC	16	1p36.12	ALPQTL2	FRA1A(1p36) com
CT 700	SCC	16	1p36.23	-	FRA1A(1p36) com
CT 702	SCC	16	1p36.23	-	FRA1A(1p36) com
CT 709	SCC	16	1p36.23	-	FRA1A(1p36) com
CT 1138	SCC	16	1p36.33	-	FRA1A(1p36) com
CT 809	SCC	16	1p36.33	-	FRA1A(1p36) com
CT 866	SCC	16	1p36.33	-	FRA1A(1p36) com
CT 864	SCC	16	1q42.13	ABCB10	FRA1H(1q42) com
CT 755	SCC	16	3q23	SLC25A36	-
CT 785	SCC	16	3q23	SLC25A36	-
			20q11.21	COX4I2	-
		18	-	-	-
CT 1210	SCC	16	3q26.2 and Episomal	MECOM	-
CT 723	SCC	16	3q26.31	-	FRA3C(3q27) com
CT 706	SCC	16	3q26.33	SOX2	FRA3C(3q27) com
CT 892	SCC	16	3q26.33	-	FRA3C(3q27) com
CT 711	SCC	16	3q28 and Episomal	LEPREL1	FRA3C(3q27) com
CT 739	SCC	16	3q28	TP63	FRA3C(3q27) com
CT 999	SCC	16	4q13.3	IL8(~3kb)	-
CT 1114	SCC	16	4q21.22	HNRPDL	-
CT 893	SCC	16	6q22.31	-	-
CT 1117	AC	16	6q22.31	-	-
		18	-	-	-
CT 1122	SCC	16	6q23.3	PDE7B	-
CT 846	SCC	16	6q23.3	MAP3K	-
CT 1019	SCC	16	7q11.21	TYW1	FRA7J(7q11) com
CT 796	SCC	16	8p11.21	-	-
CT 915	SCC	16	8p11.21	-	-
		18	-	-	-
Sample ID	**Histology**	**HPV type**	**Integration locus**	**Gene**	**Fragile site involved**
CT 1160	SCC	16	8q24.21	MYC/PVT	Flanked by FRA8E(8q24.1) rare; FRA8C(8q24.1) com; FRA8D(8q24.3) com
CT 714	SCC	16	9q22.32	C9orf3	FRA9D(9q22.1) com
CT 1162	SCC	16	9q22.32 and Episomal	C9orf3	FRA9D(9q22.1) com
		18	-	-	-
CT 753	SCC	16	9q34.11	SLC25A25	-
CT 821	SCC	18	10q23.33	-	FRA10A(10q23.3) rare
CT 859	SCC	16	10q23.31	PTEN	FRA10A(10q23.3) rare
CT 1169	SCC	16	11q13.1 and Episomal	DRAP1	FRA11H(11q13) com
CT 1183	SCC	16	11q13.1	DRAP1	FRA11H(11q13) com
CT 836	SCC	16	11q14.3	-	FRA11F(11q14.2) com
CT 871	AC	18	11q23.3	ARCN1	FRA11G(11q23.3) com
CT 1094	SCC	16	13q	-	-
CT 927	SCC	16	13q12.3	POMP	-
CT 896	SCC	16	13q22.1	-	Flanked by FRA13B(13q21) com
CT 976	SCC	16	13q22.1	KLF5	Flanked by FRA13B(13q21) com
CT 848	SCC	16	17q12 and Episomal	HNF1B	-
CT 914	SCC	16	17q23.1	VMP1	FRA17B(17q23.1) com

Table 1. Cont.

Sample ID	Histology	HPV type	Integration locus	Gene	Fragile site involved
CT 1097	SCC	16	18q11.2	-	-
CT 740	SCC	18	18q21.3	-	FRA18B(18q21.3) com
		16	Episomal	-	-
CT 906	SCC	16	19p13.3	-	FRA19B(19p13) com
CT 889	SCC	16	19q13.43 and Episomal	-	FRA19A(19q13) com
CT 918	SCC	16	20q11.21 and episomal	COX4I2	-
Sample ID	**Histology**	**HPV type**	**Integration locus**	**Gene**	**Fragile site involved**
CT 1170	SCC	16	20q11.21	COX4I2	-
		18	-	-	-
CT 839	SCC	16	20q13.13	-	-
CT 1015	SCC	16	-	-	-
CT 1082	SCC	16	-	-	-
CT 1101	SCC	16	-	-	-
CT 1107	SCC	16	-	-	-
CT 1108	SCC	16	-	-	-
CT 1135	SCC	16	-	-	-
CT 1161	SCC	16	-	-	-
CT 759	SCC	16	-	-	-
CT 787	SCC	16	-	-	-
CT 792	SCC	16	-	-	-
CT 800	SCC	16	-	-	-
CT 828	SCC	16	-	-	-
CT 868	SCC	16	-	-	-
CT 887	SCC	16	-	-	-
CT 905	SCC	18	-	-	-
CT 934	SCC	16	-	-	-
CT 935	SCC	16	-	-	-
CT 951	SCC	16	-	-	-

Key: – SCC: Squamous cell carcinoma, AC: Adenocarcinoma, Com: Common Fragile site
doi:10.1371/journal.pone.0041012.t001

array and SPF1/2 PCR. The overall HPV positivity in this cohort was found to be 95% (257/270). Further genotyping of the 79 samples, using HPV 16/18 specific primers, showed 49 samples to be positive for HPV16. None of these 79 samples were positive for HPV18. Therefore, the prevalence of HPV16 and/or HPV18 in this cohort was 69% (185/270) (Figure 1).

Association of HPV16, 18 and dual infection with clinical outcome

Kaplan-Meier survival analysis data for 125 patients with HPV type16, 18 and dual infection and with adequate clinical follow-up (median follow-up for 125 cases was 54 months), revealed that there was no significant difference between infection with these two HR-HPV types in terms of disease outcome (Figure S1).

Physical state of virus and clinical outcome

Out of 125 HPV16, HPV18 or dual HPV positive samples, a sub-set of 86 samples with good quality RNA, were taken to study the physical state and/or site of viral integration by APOT assay. In most of the cases, the viral genome was found to be integrated (n = 68), whereas in 21% (n = 18) only episomal transcripts could be identified. In 12 cases with integrated viral genome, episomal form of HPV was also detected (Figure 1). The physical state of the virus (episomal/ integrated) was associated with the disease outcome. Survival data revealed that 16 out of 18 patients with only episomal form of HPV (16 and/or 18), had disease free survival as compared to those with integrated form of the virus, indicating a good clinical outcome ($p = 0.067$, representing a borderline significance) (Figure 3). The clinical outcome of all the patients where the viral integration was studied is shown in Table S2.

Identification of viral integration site

In order to understand whether the integration event is random or there is some preference for certain sites within the chromosomes, the sequencing data for 68 cases derived from APOT assay were investigated by Blast and/or Blat. The site of integration could be predicted with a high score in 48 cases (Table S3a), for the remaining 20 cases the score was low (Table S3b). Only those cases where the integration site was predicted with high score (n = 48) were analyzed further for different features associated with the same. The sites of integration were found to be distributed throughout the genome. However,

integration was more frequent at the chromosomal loci 1p (n = 7), 3q (n = 8), 13q (n = 4), 6q (n = 4), 11q (n = 4) and 20q (n = 4) (Figure 4). Only one sample showed HPV integration at two chromosomal loci simultaneously. Some of the recurrent integration sites were also checked at the genomic level by carrying out genomic DNA PCR with HPV E7 primers as the forward primer and primers specific to a given chromosomal region as the reverse one (Figure S2).

Features associated with HPV integration

Using NCBI Fragile site Map Viewer it was observed that 60% of integrations (29/48) were located in or close to a common or rare fragile site. The rest of the integration sites were not associated with any fragile sites (Table 1). Using the UCSC Blat tool 58% of the sequences (28/48) were observed to be either within or nearby protein coding genes. These genes belonged to various categories ranging from oncogenes, transcription factors, and tumor suppressor genes (Table 1).

Discussion

It is proven beyond doubt that infection by HPV plays a major role in the etiology of cervical cancer. Reports from different parts of the sub-continent indicate a prevalence of HPV ranging from 73–99% [18–30]. In the present study we report 95% HPV positivity using three different primer sets. It is apparent from this study that a single set of primers is not sufficient to estimate the true HPV infectivity. Of the various HPV genotypes HPV16 was most common (60%), followed by infection with HPV18 alone (2%). Dual infection with HPV16/18 (6%) and HPV16/45 (3%) was also observed. These results are in concordance with other studies from the Indian subcontinent which reports 57–65% HPV16 positivity, followed by HPV18, 45, and 33 in cervical neoplasia [19,20,29].

Integration of the virus is common in late stage cervical cancers and is considered an important event in the progression of the disease. Integration generally occurs downstream of the early genes E6 and E7, often in the E1 or E2 region. The E2 gene is transcriptionally inactivated once the virus gets integrated due to disruption of its open reading frame. Viral E2 gene has been extensively studied and is known to play a role in viral replication as well as negative regulation of E6 and E7 genes [31].

Various techniques have been used to study integration of the virus, such as Ligation-mediated PCR [32], Restriction site-PCR [15] and APOT assay [6]. In order to limit our study to integration sites with a transcriptionally active viral genome, APOT assay was chosen. Also, APOT assay allows detection of integrated viral genome in clinical lesions even in the presence of a large excess of nonintegrated episomal form of viral genomes [6,33]. The frequency of viral integration into the host genome in cervical carcinomas has been reported to be as high as 100% in $HPV18^+$ tumors [34] and up to 80% in $HPV16^+$ tumors [35]. In our study we found four $HPV18^+$ samples where the virus was integrated and one $HPV18^+$ sample where the virus was episomal. The incidence of integration in $HPV16^+$ samples was higher. The mechanism of HR-HPV integration is not fully understood. It is speculated that integration might represent a chance occurrence, the probability of which increases with the frequency of double-strand breaks (DSBs) in host and viral DNA. Chromosomal fragile sites could represent the hotspots for HR-HPV integration.

The physical state and/or site of integration was studied in 86 cervical tumour samples infected with either or both of two HR-HPVs, HPV16 and HPV18, and with an adequate clinical follow-up. Episomal form of HPV was observed in 18 cases. Presence of only episomal form in these patients with advanced disease stage (predominantly FIGO stage IIIB), could indicate that either HPV integration is not solely responsible for the progression of the disease; or it could be a limitation of the technique resulting in failure of amplification of the integrant derived transcript. Recent studies have confirmed presence of only episomal form of the virus in advanced cervical squamous cell carcinomas [33,36]. In addition, since APOT works on the basis of annealing of the Frohman primer to the polyA tail, cases in which polyA tail is located at a great distance from the forward primer, might not be amenable to amplification by PCR.

In the present study, we observed 12 samples where both integrated as well as episomal forms of the virus were present. In such cases, E2 may be available in *trans* to modulate the expression levels of oncogenic E6 and E7 [37]. Also according to the report by Pett *et al.*, loss of episomes is as much important as integration of the virus into the host genome for progression of lesions to cervical neoplasia [38]. It would be interesting to see the expression of E2 and E6/E7 in cases where HR-HPV is present in episomal as well as integrated form or in samples where HPV is only episomal.

There are reports that HR-HPV proteins other than E6/E7 induce chromosomal instability and transformation [39]. It is reported that E2 stabilizes Skp2, an oncogene and this could lead to activation of S-phase entry [39]. Hamid et. al., [40] have also suggested a role for E2 in cell proliferation. Since cells in S-phase are more responsive to radiation, cancers with episomal E2 could be more responsive to radiation treatment. Comparison of the physical state of the virus (episomal/integrated) with the clinical outcome after radical radiotherapy revealed that patients with episomal form of the virus had increased disease free survival compared to those with integrated form. This observation is supported by various reports which state that the integration event is associated with a decreased disease free survival [41,42]. However, there are contrasting reports as well, according to which physical state of the virus does not correlate with disease free survival [43,44]. This needs to be studied further.

Although integration sites were distributed throughout the genome in different samples, there was a preference for certain chromosomal loci such as 1p, 3q, 6q, 11q, 13q, 6q and 20q. Certain specific regions in some of these loci such as 1p36.23, 1p36.33, 3q26, 3q28 and 20q11.21 showed repeated integrations, indicating that integration might not be a random event. Reports from studies involving western populations indicate that integration of the virus occurs most commonly at 8q chromosomal locus [3,45,46]. Integration at the 8q locus in our study was observed in only 1 out of 48 cases. This may be due to the difference in the ethnicity of the two populations.

The 3q, 13q and 20q loci besides being preferential target for HPV integration have been reported to be sites for genomic instability associated with cervical cancer. Gain of 3q and 20q, while loss of 13q has been reported in various stages of the disease [47–49]. Also more recent reports show that a significant association exists between genomic rearrangement and HPV integration [46]. It would therefore be interesting to study whether the preferential integration of the virus into these loci has a role to play in inducing genomic instability.

Most of the integrations (28/48) were found to be located within or near certain genes. This could indicate that the virus prefers transcriptionally active regions for the integration event. Such genes included oncogenes such as myc, transcription factors like TP63, MECOM, etc. This observation is supported by previous report by Wentzensen *et al* wherein they have shown involvement of tumour related genes (myc and TP63) in HPV integration

process [50]. Studies in our lab have shown that some of the genes within which integration was observed, such as ABCB10, SLC25A36, IL8, COX4I2, HNF1B, myc, demonstrated increased expression (unpublished data), thereby indicating that upon integration within or near a particular gene the virus may bring about changes in gene expression.

Integration of the virus near or within fragile sites has frequently been reported [15,45,50–52]. Fragile sites are specific regions in the chromosomes that nonrandomly undergoes break in response to certain stress. This makes genes in or near these sites susceptible to foreign DNA integration. In our study 29/48 integrations were located within or near a common or rare fragile site which is in concordance with previous reports. Another observation of our study was that patients with viral integration at chromosomal loci 3q, 13q and 20q showed the worst prognosis amongst all. Whether this can have any clinical implication in prognosis of cervical cancer would be interesting and challenging to study.

Supporting Information

Figure S1 Kaplan-Meier analysis for two HR-HPV types – HPV16 and/or HPV18 and disease outcome. Kaplan-Meier survival analysis for HPV16 and/or HPV18 in 125 patients who had a good clinical follow up was carried out. Patients with HPV16 infection alone showed a trend towards better disease free survival as compared to HPV18 infection alone and dual infection with HPV16/18.
(TIF)

Figure S2 Genomic DNA PCR of the recurrent integration sites. Representative gel images (a, b) showing HPV integration at the genomic level. Recurrent integrations at chromosomal loci 1p36.23, 3q28, 6q23.3, 8p11.21 and 11q13.1 is depicted.
(TIF)

Table S1 Primer sequences.
(DOCX)

Table S2 Clinicopathological data for all cases where the viral integration was studied.
(DOCX)

Table S3 Sequence of HPV integration sites in the genome in 68 cases. a) Integration sites for 48 cases with a high prediction score. b) Integration sites for 20 cases with a low prediction score.
(DOCX)

Acknowledgments

The authors wish to acknowledge late Dr. K.A. Dinshaw and all individuals from the Gynecology Disease Management Group, Tata Memorial Hospital, who were involved in judiciously compiling the clinical history and collecting and storing the biopsies from cervical cancer patients. Help from Ms. Sadhana Kannan (ACTREC) in statistical analysis and Ms. Tabish Hussain (ACTREC) for her valuable inputs in preparing the manuscript are gratefully acknowledged.

Author Contributions

Conceived and designed the experiments: RM PD. Performed the experiments: PD AT. Analyzed the data: PD. Contributed reagents/materials/analysis tools: RM. Wrote the paper: PD RM. Sample collection, patient follow-up, clinical inputs: UM SKS. Pathological examination of biopsies: KD.

References

1. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, et al. (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 348: 518–527.
2. Schmitt M, Bravo IG, Snijders PJ, Gissmann L, Pawlita M, et al. (2006) Bead-based multiplex genotyping of human papillomaviruses. J Clin Microbiol 44: 504–512.
3. Wentzensen N, Vinokurova S, von Knebel Doeberitz M (2004) Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. Cancer Res 64: 3878–3884.
4. Bechtold V, Beard P, Raj K (2003) Human papillomavirus type 16 E2 protein has no effect on transcription from episomal viral DNA. J Virol 77: 2021–2028.
5. Romanczuk H, Howley PM (1992) Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. Proc Natl Acad Sci U S A 89: 3159–3163.
6. Klaes R, Woerner SM, Ridder R, Wentzensen N, Duerst M, et al. (1999) Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res 59: 6132–6136.
7. Couturier J, Sastre-Garau X, Schneider-Maunoury S, Labib A, Orth G (1991) Integration of papillomavirus DNA near myc genes in genital carcinomas and its consequences for proto-oncogene expression. J Virol 65: 4534–4538.
8. Choo KB, Pan CC, Han SH (1987) Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. Virology 161: 259–261.
9. Di Luca D, Pilotti S, Stefanon B, Rotola A, Monini P, et al. (1986) Human papillomavirus type 16 DNA in genital tumours: a pathological and molecular analysis. J Gen Virol 67 (Pt 3): 583–589.
10. Jeon S, Allen-Hoffmann BL, Lambert PF (1995) Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. J Virol 69: 2989–2997.
11. Jeon S, Lambert PF (1995) Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. Proc Natl Acad Sci U S A 92: 1654–1658.
12. Yu T, Ferber MJ, Cheung TH, Chung TK, Wong YF, et al. (2005) The role of viral integration in the development of cervical cancer. Cancer Genet Cytogenet 158: 27–34.
13. Ziegert C, Wentzensen N, Vinokurova S, Kisseljov F, Einenkel J, et al. (2003) A comprehensive analysis of HPV integration loci in anogenital lesions combining transcript and genome-based amplification techniques. Oncogene 22: 3977–3984.
14. Koopman LA, Szuhai K, van Eendenburg JD, Bezrookove V, Kenter GG, et al. (1999) Recurrent integration of human papillomaviruses 16, 45, and 67 near translocation breakpoints in new cervical cancer cell lines. Cancer Res 59: 5615–5624.
15. Thorland EC, Myers SL, Persing DH, Sarkar G, McGovern RM, et al. (2000) Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. Cancer Res 60: 5916–5921.
16. Kleter B, van Doorn LJ, ter Schegget J, Schrauwen L, van Krimpen K, et al. (1998) Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. Am J Pathol 153: 1731–1739.
17. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci U S A 85: 8998–9002.
18. Basu P, Chandna P, Bamezai RN, Siddiqi M, Saranath D, et al. (2011) MassARRAY spectrometry is more sensitive than PreTect HPV-Proofer and consensus PCR for type-specific detection of high-risk oncogenic human papillomavirus genotypes in cervical cancer. J Clin Microbiol 49: 3537–3544.
19. Basu P, Roychowdhury S, Bafna UD, Chaudhury S, Kothari S, et al. (2009) Human papillomavirus genotype distribution in cervical cancer in India: results from a multi-center study. Asian Pac J Cancer Prev 10: 27–34.
20. Bhatla N, Lal N, Bao YP, Ng T, Qiao YL (2008) A meta-analysis of human papillomavirus type-distribution in women from South Asia: implications for vaccination. Vaccine 26: 2811–2817.
21. Chatterjee R, Mandal B, Bandyopadhyay S (2003) Detection of HPV DNA in cervical carcinomas by PCR and hybrid capture assay. Indian J Pathol Microbiol 46: 596–599.
22. Franceschi S, Rajkumar T, Vaccarella S, Gajalakshmi V, Sharmila A, et al. (2003) Human papillomavirus and risk factors for cervical cancer in Chennai, India: a case-control study. Int J Cancer 107: 127–133.
23. Peedicayil A, Abraham P, Sathish N, John S, Shah K, et al. (2006) Human papillomavirus genotypes associated with cervical neoplasia in India. Int J Gynecol Cancer 16: 1591–1595.
24. Pillai RM, Babu JM, Jissa VT, Lakshmi S, Chiplunkar SV, et al. (2010) Region-wise distribution of high-risk human papillomavirus types in squamous cell carcinomas of the cervix in India. Int J Gynecol Cancer 20: 1046–1051.

25. Saranath D, Khan Z, Tandle AT, Dedhia P, Sharma B, et al. (2002) HPV16/18 prevalence in cervical lesions/cancers and p53 genotypes in cervical cancer patients from India. Gynecol Oncol 86: 157–162.
26. Travasso CM, Anand M, Samarth M, Deshpande A, Kumar-Sinha C (2008) Human papillomavirus genotyping by multiplex pyrosequencing in cervical cancer patients from India. J Biosci 33: 73–80.
27. Singh A, Datta P, Jain SK, Bhatla N, Dutta Gupta S, et al. (2009) Human papilloma virus genotyping, variants and viral load in tumors, squamous intraepithelial lesions, and controls in a north Indian population subset. Int J Gynecol Cancer 19: 1642–1648.
28. Deodhar K, Gheit T, Vaccarella S, Romao CC, Tenet V, et al. (2012) Prevalence of human papillomavirus types in cervical lesions from women in rural Western India. J Med Virol 84: 1054–1060.
29. Grace Nirmala J, Narendhirakannan. (2012) RT Detection and genotyping of high-risk HPV and evaluation of anti-oxidant status in cervical carcinoma patients in Tamil Nadu State, India – a case control study. Asian Pac J Cancer Prev 12: 2689–2695.
30. Sowjanya AP, Jain M, Poli UR, Padma S, Das M, et al. (2005) Prevalence and distribution of high-risk human papilloma virus (HPV) types in invasive squamous cell carcinoma of the cervix and in normal women in Andhra Pradesh, India. BMC Infect Dis 5: 116.
31. Xue Y, Bellanger S, Zhang W, Lim D, Low J, et al. (2010) HPV16 E2 is an immediate early marker of viral infection, preceding E7 expression in precursor structures of cervical carcinoma. Cancer Res 70: 5316–5325.
32. Luft F, Klaes R, Nees M, Durst M, Heilmann V, et al. (2001) Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells. Int J Cancer 92: 9–17.
33. Vinokurova S, Wentzensen N, Kraus I, Klaes R, Driesch C, et al. (2008) Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. Cancer Res 68: 307–313.
34. Corden SA, Sant-Cassia LJ, Easton AJ, Morris AG (1999) The integration of HPV-18 DNA in cervical carcinoma. Mol Pathol 52: 275–282.
35. Melsheimer P, Vinokurova S, Wentzensen N, Bastert G, von Knebel Doeberitz M (2004) DNA aneuploidy and integration of human papillomavirus type 16 e6/e7 oncogenes in intraepithelial neoplasia and invasive squamous cell carcinoma of the cervix uteri. Clin Cancer Res 10: 3059–3063.
36. Gray E, Pett MR, Ward D, Winder DM, Stanley MA, et al. (2010) In vitro progression of human papillomavirus 16 episome-associated cervical neoplasia displays fundamental similarities to integrant-associated carcinogenesis. Cancer Res 70: 4081–4091.
37. Arias-Pulido H, Peyton CL, Joste NE, Vargas H, Wheeler CM (2006) Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. J Clin Microbiol 44: 1755–1762.
38. Pett MR, Herdman MT, Palmer RD, Yeo GS, Shivji MK, et al. (2006) Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. Proc Natl Acad Sci U S A 103: 3822–3827.
39. Bellanger S, Tan CL, Nei W, He PP, Thierry F (2009) The human papillomavirus type 18 E2 protein is a cell cycle-dependent target of the SCFSkp2 ubiquitin ligase. J Virol 84: 437–444.
40. Hamid NA, Brown C, Gaston K (2009) The regulation of cell proliferation by the papillomavirus early proteins. Cell Mol Life Sci 66: 1700–1717.
41. Kalantari M, Karlsen F, Kristensen G, Holm R, Hagmar B, et al. (1998) Disruption of the E1 and E2 reading frames of HPV 16 in cervical carcinoma is associated with poor prognosis. Int J Gynecol Pathol 17: 146–153.
42. Vernon SD, Unger ER, Miller DL, Lee DR, Reeves WC (1997) Association of human papillomavirus type 16 integration in the E2 gene with poor disease-free survival from cervical cancer. Int J Cancer 74: 50–56.
43. Holm R, Kraus I, Skomedal H, Langerod A, Kristensen GB, et al. (2008) Human papillomavirus DNA and e6/e7 mRNA status in relation to survival of patients treated for cervical squamous cell carcinoma. Open Virol J 2: 74–81.
44. Nambaru L, Meenakumari B, Swaminathan R, Rajkumar T (2009) Prognostic significance of HPV physical status and integration sites in cervical cancer. Asian Pac J Cancer Prev 10: 355–360.
45. Kraus I, Driesch C, Vinokurova S, Hovig E, Schneider A, et al. (2008) The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. Cancer Res 68: 2514–2522.
46. Peter M, Stransky N, Couturier J, Hupe P, Barillot E, et al. (2010) Frequent genomic structural alterations at HPV insertion sites in cervical carcinoma. J Pathol 221: 320–330.
47. Rao PH, Arias-Pulido H, Lu XY, Harris CP, Vargas H, et al. (2004) Chromosomal amplifications, 3q gain and deletions of 2q33–q37 are the frequent genetic changes in cervical carcinoma. BMC Cancer 4: 5.
48. Scotto L, Narayan G, Nandula SV, Arias-Pulido H, Subramaniyam S, et al. (2008) Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. Genes Chromosomes Cancer 47: 755–765.
49. Wilting SM, de Wilde J, Meijer CJ, Berkhof J, Yi Y, et al. (2008) Integrated genomic and transcriptional profiling identifies chromosomal loci with altered gene expression in cervical cancer. Genes Chromosomes Cancer 47: 890–905.
50. Wentzensen N, Ridder R, Klaes R, Vinokurova S, Schaefer U, et al. (2002) Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions. Oncogene 21: 419–426.
51. Matovina M, Sabol I, Grubisic G, Gasperov NM, Grce M (2009) Identification of human papillomavirus type 16 integration sites in high-grade precancerous cervical lesions. Gynecol Oncol 113: 120–127.
52. Dall KL, Scarpini CG, Roberts I, Winder DM, Stanley MA, et al. (2008) Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under noncompetitive conditions. Cancer Res 68: 8249–8259.