Chapter 3

Methods
3.1 STUDY DESIGN

To address the objectives of this study, the following study design was adopted:

- HPV Negative Histologically Normal Control Samples
- HPV16 Positive Histologically Normal Non-malignant Samples
- HPV16 Positive Cervical Cancers
- Episomal HPV16 genomes
- Integrated HPV16 genomes

Microarray based gene expression profiling

Validation of expression profiling of HOX cluster genes and candidate long noncoding RNAs by quantitative Real Time PCR

Identification of the functional relevance and mechanistic implications of one of the HOX cluster encoded, PRC2-bound lncRNA, HOTAIR

- Effect of HOTAIR expression levels (low and high HOTAIR) in cervical cancer development
- Effect of sequence variations in HOTAIR on its expression and association with cervical cancer development
- Effect of HPV16 E7 in regulating HOTAIR expression and function
- Effect of HPV16 E7 in regulating gene expression through HOTAIR interacting, PRC2 complex
3.2 SUBJECTS AND SAMPLES

3.2.1 Statement of ethics

The samples were collected with informed consent from the participants, approved by the ethical committee for human experimentation of National Institute of Biomedical Genomics. All the experiments were performed in accordance with the approved guidelines and regulations.

3.2.2 Subject recruitment for identification of various categories of cervical samples

The biopsy samples were collected from disease-free married women who were attending Gynecology-OPD of various hospitals in West Bengal (Calcutta Medical College Hospital, Colootala St, Bowbazar, Kolkata-700012, College of Medicine & Jawaharlal Nehru Medical College Hospital, Shilpanchal Station Road, Kalyani-741235, Nadia) or undergoing hysterectomy for various reasons other than cancers such as prolapse, fibroid, cyst etc. Information on clinical complaints and demographic and lifestyle variables were obtained using a questionnaire (attached as Annexure I) by female interviewers. All samples (biopsies, punch or surgical, collected by the collaborating Gynecologist) and information on the subjects were collected (by the Gynecologist and us) with informed consent, and were included in the study, after approval of the Institutional Review Board for human experimentation. Pregnant women, women having menstrual bleeding and those with recent childbirth / miscarriage / abortions (< 4 months), were not enrolled in the study.

3.2.3 Malignant tissues from Cervical Cancer patients

The CaCx samples used for this study were derived from married subjects aged 28 - 75 years, attending a cancer referral hospital (Saroj Gupta Cancer Centre and Research Institute, South 24 Parganas, West Bengal, India). Only those women who had not undergone any treatment for CaCx were enrolled for the study. A questionnaire (same as above) was used to collect information from patients on demographic, life-style and reproductive factors, by female interviewers. These included age, socioeconomic status, educational level, age of marriage, parity, age at first childbirth, contraceptive practice adopted tobacco usage habits etc. The clinical information and cervical status of each
patient was noted from the hospital records and included in the same questionnaire. All the samples (biopsies, punch or surgical, collected by the collaborating Gynecologic Oncologist) and information on subjects were collected (by the Gynecologic Oncologist) with informed consent, after approval of the Institutional Review Board for human experimentation.

3.2.4 Sample collection and storage

All the biopsy samples were divided into three sections. One section of the biopsy tissue (from normal subjects or subjects with CaCx) was kept in 10% formalin (diluted in Phosphate Buffered Saline) for histopathological confirmation. The second section was stored in RNALater (for RNA based studies) and the third section was stored in Phosphate Buffered Saline (for DNA based analysis). All the samples were transported from the hospitals to the laboratory in dry-ice. The samples for histopathological examination and those for DNA and RNA isolation were all marked with the same code number for a particular individual.

3.2.5 Histopathological analysis of cervical tissue biopsies

The formalin fixed cervical biopsy samples were paraffinized and the sections were stained using standard protocols by a pathological laboratory and histopathological reading was performed. Our collaborating pathologist, Dr. Sudipta Roy (Sri Aurobindo Seva Kendra, 1H, Gariahat Road (S) Jodhpur Park, Kolkata - 700068, West Bengal, India), interpreted the biopsy samples from all the subjects. The readings were interpreted as well differentiated, moderately differentiated, and poorly differentiated squamous cell carcinoma (SCC) or adenocarcinoma (ADC) with infiltration or carcinoma in situ. Samples identified as SCC with infiltration were analysed in this study. FIGO classification and histology were used for staging and morphological classification of tumor tissues. The findings from tumor staging could help to predict the severity of lesions. The histopathological/cytological classifications of cervical tumors, premalignant lesions and normal cervical tissues were based on the WHO recommendations.

The current study was conducted on histopathologically confirmed invasive squamous cell carcinomas (SCCs) and clinically diagnosed as tumor stage III and above as per FIGO classification, and moderately differentiated. Only those samples corresponding
to normal histology, without any inflammation or minimal inflammation as confirmed by
the oncopathologist, were selected for further analysis.

3.3 MOLECULAR CHARACTERIZATION OF SAMPLES

3.3.1 DNA isolation

DNA isolation from cervical tissues was done by QIAamp DNA mini kit (Cat # 51304). Tissues stored in -80°C were equilibrated at room temperature (15°C-25°C) and approximately 25 mg of each tissue was minced thoroughly with sterile scalpel blades on autoclaved glass blocks. It was then placed in a 1.5 ml microcentrifuge tube and 180 μl of ATL buffer (lysis buffer from Qiagen, pre-warmed at 37°C to remove the precipitate) was added. The tissue sections were minced into very small pieces, to decrease lysis time. Thereafter, 20 μl of Proteinase K was added to the lysate, mixed by vortexing and incubated overnight at 56°C in a shaking water bath, until the tissue was completely lysed. After brief centrifugation, 200 μl of Buffer AL was added to the sample, mixed by pulse-vortexing for 15 secs, and incubated at 70°C for 10 mins. After brief centrifugation, 200 μl of absolute ethanol (MERCK, Germany) was added to the sample and mixed by pulse-vortexing for 15 secs. The lysate was then applied to the QIAamp Spin Column (in a 2 ml collection tube), without wetting the rim and centrifuged at 6,000 x g (8000 rpm) for 1 min. The QIAamp Spin Column was then placed in a clean 2 ml collection tube provided with the kit, and the tube containing the filtrate was discarded. Without wetting the QIAamp Spin Column, 500 μl of Buffer AW1 was added and centrifuged at 6000 x g (8000 rpm) for 1 min. The QIAamp Spin Column was then placed in a clean 2 ml collection tube, and the tube containing the filtrate was discarded. Without wetting the QIAamp Spin Column, 500 μl of Buffer AW2 was then added and centrifuged at full speed (20,000 x g; 14,000 rpm) for 3 mins and the filtrate was discarded. The Column was then placed in the same 2 ml collection tube and centrifuged at 20,000 x g (14,000 rpm) for 1 min, then placed in a clean 1.5 ml microcentrifuge tube and 200 μl of Buffer AE (DNA elution buffer from Qiagen) was added. The column was then incubated at room temperature for 1 min, and centrifuged at 6000 x g (8000 rpm) for 1 min to elute the DNA, which was stored at -20°C for long term usage.
3.3.2 Estimation of DNA concentration and purity

(A) Nanodrop based DNA quantification: Absorbance measurements made on the Thermo NanoDrop™ 8000 Spectrophotometer will include the absorbance of all molecules in the sample that absorb at the wavelength of interest. Nucleic acids absorb light at a wavelength of 260 nm. If a 260 nm light source shines on a sample, the amount of light that passes through the sample can be measured, and the amount of light absorbed by the sample can be inferred. For double-stranded DNA, an Optical Density (OD) of 1 at 260 nm correlates to a DNA concentration of 50 ng/μl. The ratio of absorbance at 260 nm and 280 nm is used to assess the purity of DNA and RNA. A ratio of ~1.8 is generally accepted as “pure” for DNA and a ratio of ~2.0 is generally accepted as “pure” for RNA. A ratio substantially lower than 1.8 is indicative of the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm. Thus, based on the NanoDrop 8000 Spectrophotometer readings, quality and concentration of stock DNA was determined. Working solutions (100 ng/μl) of the DNA, to be used for PCR were prepared by diluting the stock with TE buffer and stored in aliquots to prevent repeated freezing and thawing.

(B) Fluorometer based DNA quantification: The Qubit fluorometer is an instrument used for quantification of DNA, RNA, and protein and used in many different applications. The Qubit 2.0 fluorometer uses fluorescent dyes to determine the concentration of nucleic acids and proteins in a sample. Once added to a solution of DNA, the Qubit DNA dye binds to the DNA within seconds and reaches equilibrium in less than two minutes. At a specific amount of the dye, the amount of fluorescence signal from this mixture is directly proportional to the concentration of DNA in the solution. The Qubit fluorometer picks up this fluorescence signal and converts it into a DNA concentration measurement, using DNA standards of known concentration. The Qubit fluorometer uses DNA standards to derive the relationship between DNA concentration and fluorescence. For quantification of tissue-isolated DNA samples, Qubit dsDNA BR Assay was used (assay range: 100 pg/μl – 1 μg/μl).

3.3.3 Screening of cervical tissue DNA samples for HPV

(A) Detection for presence of HPV infection: DNA from cervical biopsies and scrapes were screened for presence of HPV infection by PCR, using L1 consensus primers,
MY011 and MY09. The 10 μl PCR reaction mixture included 1 μl of 10X PCR Buffer [Roche: 50 mM KCl, 50 mM Tris-HCl (pH 8.3), 50 mM (NH₄)₂SO₄], 2 mM magnesium chloride (MgCl₂), 30 ng of each primer, 200 μM deoxyribonucleotide triphosphates (dNTPs), 0.5 U FastStart Taq DNA polymerase (Roche) and 100 ng of DNA. Samples were amplified for 30 cycles at 95°C for 1 min, 57°C for 2 mins and 72°C for 1 min 30 secs, with initial preheating at 95°C for 5 mins and a final extension at 72°C for 5 mins. The amplified products were then subjected to electrophoresis on a 2% agarose gel and amplified bands (~ 450 bp) were visualized under UV light after staining with ethidium bromide (Figure 3.1).

**Figure 3.1: Gel electrophoresis of L1 PCR products.** Lane 1: 50 bp DNA ladder; Lanes 2-7: Samples showing specific band for L1 at 450 bp; Lane 8: L1 (-)ve sample; Lane 9: positive control; Lane 10: negative control

L1-amplicon negative samples were re-amplified with GP5/6 primers, nested to MY011 and MY09 primers, for further screening of samples. The 10 μl PCR reaction mixture consisted of 1 μl of 10X PCR Buffer, 1 mM magnesium chloride (MgCl₂), 25 ng of each primer, 100 μM deoxyribonucleotide triphosphates (dNTPs), 0.25 U FastStart Taq DNA polymerase (Roche) and 0.3 μl of L1 amplified PCR product. Samples were amplified for 40 cycles at 95°C for 30 secs, 42°C for 1 min and 72°C for 30 secs, with initial preheating at 95°C for 5 mins and a final extension at 72°C for 5 mins. The amplified products were subjected to electrophoresis on a 2% agarose gel and amplified bands (150 bp) were visualized under UV light after staining with ethidium bromide (Figure 3.2). The
primer sequences are provided in Table 3.1. The samples that did not show specific DNA band-size after PCR, with either L1 (450 bp) or GP5/6 (150 bp) primers, were considered as HPV negative [HPV (-)ve] samples. Those samples that showed specific DNA band-size either with L1 or GP5/6 primers were considered to be HPV positive [HPV (+)ve].

Figure 3.2: Gel electrophoresis of GP5/6 PCR products. Lane 6: 50 bp DNA ladder; Lanes 3,4,7,8: samples showing specific band for GP5/6 at 150 bp; Lane 2,5: GP5/6 (-)ve sample; Lane 1: positive control; Lane 9: negative control.

(B) **Testing cervical tissue DNA samples for the presence of HPV16:** HPV16 detection was carried out using type specific primers amplifying a region within the E6 gene. The reaction was carried out with 200 ng of DNA in a final reaction volume of 20 μl (Duttagupta et al, 2002) containing 2 μl of 10X PCR buffer, 2 mM magnesium chloride (MgCl₂), 200 μM deoxynucleotidetriphosphates and 0.5 U FastStart Taq polymerase (Roche) along with 4 ng primers for HPV 16 PCR. HPV 16 PCR followed a touch-down program with two sets of separate thermal cycling, an initial heating at 95°C for 5 mins and a final elongation at 72°C for 5 mins. The first set consisted of 14 cycles of amplification at 95°C, 61°C and 72°C for 30 secs each, while the second set consisted of 23 cycles of amplification at 95°C, 55°C and 72°C for 30 secs each. The annealing temperature of the first set was reduced by 0.5°C per cycle. The primer sequences are provided in Table 3.1. 5 μl of the reaction product was electrophoresed on 2% agarose gel to check amplified bands of expected size [HPV16: 116bp (Figure 3.3)]. A negative control (water instead of template...
DNA) was included for each set of PCR and positive control DNA from Caski/SiHa cells were used for amplification specific for HPV16. The primer sequences and amplicon lengths for PCR related to HPV screening and typing have been provided in Table 3.1.

Figure 3.3: Gel electrophoresis of HPV16 E6 PCR products. Lane 4: 50 bp ladder; Lanes 3, 5-8: HPV16 (+)ve samples showing specific band size at 116 bp; Lane 9: HPV 16 (-)ve sample; Lane 2: positive control; Lane 1: negative control

3.3.4 Estimation of HPV16 load and the E2 status in cervical cancer samples

HPV16 positive samples were used for viral load assessment by amplifying a small region within the E6 gene. For such assays, 7900 HT Real Time platform (Applied Biosystems) was employed to conduct Taqman hydrolysis probe based real time PCR (absolute quantification) technology (chemistry: 5'-nuclease activity of Taq polymerase cleaves DNA-bound probe and releases flurochrome from quencher, both tagged with the probe at opposite ends). E6, being a stable viral oncogene, was targeted for this assay. Real time PCR of GAPDH gene was performed to ensure the integrity of the DNA samples. The DNA samples used for this study all portrayed the presence of GAPDH gene amplicon.

(A) Determination of E6 copy numbers: 100 ng of HPV16 positive genomic DNA was used for qRT-PCR in a reaction mixture of 20 μl containing 25 ng of primers (Peitsaro et al, 2002) in Taqman Universal Master Mix (PE Applied Biosystems, Perkin-Elmer). The primer sequences, PCR conditions and amplicon sizes are summarized in Table 3.2 of primer sequences and conditions. Standard curve plot (R² = 0.99) of threshold cycles (Ct)
against log (copy number), were generated using serial dilutions of $1.75 \times 10^9$, $1.75 \times 10^7$, $1.75 \times 10^5$ and $1.75 \times 10^3$ copies of the HPV16 plasmid insert (pUC19 plasmid vector with HPV16 ref. sequence insert = 10.582 kb; concentration = 20 x10³ ng/μl) along with 100 ng genomic DNA (placental DNA). Viral copy number per 100 ng of genomic DNA was calculated by the following formula and was then log transformed (natural log):

\[
\text{Number of copies per } \mu l = \frac{\text{amount of plasmid in gm}}{\text{weight of one molecule of plasmid in gm}} = 1.75 \times 10^{12} \text{ copies} = 1.75 \times 10^{12} \text{ copies/μl} \quad [\text{Amount of plasmid in gm} = 20 \times 10^3 \times 10^{-9}]
\]

\[1\text{kb of double stranded DNA}= 6.5 \times 10^5 \text{ Daltons};\]
\[\text{Avogadro's number}= 6.023 \times 10^{23}; \text{1 Dalton} = (1/6.023 \times 10^{23}) \text{ gm}\]

\[\text{[Weight of one molecule of plasmid in gm} =10.582 \times 6.5 \times 10^5/6.023 \times 10^{23}\]

Three non-template controls (NTC) were used in each assay to check for the chances of non-specific amplification. The viral copy number values (per 100 ng of genomic DNA) of the unknown samples were interpreted from the standard curve, which showed a linear relationship between the threshold cycle values plotted against the log of the copy numbers, over the entire range of dilutions. The slope of the standard curve was used to determine the efficiency [Efficiency = $10^{(-1/\text{slope})-1}$] of the PCR reaction. An assay in which 90-100% efficiency was obtained was considered for analysis. Each such assay was done at least twice, with three replicates per sample in each assay (Figure 3.4).
Figure 3.4: Representative Amplification plot of HPV16 E6 transcript and Standard Curve Plot for viral copy number estimation. Efficiency ~ 100%; \( R^2 = 0.99 \)

(B) Determination of E2 copy numbers: In general, disruption of E2 gene releases the repressive effect of E2 protein on viral oncogene (E6 and E7) expression, leading to host cellular transformation. The most commonly reported (Arias-Pulido et al, 2006) region that harbours majority of the disruption-events is within the HPV16 E2 gene, coding for the hinge region of the E2 protein. A small portion (position on HPV16 genome: 3361 to 3442; amplicon length: 82bp) of this region was targeted to determine the presence and copy numbers of the E2 gene by qRT-PCR (Peitsaro et al, 2002) keeping efficiency of the assay around 100%. HPV16 positive genomic DNA (100 ng) was used as template DNA in a reaction-mixture of 20 μl containing 25 ng of primers. The PCR-program used and the method of estimation of E2 copy number was the same as that used for viral E6 copy number estimation. The primer sequences, PCR conditions and amplicon sizes are summarized in Table 3.2 of primer sequences and conditions. Copy numbers of E2 were determined from the standard plot \( (R^2=0.98) \) and log transformed (natural log), which was obtained in the same way as in case of qRT-PCR of E6. Representative amplification and standard curve plots have been depicted in Figure 3.5.
Figure 3.5: Representative Amplification plot of HPV16 E2 transcript and Standard Curve Plot for E2 copy number estimation. Efficiency ~ 100%; R² = 0.98

3.3.5 RNA isolation

Total RNA was isolated from fresh cervical tissue stored in RNALater using the Qiagen RNeasy mini kit according to the manufacturers’ protocol and subjected to DNase-treatment followed by re-purification. RNA isolation from cervical tissues was done by RNeasy mini kit (Cat # 74104). Tissues stored in -80°C were equilibrated to 37°C and approximately 25mg of each tissue was taken and immediately placed in liquid nitrogen, followed by grinding thoroughly with a mortar and pestle. After grinding, 500 μl of Buffer RLT was added and the tissue was homogenized using a mortar and pestle, until it was uniformly homogeneous. The lysate was directly added into a QIAshredder spin column (Cat # 79654) placed in a 2 ml collection tube and centrifuged for 2 mins at full speed. The filtrate was then collected and centrifuged for 3 mins at full speed. The supernatant was carefully removed by pipetting, and transferred to a new microcentrifuge tube for subsequent RNA isolation. One volume of 70% ethanol was added to the cleared lysate and mixed immediately by pipetting. The lysate was transferred to an RNeasy spin column placed in a 2 ml collection tube and centrifuged for 15 secs at 8000 x g (10,000 rpm). The flow-through was discarded and 700 μl of Buffer RW1 was added to the Spin Column and centrifuged at 8000 x g (10,000 rpm) for 15 secs. The flow-through was discarded and 500 μl of Buffer RPE was added to the column without wetting the rim and centrifuged at 8000
x g (10,000 rpm) for 15 secs. The flow-through was discarded and again 500 μl of Buffer RPE was added to the column without wetting the rim and centrifuged at 8000 x g (10,000 rpm) for 2 mins. The flow-through was discarded. The Column was then placed in the same 2 ml collection tube and centrifuged at full speed for 1 min to eliminate any possible carryover of Buffer RPE. The column was then placed in a clean 1.5 ml microcentrifuge tube and 40 μl RNase free water (Qiagen) was added. The column was then incubated at room temperature for 1 min, and centrifuged at 6000 x g (8000 rpm) for 1 min to elute the RNA. The RNA was then mixed with 1X DNase I buffer and 1 U of DNase I enzyme (Fermentas) and incubated at 37°C for 30 mins to remove the DNA contamination. The RNA was then cleaned up in RNeasy spin column and eluted by adding RNase free water, which was used for cDNA synthesis.

3.3.6 Estimation of RNA concentration and purity

(A) **Fluorometer based RNA quantitation:** For quantification of RNA samples, Qubit RNA BR Assay was used (assay range: 1 ng/μl – 1 μg/μl). The Qubit RNA BR Assay provides an accurate and selective method for the quantitation of high-abundance RNA samples. The assay is highly selective for RNA and will not quantitate DNA, protein or free nucleotides. The mode of estimation of RNA concentration is similar to that explained for DNA quantitation.

(B) **Assessment of RNA quality by Agilent Bioanalyzer 2100:** Each RNA sample was tested for RNA integrity by using Agilent 6000 Nano kit developed for RNA samples. The RNA integrity number (RIN) is a software tool designed to help scientists estimate the integrity of total RNA samples. The expert software automatically assigns an integrity number to eukaryote total RNA sample. Using this tool, sample integrity is no longer determined by the ratio of the ribosomal bands, but by the entire electrophoretic trace of the RNA sample. This includes the presence or absence of degradation products. In this way, interpretation of an electropherogram is facilitated, comparison of samples is enabled and repeatability of experiments is ensured. The representative plots obtained by running good quality or degraded RNA sample are shown in Figure 3.6.
3.3.7 cDNA preparation

C DNA was prepared using random hexamers as well as oligo- dT primers. For cDNA preparation using oligo-dT primers, one microgram of DNaseI treated total RNA from each sample was reverse transcribed using the primer (dT)$_{17}$-P3, i.e., an oligo (dT)$_{17}$-primer coupled to a linker sequence (5'-GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT-3') (Zolotukhin et al, 1996) in a 20 µl reaction mixture. In brief, each RNA sample was mixed with 400 ng of oligo-(dT)-P3 primer and incubated at 70°C for 10 minutes. The mix (10 µl) was quickly chilled on ice and then mixed with equal volume of a mixture of 2X reverse transcriptase buffer, 8 mM dNTPs (with DTT), 1X random primers, 20 U RNase inhibitor and 50 U MultiScribe™ reverse transcriptase (High capacity cDNA Reverse Transcription kit, Applied Biosystems) and reverse transcribed at 25°C for 10 mins followed by 37°C for 120 mins and followed by heat inactivation at 85°C for 2 mins. Reverse transcription reaction, with mRNA and all reagents but no reverse transcriptase, was performed for the samples as negative controls. All cDNA samples were stored at –20°C for long term usage in small aliquots.
3.3.8 Identification of integration status of HPV16 in cervical samples

Amplification of Papillomavirus Oncogenic Transcript (APOT) was used primarily to get amplified products of the entire region from E7 to E2 using total cDNA formed with an oligo-dT primer (dT)_{17}-P3. Subsequently, Taqman assay employing probe-primer sets for E7 and E4 was done to (i) check the intactness of viral genomes and (ii) to quantify the differences, if any, between E7 and E4 expression among CaCx samples with episomal and integrated HPV16 genomes. **Table 3.3** of primer sequences shows the probe-primer sequence and protocol for the PCR following the methodology established in our laboratory (Das Ghosh *et al.*, 2012). E7 is always (both in episomal and integrated conditions) found to be intact owing to its role in oncogenicity, and it does not harbour any splice junction site. So, quantification of the E7 region from the viral cDNA pool marks viral presence irrespective of its genomic status (episomal or integrated). On the other hand, intact E2 gene (formed by particular splicing) retains E4 and represses oncogenic transcription. Such a form of E2 holds the hinge-coding region (E2 protein has 3 parts – transactivation domain, DNA binding domain and a hinge in between) within the E4. Hinge-coding region reportedly harbours the majority of the disruption sites for viral integration into the host genome. So, quantification of the hinge-coding region of E4 from the viral cDNA pool marks the presence of an intact E2 gene. So probe-primer sets were designed for E7 (Kalantari *et al.*, 2008) and E4 for Taqman-based qRT-PCR on the APOT nested PCR products (**Figure 3.7**).

![Figure 3.7: APOT coupled Taqman assay](image.png)

*Figure 3.7: APOT coupled Taqman assay.* The left panel shows presence of E4 in episomes and absence of E4 (replaced by cellular DNA) in integrated samples. The right panel shows positions of E7- and E4-specific qRT-PCR primers.
A detail of the protocol is as follows:

(A) **APOT assay:** For the first strand cDNA synthesis using (dT)$_{17}$-P3 primer, the reaction mixture (20 µl) contained ≤ 1 µg RNA template, 400 ng primers, 5X RT buffer [Fermentas; 250 mM Tris-HCl (pH 8.3 at 25°C), 250 mM KCl, 20 mM MgCl$_2$, 50 mM DTT], 10 mM dNTP, 0.1 M DTT and 200 U of MMuLV reverse transcriptase (Fermentas)]. The thermal cycler programme for the reaction included a heat-denaturation at 70°C for 10 mins, quick chilling on ice followed by reverse transcription at 42°C for 60 mins and inactivation at 90°C for 5 mins. For the 1st PCR with P1 and P3, the reaction mixture (20 µl) contained the 5X diluted cDNA, 0.25 µM primers, the same 5X RT buffer (Fermentas), 0.96 mM MgCl$_2$, 0.2 mM dNTP and 1.5 U thermostable DNA polymerase (Roche). The 2nd PCR with P2 and (dT)$_{17}$-P3 was nested within the amplicon of the 1st PCR and had a similar reaction mix (Figure 3.8). The primers and PCR programmes are provided in Table 3.3.

![Gel electrophoresis picture showing P2-(dT)$_{17}$-P3 products.](image)

**Figure 3.8:** Gel electrophoresis picture showing P2-(dT)$_{17}$-P3 products. Lanes 4, 7 and 9: samples showing presence of 1050 bp band-size indicating E2-intact viral genome; Lanes 2, 3, 5, 8, 11, 12: samples without specific band-sizes, but integrated HPV16 derived bands were present; Lane 1: negative control; Lane 6: 1 kb ladder.

(B) **Taqman assays:** Taqman-based quantitative Real Time PCR was performed for E7 and E4 regions with similar PCR conditions. The reaction mix (10 µl) contained 2X TaqMan® Universal PCR Master mix (Applied Biosystems; includes AmpliTaq® DNA polymerase and AmpErase® UNG), 3 ng primers and 2 µM probe. The primers and PCR
programmes are provided in Table 3.3. The presence of the particular repressor splice variant of E2 was checked by gel electrophoresis of P2-(dT)_{17}-P3 products, where presence of band sizes of 250 bp, 870 bp and 1400 bp represent the transactivator E2 (unspliced form), whereas, presence of 530 bp, 610 bp and 1050 bp represent repressor E2 splice variant. Of these, the band size of 1050 bp is considered as specific for repressor E2-splice variant coded from intact E2 gene (Figure 3.8). Presence of bands of lengths other than these would indicate presence of integration derived transcripts (Klaes et al, 1999). Different sizes of episome-derived bands appear, because of different polyA stretches (present between the splice-donor and splice acceptor sites and also after the splice acceptor site) that can bind (dT)_{17}-P3 yielding different sizes of first strand cDNAs. Differential amplification of E7 and E4 in samples with episomal and integrated HPV16 genomes could be indicated by differences in Ct-values of E7- and E4-specific qRT-PCR (Figure 3.9).

![Figure 3.9: APOT-cum-Taqman assay based confirmation of HPV16 integration status of the cervical cancer samples.](image)

(a) A representative amplification plot showing both E7 and E4 expression among CaCx samples with episomal HPV16 genome. (b) A representative amplification plot showing only E7 expression and no E4 expression among CaCx samples with integrated HPV16 genome.

### 3.4 GENE EXPRESSION ANALYSIS OF VIRAL AND HOST TRANSCRIPTS

#### 3.4.1 Quantitation of HPV16 E7 expression by Taqman based qRT-PCR assay

Expression of HPV16 E7 gene was also estimated by qRT-PCR analysis. The reaction mixture in each case (10 μl) contained 2 μl cDNA, 50 ng primers corresponding
to E7, 0.1 μM probe and 2X TaqMan® Universal PCR Master Mix, (Applied Biosystems Inc, USA). Human 18S rRNA as Endogenous Control (FAM Probe, Primer Limited) (Applied Biosystems Inc, USA) was used as internal control (Figure 3.10). The primer and probe sequences, PCR conditions and amplicon sizes are summarized in Table 3.3. The negative controls consisted of an aliquot from a reverse transcription reaction (i) containing all reagents except mRNA and (ii) including mRNA and all reagents but no Reverse Transcriptase. The PCR-controls were NTC (non-template control) as well as separate aliquots from Reverse Transcription reactions with (i) all reagents except mRNA, (ii) mRNA and all reagents but no Reverse Transcriptase, and (iii) HPV-negative cellular mRNA.

Figure 3.10: Representative Taqman amplification plots of HPV16 E7 and 18s rRNA. (a) Amplification plot of HPV16 E7 (b) Amplification plot of 18s rRNA.

3.4.2 Quantitation of host transcript expression by SYBR Green based qRT-PCR assay

For real-time PCR of Power SYBR Green based estimation of host transcripts, the reaction mix comprised of 10-30 ng of forward and reverse primers, 1 μl undiluted cDNA, 5 μl Power SYBR Green PCR Master Mix (Applied Biosystems), and the reaction volume was made up to 10 μl using nuclease-free water. The real time PCR program included initial denaturation at 95°C for 10 mins, followed by 40 cycles of denaturation at 95°C for 15 secs and annealing at 60°C for 1 min. Dissociation curve analysis was done, in order to rule out
the occurrence of nonspecific amplification and primer dimer formation. The PCR controls were NTC (non-template control) as well as separate aliquots from Reverse Transcription reactions with (i) all reagents except mRNA, (ii) mRNA and all reagents but no Reverse Transcriptase. Each assay was performed at least thrice, with three replicates per sample in each assay, on MicroAmp optical 96-well plates using a 7900 HT PCR System (ABI). Relative expressions of the host transcripts were calculated using GAPDH as the endogenous control, and calibrated to the control samples. The details of primer sequences, PCR conditions and amplicon sizes are summarized in Table 3.4. Representative amplification plots and dissociation curves are depicted in Figure 3.11 and Figure 3.12.
Figure 3.11: Representative amplification plots of host transcripts under investigation (HOXA13, HOXC8, HOXC9, HOXC11, HOXC12, HOXD10, HOTAIR, HOXA11-AS1, HOTTIP, EZH2, SUZ12, E-Cadherin and Vimentin) and endogenous control GAPDH.
Figure 3.12: Representative dissociation curves of host transcripts under investigation (HOXA13, HOXC8, HOXC9, HOXC11, HOXC12, HOXD10, HOTAIR, HOXA11-AS1, HOTTIP, EZH2, SUZ12, E-Cadherin and Vimentin) and endogenous control GAPDH.
3.4.3 Quantitation of miRNA expression by SYBR Green based qRT-PCR assay

Power SYBR Green based miRNA assays for the selected miRNAs were undertaken, employing cDNA prepared from total RNA samples, using specific miRNA stem loop primers and reagents from TaqMan miRNA Reverse Transcription Kit (ABI; Cat#4366596). The 15 µl reverse transcription reactions consisted of 100 ng of total RNA, 5 U MultiScribe Reverse Transcriptase, 0.5 mM of each dNTP, 1X reverse transcription buffer, 4 U RNase inhibitor, and nuclease-free water. This was performed at 16°C for 30 mins and at 42°C for 60 mins, followed by termination at 85°C for 2 mins. For real-time PCR of Power SYBR Green based miRNA assays, the reaction mix comprised of 0.5 µl miRNA assay primer, 0.5 µl universal reverse primer (stem loop specific), 1.5 µl undiluted cDNA, 5 µl Power SYBR Green PCR Master Mix (Applied Biosystems) and 2.5 µl nuclease-free water. The real time PCR program included initial denaturation at 95°C for 10 mins, followed by 40 cycles of denaturation at 95°C for 15 secs and annealing at 60°C for 1 min. Dissociation curve analysis was done, in order to rule out the occurrence of nonspecific amplification and primer dimer formation. The PCR controls were NTC (non-template control) as well as separate aliquots from Reverse Transcription reactions with (i) all reagents except mRNA, (ii) mRNA and all reagents but no Reverse Transcriptase. Each assay was performed at least thrice, with three replicates per sample in each assay, on MicroAmp optical 96-well plates using a 7900 HT PCR System (ABI). Relative expressions of the miRNAs were calculated using miR-127 (miRNA control assay which showed uniform expression across all samples) as the endogenous control, and calibrated to the control samples. The details of stem loop primer sequences, miRNA assay primer sequences, PCR conditions and amplicon sizes are summarized in Table 3.5. Representative amplification plots and dissociation curves are depicted in Figure 3.13.
Figure 3.13: Representative amplification plots and dissociation curves of miR-22 and endogenous control miR-127. (a)-(c) Amplification plots of miR-22 and miR-127, respectively. (b)-(d) Dissociation curves of miR-22 and miR-127, respectively.

3.5 MICROARRAY BASED GLOBAL GENE EXPRESSION PROFILING

Microarray based transcriptome analysis of host genes was done using samples from HPV negative controls (n=11), HPV16 positive non-malignant samples (n=11) and CaCx cases (n=20) employing Illumina HumanHT-12_v4 Expression BeadChip array based platform. The raw data was generated by the GenomeStudio software as text file formats and analysed by the Bioconductor based “limma” package.

3.5.1 Gene expression assay work plan

1) **Selection of RNA Samples:** RNA was isolated using Qiagen RNeasy mini kit and was quantitated using Qubit RNA BR assay kit. The samples were then checked for quality.
using the Agilent Bioanalyzer Nano Chip for RNA integrity analysis and to estimate the RIN value. Only samples which had RIN value of 6 and above, were selected for further work.

2) **RNA amplification and gene expression array**: After quantitation, the RNA was amplified using the Illumina TotalPrep RNA Amplification kit, which was used to convert RNA into cDNA, followed by synthesis of biotinylated RNAs by in vitro transcription. Subsequently, the cRNA was quantified using Nano Drop-8000 and samples with a minimum concentration of 150 ng/µl were further used for the expression profiling, with the requisite amount of 750 ng.

Based on the above requirements, the selected samples were used for gene expression profiling, using appropriate gene expression array platform. The entire workflow is depicted in **Figure 3.14**.

![Figure 3.14. Work flow of transcriptome analysis by microarray based method](image)

3.5.2 Microarray data analysis

**Input files required**

1) **Sample probe profile file**: This contained the non-normalized summary values as output by GenomeStudio. In this file each row is a different probe in the experiment and the columns give different measurements for the gene. For each array, the summarized expression level (AVG Signal), standard error of the bead replicates (BEAD STDERR),
number of beads (Avg NBEADS) and a detection p-value (Detection Pval) was recorded, which estimated the probability of a gene being detected above the background level.

2) **Control probe profile**: This file contained the summarized data for all of the controls on each array, which was useful for calibration purposes.

**Reading of input files in bioconductor based limma package**

The input files or image files were read by “read.ilmn” function in this package.

**Preprocessing and quality assessment of the probes**

Negative control probes, which measure background signal on each array, were used to assess the proportion of expressed probes that are present in the samples. The proportion of expressed probes was estimated, by comparing the empirical intensity distribution of the negative control probes with that of the regular probes. A mixture model was fitted to the data from each array, to infer the intensity distribution of expressed probes and estimate the expressed proportion.

**Background correction and normalization**

The normal-exponential convolution model was used in background correction of Illumina probe expression data, which simplified the parameter estimation process for background parameters in this model. By applying this approach, the data was found to be improved in terms of bias-variance trade-off and reduced false positives. The “neqc” function in “limma” used such a convolution model for the intensities from each sample, before quantile normalization and log2 transformation of the data, to standardize the signal between samples.

**Probes annotation and filtering**

Filtering non-responding probes from further analysis can improve the power to detect differential expression. The probes with multiple genomic matches or matches to non-transcribed genomic locations are likely to be unreliable and hence those were removed.
The “illuminaHumanv4.db” annotation package within the “AnnotationDbi” package provides access to the reannotation information. This information was used as a basis for filtering promiscuous or un-informative probes from further analysis.

Four basic annotation quality categories (‘Perfect’, ‘Good’, ‘Bad’ and ‘No match’) were defined, which correlated with expression level and measures of differential expression. The probes assigned as ‘Bad’ or ‘No match’ quality score, were removed after normalization, which is similar to removing lowly-expressed probes, but with the additional benefit of discarding probes with a high expression level caused by non-specific hybridization.

**Batch effect correction**

The batch effects were corrected to remove the technical variability within the experiments.

**Differential gene expression analysis**

The differential expression method in the “limma” package was used to identify differentially expressed genes, by fitting a linear model, to summarize the values from the arrays. The functions “lmFit”, “contrasts.fit”, “eBayes” were applied to the normalized and filtered data. Finally, the “decideTests” function was used for selecting probes that show evidence for differential expression after correcting for multiple testing.

**3.5.3 Pathway analysis**

Subsequently, pathway analysis based on the differentially expressed genes was done by employing the Ingenuity Pathway Analysis software (IPA). The objective was to identify the significantly enriched and biologically relevant pathways in cervical cancer pathogenesis by comparing CaCx cases with HPV negative control samples and HPV16 positive non-malignant samples.
3.6 IMMUNOBLOT ANALYSIS OF VIRAL AND HOST PROTEINS

Homogenization of the tissue samples (~10 mg) was done in 100 µl ice cold protein lysis buffer (30 mM Tris HCl; pH 7.5, 1 mM MgCl₂, 1 mM EGTA, 0.67% β-mercaptoethanol, 0.5% CHAPs, 10% Glycerol and 0.5% Triton X-100). After overnight incubation at 4°C with shaking, and subsequent centrifugation at 12,000 rpm at 4°C for 20 minutes, the supernatant was collected. The protein concentration was determined using Bradford assay. Protein samples (30 µg) were run on 12.5% SDS PAGE in duplicate, transferred onto 0.45 µm PVDF membrane. After blocking with 5% skimmed milk, the membrane was treated with 1:200 dilution of EZH2 primary antibody (Santa Cruz Biotechnology, sc-25383), 1:200 dilution of SUZ12 primary antibody (Santa Cruz Biotechnology, sc-67105), 1:200 dilution of E7 primary antibody (Santa Cruz Biotechnology, sc-6981) overnight at 4°C. After washing, the membrane was again treated with respective secondary antibodies (1:5000 dilution, goat anti-mouse IgG-HRP, Santa Cruz Biotechnology, sc-2050 or goat anti-rabbit IgG-HRP, Santa Cruz Biotechnology, sc-2030) at 37°C for 2 hours. Protein expression was detected by chemiluminiscence based assay, after washing the membrane. Expression of GAPDH was taken as loading control. Mouse monoclonal GAPDH primary antibody (1:5000 dilution, Abcam, ab9485) and anti-mouse secondary antibody (1:5000 dilution, goat anti-mouse IgG-HRP, Santa Cruz Biotechnology, sc-2005) were used for GAPDH protein expression study. Densitometric analysis of the western blot results were performed using ImageLab Software (BioRad).

3.7 CLONING AND EXPRESSION OF HPV16 E7 INTO MAMMALIAN CELLS

3.7.1 Cloning of HPV16 E7 into mammalian expression vector pcDNA3.1(+)

Mammalian expression vector pcDNA3.1(+) was purchased from Invitrogen and used for cloning of HPV16 E7 ORF. The primers used for synthesizing the insert, the amplicon size and the PCR conditions are all listed in Table 3.6. The generated amplicons were PCR purified using the Qiagen PCR purification Kit (Cat # 28104). KpnI (NEB, Cat # R0189S) and NotI (NEB, Cat # R0142S) restriction enzymes were used for inserting the amplified fragments and the digested products were ligated using T4 DNA Ligase (Roche, Cat # 10716359001), by incubating overnight at 16°C. The plasmid was confirmed by double digestion and checking for the insert followed by transformation into DH5α cells
and plasmid isolation. The plasmid pcDNA3.1-HPV16 E7 was checked for insertion of the E7 fragment into the mammalian expression vector by Sanger sequencing.

![Schematic representation of pcDNA3.1(+)-HPV16E7 plasmid](image)

**Figure 3.15: Schematic representation of pcDNA3.1(+)-HPV16E7 plasmid.** *KpnI* and *NotI* restriction enzymes were used for inserting the HPV16 E7 ORF.

### 3.7.2 Cell culture

HPV negative cell line C33A and Caski cells were a kind gift by Professor Sudhir Krishna (Cellular Organisation and Signalling, National Centre for Biological Sciences, Bangalore, India) and SiHa cell line was gifted by Dr Madhumita Pal Roy (Chittaranjan National Cancer Institute, Kolkata, India). The cells were cultured in DMEM (Gibco, Cat# 11995-065) supplemented with 10% FBS (Gibco, Cat# 16000-044), 50 Units/ml of penicillin and 50 µg/ml of streptomycin (Gibco, Cat# 15070-063), at 37°C and 5% CO₂.

### 3.7.3 Transfection of plasmids into cell lines

C33A cells were transfected using Lipofectamine 2000 reagent (Invitrogen, Cat# 11668-027) as per the manufacturer’s protocol, using 1 µg of plasmid pcDNA3.1-HPV16 E7. The cells were harvested at three time points: 24, 48 and 72 hrs post-transfection of cells with HPV16 E7. Empty vector pcDNA3.1(+) was used for mock. The cells were washed with 1X PBS, pH 7.4 (Gibco, Cat# 14190-144), trypsinized and collected by centrifugation at 300 x g for 10 minutes. The transfected cells were used further for RNA and protein isolation. The transfection experiments were repeated thrice, with each
condition (untransfected, mock transfected and transfected) in triplicates.

### 3.8 SEQUENCING OF HOTAIR GENE TO IDENTIFY CERVICAL CANCER ASSOCIATED SEQUENCE VARIATIONS

#### 3.8.1 Deep sequencing of HOTAIR gene

**A) PCR based amplification of the target region and library preparation:** Long PCR amplification of the HOTAIR gene was done using two sets of primers having an overlap of ~1kb (Figure 3.16).

![Figure 3.16: Strategy for overlapping primer design for HOTAIR deep sequencing.](image)

The primer sequences, amplicon length and PCR conditions used for generating fragments using the Expand Long Template PCR enzyme mix (Roche) are listed in Table 3.7. The amplicon concentrations were estimated using the Qubit® dsDNA BR Assay Kit (Life Technologies, Invitrogen Division, Darmstadt, Germany).
Figure 3.17: Representative HOTAIR gene amplification products for Ion Torrent amplicon sequencing on 0.8% agarose gel. (a) Lane 1: No Template Control (NTC)PCR; Lanes 2-3: 6613 bps product of HOTAIR genes; Lane 4: GeneRuler 1kb ladder. (b) Lane 1: No Template Control (NTC)PCR; Lanes 2-8: 7868 bps product of HOTAIR genes; Lane 9: GeneRuler 1kb ladder.

The amplicons were pooled in equimolar amounts and 1µg of the pooled amplicon was sheared using S220 focused ultrasonicator (Covaris) to generate fragments ranging from 50-500 bps. The fragments were used for barcoded library preparation using the Ion Plus Fragment Library Kit and Ion Xpress™ Barcode Adapters 1–16 kit (Cat# 4471250) and Ion Xpress™ Barcode Adapters 17-32 kit (Cat# 4474009). The adapter ligated libraries were then size-selected using E-Gel® SizeSelect™ 2% Agarose gel in the E-Gel® iBase™ unit (Life Technologies (Invitrogen), Cat# G6610-02). The size-selected fragments were assessed for library quality using Agilent High Sensitivity DNA Kit (Agilent Technologies, Cat# G2938-90321) employing the Agilent Bioanalyzer™ (Agilent Technologies) to confirm fragment size around 330 bp (200 bp library +barcoded adapters) as shown in Figure 3.18.
Figure 3.18: Bioanalyzer™ instrument analysis of a 200 base-read library. Peaks at 35 and 10380 bp represent low and high molecular weight markers. Peaks around 330 bp represent size selected adapter ligated 200 base–read library (200 base read library + adapter = 330 bp) used for amplicon sequencing.

The libraries were then quantified by qRT-PCR using Ion library Quantitation Kit (Cat# 4468802) and the dilution (Template Dilution Factor) required for template preparation (PCR-mediated addition of libarary molecules onto Ion Sphere™ Particles) was estimated subsequently. The template preparation using the Ion OneTouch™ System.

The barcoded library pool was then clonally amplified onto the proprietary Ion Sphere™ particles and clonal amplification was accomplished by emulsion PCR (emPCR) using Ion Xpress™ Template 200 Kit (Cat# 4471253). The Ion Sphere™ particles coated with template were applied to the Ion 314™ Chip and sequencing was done by the Ion PGM™ 200 Sequencing Kit (Cat# 4474004) in Ion PGM™ sequencer according to manufacturer’s protocol.

Once data was generated on the Ion PGM™ sequencer, it was automatically transferred to the Torrent Server, where the data were run through signal processing and base calling algorithms that produce the DNA sequences associated with individual reads.

(B) Quality checking and identification of variations in HOTAIR gene by Ion Torrent algorithm: The depth of coverage needed to ensure enough statistical confidence to make accurate mutation calls, depend on the expected frequency of the mutation within the sample. The mean coverage of depth was 310X and mean length of the libraries were 233
bp (Figure 3.19), which was acceptable for evaluating the evidence for a variant in Torrent variant caller algorithm.

**Figure 3.19: Read length histogram of Ion Torrent data.** The mean read length was estimated by taking an average of the read lengths observed for all samples, which can be confirmed by the observation of a peak around 230 bps.

The reads generated from each sample was aligned with the HOTAIR gene sequence (gi|568815586: 53968396-53975008, *Homo sapiens* chromosome 12, GRCh 38 Primary Assembly), as reference. The binary alignment and the indexing files (BAM and BAI file format) were then downloaded from torrent server and visualized in IGV v4.4 (Broad Institute). The quality of the reads and depth of coverage was calculated to ensure enough statistical confidence, to make accurate variant calls within the sample. The variants were then analysed by Torrent variant caller algorithm (DiBayes Algorithm) for evaluating the evidence for a variant by looking at a pileup of the reads.

### 3.8.2 Sanger sequencing based validation of findings from deep sequencing of HOTAIR gene

The respective region harbouring the SNP rs2366152 was amplified using the sets of primers described in Table 3.7 and the fragment generated was analysed by gel electrophoresis (Figure 3.20). Next, 25 ng of generated amplicons were sequenced using the ABI 3500 XL genetic analyser (Life Technologies).
Figure 3.20: Representative agarose gel (2%) visualisation of HOTAIR gene amplification products (spanning rs2366152) for Sanger sequencing. Lane 1: No Template Control (NTC) PCR; Lanes 2-6: 482 bps product of HOTAIR genes; Lane 7: GeneRuler 50 bp ladder.

3.9 RNA IMMUNOPRECIPITATION

C33A (untransfected and HPV 16 E7 transfected) cells were harvested by trypsinization and resuspended in 2 ml PBS, 2 ml nuclear isolation buffer (1.28 M sucrose; 40 mM Tris-HCl pH 7.5; 20 mM MgCl2; 4% Triton X-100), and 6 ml water on ice, for 20 mins (with frequent mixing). Nuclei were pelleted by centrifugation at 2,500 x g for 15 mins. Nuclear pellet was re-suspended in 1 ml RIP buffer (150 mM KCl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP-40, Protease Inhibitor Cocktail, 100 U/ml RNaseOUT Ribonuclease Inhibitor; Life technologies). Re-suspended nuclei were split into two fractions of 500 µl each (for Mock and IP) and were mechanically sheared using a dounce homogenizer with 15–20 strokes. Nuclear membrane and debris were pelleted by centrifugation at 13,000 rpm for 10 mins. Antibody to HPV16 E7 (25 µg) or SUZ12 (15 µg) was added to supernatant for pull down along with mock controls and incubated for 2 hrs at 4°C with gentle rotation. Protein A Agarose beads (40 µl) were added and incubated for 1 hr at 4°C with gentle rotation. Beads were pelleted at 2,500 rpm for 30 secs, the supernatant was removed, and beads were re-suspended in 500 µl RIP buffer and repeated for a total of three RIP washes, followed by one wash in PBS. Beads were re-suspended in 1 ml of Trizol. Co-precipitated RNAs were isolated, followed by SYBR green based qRT-PCR for detection of HOTAIR transcript. Protein isolated by the beads was detected by
western blot analysis, to confirm the specificity of the pull-down (Figure 3.21).

![Image: Western Blot](image.png)

**Figure 3.21: Western Blot confirming the specificity of RNA Immunoprecipitation.**

### 3.10 CHROMATIN IMMUNOPRECIPITATION-QPCR

SiHa and C33A cells were cultured as explained above. SiHa cells, untransfected and HPV16 E7 expressing C33A cells were used for comparison of the H3K4me3 and H3K27me3 profiles of the HOX cluster members, showing deregulated expression among cancers. The cells were cross-linked using formaldehyde at 1% final concentration and incubated at room temperature for 10 mins. The cross-linking was quenched using glycine at a final concentration of 0.125 M and incubated at room temperature for 5 mins, followed by washing twice with 1X PBS. The cells were harvested by scraping in 1X PBS (with Protease Inhibitor cocktail) and pelleting the cells by centrifugation at 300 x g for 5 mins, at 4°C. The cells were re-suspended in 1ml of SDS lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl pH 8.1 and protease inhibitors before use) and incubated on ice for 15 mins. The lysate was sonicated at 30% amplitude, 14 X 10s impluses, 20s pauses followed by centrifugation at 14,000 rpm at 4°C. The lysate was pre-cleared with 80 µl of 50% slurry of Protein A Agarose beads for 60 mins at 4°C, followed by centrifugation at 4,000 rpm for
1 min at 4°C. Specific antibody (5µg) was added to the supernatant and precipitated overnight at 4°C with rotation. After overnight rotation, 60 µl of blocked Protein A Agarose beads (1 ml 50% bead slurry, 0.5 mg BSA) were added to the antibody/protein complex followed by rotation for 60 mins at 4°C. The beads were pelleted by centrifugation at 1,000 rpm for 1 min at 4°C, followed by sequential washing with 1X Low Salt Immune Complex wash Buffer (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl), 1X High Salt Immune Complex wash Buffer (0.1% SDS, 1% TritonX-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 500 mM NaCl), 1X LiCl Immune Complex wash Buffer (0.25 M LiCl, 1% NP-40, 1% Deoxycholic acid sodium salt, 1 mM EDTA, 10 mM Tris-HCl pH 8.1) and 1X TE (10 mM Tris pH 8.1, 1 mM EDTA). The beads were eluted in freshly prepared elution buffer (1% SDS, 0.1 M NaHCO₃) and incubated at room temperature for 30 mins with agitation. The DNA was reverse cross-linked by adding 12 µl of 5 M NaCl per 100 µl elute and incubation at 65°C overnight, followed by RNase A treatment and purified using PCR purification kit.

The purified DNA was then used for SYBR green based qPCR estimation for identification of enrichment of H3K4me3 and H3K27me3 marks in the promoter regions of selected HOX cluster genes, which show altered expression in cancers, based on our microarray based analysis. The primers used for the qPCR enrichment study are listed in Table 3.7. The amplification plots and dissociation curves for the qPCR reactions are provided in the Figure 3.22 and Figure 3.23, respectively. The primers targeting RPL-30 and EVX-1 promoters were used as positive control for H3K4me3 and H3K27me3 enrichment, respectively. While, ZNF333-3' served as a negative control for H3K4me3 and H3K27me3 enrichment, which did not show any amplification.
Figure 3.22: Representative amplification plots of gene promoters under investigation (HOXA13, HOXB13, HOXC11, HOXD10, RPL30, EVX1 and ZNF333-3’).
Figure 3.23: Representative dissociation curves for checking the specificity of amplicons from the gene promoter regions under investigation (HOXA13, HOXB13, HOXC11, HOXD10, RPL30, EVX-1 and ZNF333-3').
3.11 IN SILICO ANALYSIS

3.11.1 RPISeq

RPISeq is a family of machine learning classifiers for predicting RNA-protein interactions, using only sequence information (http://pridb.gdcb.iastate.edu/RPISeq/). RPISeq predictions are based on Random Forest (RF) or Support Vector Machine (SVM) classifiers trained and tested on two non-redundant benchmark datasets of RNA-protein interactions (RPI2241 and RPI369) extracted from PRIDB, a comprehensive database of RNA-protein complexes extracted from the PDB.

3.11.2 catRAPID

catRAPID is an algorithm to estimate the binding propensity of protein-RNA pairs (http://service.tartaglialab.com/page/catrapid_group). By combining secondary structure, hydrogen bonding and van der Waals contributions, catRAPID predicts protein-RNA associations with great accuracy. The catRAPID web server provides users with distinct modules to run their predictions: catRAPID graphics, catRAPID fragments, catRAPID strength, catRAPID omics, catRAPID express.

3.11.3 ImageLab (BioRad)

Densitometric analysis of each band of protein (considered for the study) in immunoblots, were performed using ImageLab™ software.

3.11.4 RNAsnp

RNAsnp algorithm predicts the effect of SNPs on large (>1000 nts) RNA sequences (http://rth.dk/resources/rnasnp/). This tool was employed with a folding window of 200 and keeping the length over which the local structural changes should be evaluated to 20.
3.11.5 SNPfold

SNPfold is an algorithm that computes the potential effect of Single Nucleotide Polymorphisms (SNPs) on RNA structure and it was used to analyze the effect of the SNPs in HOTAIR gene, on RNA structure and stability (http://ribosnitch.bio.unc.edu/snpfold/SNPfold.html).

3.11.6 IncRNASNP

IncRNASNP is a database of SNPs in IncRNAs and their potential functions in humans and mouse (http://bioinfo.life.hust.edu.cn/IncRNASNP). The database provides information about any GWAS based associations identified for a particular SNP. The database also predicts changes in local secondary structure and probable gain or loss of miRNA binding sites, which are predicted based on the TargetScan and miRanda.

3.11.7 Ingenuity Pathway Analysis (IPA)

IPA is used to analyze data derived from gene expression experiments and others that generate gene lists, in order to gain insight into molecular and chemical interactions, cellular phenotypes, and disease processes within a system. This tool was used to identify pathways significantly associated with CaCx pathogenesis.

3.11.8 ENCODE database

The Encyclopedia of DNA Elements (ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active (https://www.encodeproject.org/).
3.12 STATISTICAL ANALYSIS

3.12.1 Hypothesis testing

Statistical tests are used in arriving at decisions concerning populations on the basis of sample information. *Statistical Hypothesis* is defined as assertion or conjecture about the parameter or parameters of a population. They may also concern the type, nature or probability distribution of the population. *Null Hypothesis* states that there is no difference between two procedures and is denoted by $H_0$. The null hypothesis is always tested, i.e., we want to either accept or reject the null hypothesis. *Alternative Hypothesis* states that there is a difference between the procedures and is denoted by $H_A$. *Test Statistic* is the random variable $X$, whose value is tested to arrive at a decision.

3.12.2 Kolmogorov-Smirnov test (K-S test)

This is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test). The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution (in the two-sample case) or that the sample is drawn from the reference distribution (in the one-sample case). Such analysis was carried out using SPSS v16.0 for windows.

3.12.3 Mann-Whitney U test

The Mann-Whitney $U$ test evaluates whether the medians on a test variable differ significantly between two groups. It is used to test whether two independent groups of observations are drawn from the same or identical distributions. An advantage with this test is that the two groups under consideration may not necessarily have the same number of observations or follow normal distribution. For the current study, this test was done to compare whether two medians were same or not for two distributions considered for analysis. Mann-Whitney U test was carried out using SPSS v16.0 for windows.
3.12.4 Two-sample t-test

This test was used for comparing the means of two groups, where the variables under consideration followed normal distribution with equal variances. In simple terms, this test compares the actual difference between two means in relation to the variation in the data (expressed as standard deviation of the difference between the means).

The $t$ statistic to test whether the means are different can be calculated as follows:

$$t = \frac{\bar{X} - \bar{Y}}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$, $s_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$, and $s_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$.

The above test statistic follows a $t$ distribution with $n_1 + n_2 - 1$ degrees of freedom, under the null hypothesis that the two means are equal. This test was carried out using SPSS windows v16.0.

3.12.5 Linear regression

When the response variable is continuous in nature and depends on other variables, linear regression is performed to see the influence of the explanatory variables on the response variables. The regression coefficient is a constant that represents rate of change of the response variable, as a function of changes in the explanatory variable,

$$y_i = \alpha + \beta x_i + e_i, \ i = 1, 2, \ldots, n$$

where associated with the $i$-th unit, $y_i$ is the response variable, $x_i$ is the explanatory variable, $e_i$ is the error assumed to follow a normal distribution with mean 0 and variance $\sigma^2$; $\alpha$ and $\beta$ are the parameters to be estimated through least square method.
3.12.6 Test for association (Chi-square test)

A null hypothesis (\(H_0\)), relevant to this study was formulated as a specific variable had no association with the disease. To obtain the statistical significance of the difference between observed proportions of a dichotomous variable in two groups, a 2×2 contingency Table was used for \(\chi^2\) test. In this study, the relevant dichotomous categories were risk and non-risk and the two groups were the cases (affected) and controls (unaffected). The data were first cross tabulated in classes. The null hypothesis to be tested was that - The risk variable is independent of the group (cases or controls). To test this null hypothesis, the observed frequency, \(O_i\), was compared with the expected frequency, \(E_i\), which was calculated assuming that the null hypothesis was true. The test statistic was,

\[
X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}
\]

where,
\(X^2\) = Pearson's cumulative test statistic, which asymptotically approaches a \(\chi^2\) distribution.
\(O_i\) = an observed frequency;
\(E_i\) = an expected (theoretical) frequency, asserted by the null hypothesis;
\(n\) = the number of cells in the Table.

The chi-square statistic was then used to calculate a p-value, by comparing the value of the statistic to a chi-squared distribution.

3.12.7 Multiple testing corrections

The incidence of false positives is proportional to the number of tests performed and the critical significance level (p-value cut-off). The chance that false positives are going to be sampled is higher, when more SNPs are applied for statistical testing. This is why it is important to correct the p-value of each SNP when performing a statistical test on a group or SNPs. Multiple testing correction adjusts the individual p-value for each SNP to keep the overall error rate (or false positive rate) to less than or equal to the user-specified p-value cut-off or error rate. Benjamini and Hochberg False Discovery Rate (FDR) was used as correction for multiple testing. Among a list of rejected hypotheses, FDR can control the expected proportion of incorrectly rejected null hypotheses (that is, type I errors). It is a less conservative procedure for comparison.
3.12.8 Receiver Operating Characteristic (ROC) Curve Analysis

Receiver operating characteristic (ROC), or ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity or the sensitivity index. The false-positive rate is also known as the fall-out and can be calculated as (1–specificity). The area under the curve (AUC) is estimated and then the point showing the best possible combination of sensitivity and 1-specificity is selected as the cut-off point to group clinical samples into two sub-categories. Although ROC analysis is a good tool for bilateral distribution of a set of values, there are certain limitations of this analysis which include the requirement of large sample size to allow for the effects to be real and significant.
Tables of primer sequences and their respective PCR conditions

Table 3.1: Primer Sequences for Sample categorisation based on their HPV Status

<table>
<thead>
<tr>
<th>Target Name</th>
<th>Primer Sequence (5'-3')</th>
<th>Product Length (bps)</th>
<th>PCR Conditions</th>
</tr>
</thead>
</table>
| L1          | Forward: GCM CAG GGW CAT AAY AAT GG  
Reverse: CGT CCM ARR GGA WAC TGA TC | 450 | No of cycles: 15  
Denaturation: 95 °C/30 s  
Annealing: 57 °C/2 min  
Extension: 72 °C/1 min 30 s |
| GP 5/6      | Forward: TTG GTT ACT GTG GTA GAT ACT AC  
Reverse: GAA AAA TAA ACT GTA AAT CAT ATT C | 150 | No of cycles: 40  
Denaturation: 95 °C/30 s  
Annealing: 42 °C/1 min  
Extension: 72 °C/30 s |
| HPV16       | Forward: TCA AAA GCC ACT GTG TCC TG  
Reverse: CGT GTT CTT GAT GAT CTG CA | 116 | No of cycles: 14  
Denaturation: 95 °C/30 s  
Annealing: 61 °C/30 s  
Extension: 72 °C/30 sec  
No of cycles: 24  
Denaturation: 95 °C/30 s  
Annealing: 55 °C/30 s  
Extension: 72 °C/30 sec |
| HPV18       | Forward: ACC TTA ATG AAA AAC CAC GA  
Reverse: CGT CGT TGG AGT CGT TCC TG | 100 | No of cycles: 40  
Denaturation: 94 °C/30 s  
Annealing: 53 °C/30 s  
Extension: 72 °C/30 s |

Table 3.2: Primer and probe sequences for APOT-cum-Taqman assay

<table>
<thead>
<tr>
<th>Target Name</th>
<th>Primer Sequence (5’-3’)</th>
<th>Product Length (bps)</th>
<th>PCR Condition</th>
</tr>
</thead>
</table>
| HPV 16 E7 – E5 (701 – 4218; includes a Splice junction) | \( P1: \) CGG ACA GAG CCC ATT ACA AT \( P3: \) GAC TCG AGT CGA CAT CG | Variable (depending on polyA stretch) | Number of cycles: 30  
Denaturation: 94°C / 30 s  
Annealing: 59°C / 30 s  
Elongation: 72°C / 4m |
| | \( P2: \) CCT TTT GTT GCA AGT GTG ACT CTA CG \( (dT)17P3: \) GAC TCG AGT CGA CAT CGA TTTTTTTTTTTTT T | (Depends on polyA stretch) | Number of cycles: 30  
Denaturation: 94°C / 30s  
Annealing: 67°C / 30s  
Elongation: 72°C / 4m |
| | | | |
| HPV 16 E4 (3439 - 3556) | Forward: AAG TGT GAC TCT ACG CTT CCG TT  
Reverse: GCC CAT TAA CAG GTC TTC TTC CAAA  
Probe: 5’-FAM-TGC GTA CAA AGC ACA CAC GTA GAC ATT CGT A -BHQ -3’ (Kalantari et al, 2008) | 78 | UNG-activation: 50°C / 2 m  
Initial denaturation: 95°C / 10m  
Number of cycles: 40  
Denaturation: 95°C / 15s  
Annealing: 60°C / 1m |
| | | | |
Table 3.3: Primer and probe sequences for estimation of HPV16 load

<table>
<thead>
<tr>
<th>Target Name</th>
<th>Primer Sequence (5'-3')</th>
<th>Product Length (bps)</th>
<th>PCR Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV16 E2</td>
<td>FP: AAC GAA GTA TCC TCT CCT GAA ATT ATT AG</td>
<td>82</td>
<td>UNG-activation: 50°C / 2 m</td>
</tr>
<tr>
<td></td>
<td>RP: CCA AGG CGA CGG CTT TG</td>
<td></td>
<td>Initial denaturation: 95°C / 10m</td>
</tr>
<tr>
<td></td>
<td>Probe: 5'-FAM-CAC CCC GCC GCG ACC CAT A-DQ-3' (Peitsaro et al., 2002)</td>
<td></td>
<td>Number of cycles: 40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Denaturation: 95°C / 15s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Annealing: 60°C / 1m</td>
</tr>
<tr>
<td>HPV16 E6</td>
<td>FP: GAG AAC TGC AAT GTT TCA GGA CC</td>
<td>81</td>
<td>UNG-activation: 50°C / 2 m</td>
</tr>
<tr>
<td></td>
<td>RP: TGT ATA GTT GTT TGC AGC TCT GTG C</td>
<td></td>
<td>Initial denaturation: 95°C / 10m</td>
</tr>
<tr>
<td></td>
<td>Probe: 5'-VIC-CAG GAG CGA CCC AGA AAG TTA CCA CAG TT-DQ-3' (Peitsaro et al., 2002)</td>
<td></td>
<td>Number of cycles: 40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Denaturation: 95°C / 15s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Annealing: 60°C / 1m</td>
</tr>
</tbody>
</table>

Table 3.4: Primer sequences for gene expression estimation of host transcripts

<table>
<thead>
<tr>
<th>Target Name</th>
<th>Primer Sequence (5'-3')</th>
<th>Product Length (bps)</th>
<th>PCR condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOTAIR/RT</td>
<td>Forward: GGT CCT GCT CCG CTT CGC AG</td>
<td>116</td>
<td>UNG-activation: 50°C / 2 m</td>
</tr>
<tr>
<td></td>
<td>Reverse: ACG CCC CTC CTT CCT CTC GC</td>
<td></td>
<td>Initial denaturation: 95°C / 10m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of cycles: 40</td>
</tr>
<tr>
<td>Gene</td>
<td>Forward</td>
<td>Reverse</td>
<td>Product Size (bp)</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>HOXA11-AS1/RT</td>
<td>GGC TGC CCT GCT GCT CA</td>
<td>TGG ACC TCC CAG GAG AAG TCT</td>
<td>137</td>
</tr>
<tr>
<td>HOTTIP/RT</td>
<td>AGC CAC AAG CAG GTA CCA CAA AGC</td>
<td>GTG GGG TGG AGG TCT CCT TGC T</td>
<td>140</td>
</tr>
<tr>
<td>HOXA13/RT</td>
<td>AGT CGC GCC ACG AAC CCT TG</td>
<td>ATG GGA GAC CAC GTC GGG CA</td>
<td>149</td>
</tr>
<tr>
<td>HOXC8/RT</td>
<td>ATG GAT GAG ACC CCA CGC TCC</td>
<td>TCG GTC AGT CCC AGG GCA TGA G</td>
<td>150</td>
</tr>
<tr>
<td>HOXC9/RT</td>
<td>CCG CAG ACA CTG CCC TCG CC</td>
<td>GCC ACG GGG TTG CTG GGG TC</td>
<td>92</td>
</tr>
<tr>
<td>Gene</td>
<td>Forward</td>
<td>Reverse</td>
<td>Forward Primers</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>---------</td>
<td>-----------------</td>
</tr>
<tr>
<td>HOXC11/RT</td>
<td>AAA GGA GCC GCC CCC AAC GC</td>
<td>CAT CCG GGA CAG CTG CAG CC</td>
<td>138</td>
</tr>
<tr>
<td>HOXC12/RT</td>
<td>CCC CGG CAG CTT GGT ATC GC</td>
<td>TGA GAG TTC CCT CCG GCG CT</td>
<td>187</td>
</tr>
<tr>
<td>HOXD10/RT</td>
<td>GCA GGA GAA GGA AAG CAA AGA GGA A</td>
<td>CGC TCG CGG GTG AGG TAC AT</td>
<td>159</td>
</tr>
<tr>
<td>EZH2/RT</td>
<td>CCC TGA CCT CTG TCT TAC TTG TGG A</td>
<td>CGT CAG ATG GTG CCA GCA ATA</td>
<td>119</td>
</tr>
<tr>
<td>SUZ12/RT</td>
<td>AAA CGA AAT CGT GAG GAT GG</td>
<td>CCA TTT CCT GCA TGG CTA CT</td>
<td>115</td>
</tr>
<tr>
<td>Target Name</td>
<td>Primer Sequence (5’-3’)</td>
<td>Product Length (bps)</td>
<td>PCR condition</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------</td>
<td>----------------------</td>
<td>----------------------------</td>
</tr>
</tbody>
</table>
| E-Cadherin/RT       | *Forward:* TAC CCT GGT GGT TCA AGC TG  
                       *Reverse:* CCT GAC CCT TGT ACG TGG TG | 128                  | UNG-activation: 50°C / 2 m  
                       Initial denaturation: 95°C / 10m  
                       Number of cycles: 40  
                       Denaturation: 95°C / 15s  
                       Annealing: 60°C / 1m        |
| Vimentin/RT         | *Forward:* AAT GGC TCG TCA CCT TCG TG  
                       *Reverse:* CAG AGA AAT CCT GCT CTC CTC G | 115                  | UNG-activation: 50°C / 2 m  
                       Initial denaturation: 95°C / 10m  
                       Number of cycles: 40  
                       Denaturation: 95°C / 15s  
                       Annealing: 60°C / 1m        |
| GAPDH/RT            | *Forward:* CAG CCT CAA GAT CAT CAG CA  
                       *Reverse:* TGT GGT CAT GAG TCC TTC CA | 106                  | UNG-activation: 50°C / 2 m  
                       Initial denaturation: 95°C / 10m  
                       Number of cycles: 40  
                       Denaturation: 95°C / 15s  
                       Annealing: 60°C / 1m        |
| Table 3.5: Primer sequences for gene expression estimation of miR-22 and miR-127 |

<table>
<thead>
<tr>
<th>Target Name</th>
<th>Primer Sequence (5’-3’)</th>
<th>Product Length (bps)</th>
<th>PCR condition</th>
</tr>
</thead>
</table>
| Loop primer for cDNA synthesis of mature miR-22 | CTC AAC TGG TGT CGT GGA GTC GCC AAT TCA GTT GAG TAA AGC TT | -                    | 16°C / 30 min  
                       42°C / 60 min  
                       85°C / 2 min     |
| miR-22/RT           | *Forward:* ACA CTC CAG CTG GGA GTT CTT CAG TGG C  
                       *Reverse:* TGG TGT CGT GGA GTC G | 76                   | Initial denaturation: 95°C / 10 min  
                       Number of cycles: 40  
                       Denaturation: 95°C / 15s  
                       Annealing: 60°C / 1m      |
### Loop primer for cDNA synthesis of mature miR-127

<table>
<thead>
<tr>
<th>Primer Sequence (5’-3’)</th>
<th>Product Length (bps)</th>
<th>PCR condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTC AAC TGG TGT CGT GGA GTC GGC AAT TCA GTT GAG AGC CAA GC</td>
<td>6613</td>
<td>Number of cycles: 30 Denaturation: 94°C / 30 s Annealing: 58°C / 30 s Elongation: 72°C / 4 min</td>
</tr>
</tbody>
</table>

### miR-127/RT

<table>
<thead>
<tr>
<th>Primer Sequence (5’-3’)</th>
<th>Product Length (bps)</th>
<th>PCR condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACA CTC CAG CTG GGT CGG ATC CGT CTG AGC</td>
<td>78</td>
<td>Initial denaturation: 95°C / 10m Number of cycles: 40 Denaturation: 95°C / 15s Annealing: 60°C / 1m Dissociation stage: 95°C/15s, 60°C/15s, 95°C/15s</td>
</tr>
</tbody>
</table>

### Table 3.6: Primer Sequences for HOTAIR sequencing

<table>
<thead>
<tr>
<th>Target Name</th>
<th>Primer Sequence (5’-3’)</th>
<th>Product Length (bps)</th>
<th>PCR condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOTAIR Seq1</td>
<td>Forward: CCC TGC CAG GTC TAG CG Reverse: GTC TAG ACA ATA GAT GGC TGT</td>
<td>6613</td>
<td>Number of cycles: 30 Denaturation: 94°C / 30 s Annealing: 58°C / 30 s Elongation: 72°C / 4 min</td>
</tr>
<tr>
<td>HOTAIR Seq2</td>
<td>Forward: AGC TAA TCA ATC CAA AAG GAA Reverse: TCT ATT ATG GGT ACA TTG TTG</td>
<td>7868</td>
<td>Number of cycles: 30 Denaturation: 94°C / 30 s Annealing: 58°C / 30 s Elongation: 72°C / 4 min</td>
</tr>
<tr>
<td>HOTAIR 10</td>
<td>Forward: AAC TTC CTC CTG CTA TTA AGA Reverse: TCT ATT ATG GGT ACA TTG TTG</td>
<td>482</td>
<td>Number of cycles: 30 Denaturation: 94°C / 30 s Annealing: 56°C / 30 s Elongation: 72°C / 1 min</td>
</tr>
</tbody>
</table>
### Table 3.7: Primer Sequences for ChIP-qPCR:

<table>
<thead>
<tr>
<th>Target Name</th>
<th>Primer Sequence (5’-3’)</th>
<th>Product Length (bps)</th>
<th>PCR condition</th>
</tr>
</thead>
</table>
| HOXA13/TS  | *Forward:* CCA AGC CTC CAG TGG AAT GT  
*Reverse:* CCT TCG GCT TCG ATG GAG TT | 119 | UNG-activation: 50°C / 2 m  
Initial denaturation: 95°C / 10m  
Number of cycles: 40  
Denaturation: 95°C / 15s  
Annealing: 60°C / 1m |
| HOXB13/TS  | *Forward:* AGG AAG GGT ACC GGG AAT GA  
*Reverse:* TCA GAG GAG AAC CCG GGA AT | 129 | UNG-activation: 50°C / 2 m  
Initial denaturation: 95°C / 10m  
Number of cycles: 40  
Denaturation: 95°C / 15s  
Annealing: 60°C / 1m |
| HOXC11/TS  | *Forward:* CAG CTC TGG CCA GGA AAA GA  
*Reverse:* ACT TCT TTC CCC ACT GCC TG | 141 | UNG-activation: 50°C / 2 m  
Initial denaturation: 95°C / 10m  
Number of cycles: 40  
Denaturation: 95°C / 15s  
Annealing: 60°C / 1m |
| HOXD10/TS  | *Forward:* CGC GTA GTA GAT GTC GCT GT  
*Reverse:* ACC ACA TGA CAA CCA AGC CA | 127 | UNG-activation: 50°C / 2 m  
Initial denaturation: 95°C / 10m  
Number of cycles: 40  
Denaturation: 95°C / 15s  
Annealing: 60°C / 1m |
| RPL-30/TSS | *Forward:* CAA GGC AAA GCG AAA TTG GT  
*Reverse:* GCC CGT | 73 | UNG-activation: 50°C / 2 m  
Initial denaturation: |
<table>
<thead>
<tr>
<th></th>
<th>Forward:</th>
<th>Reverse:</th>
<th>Number of cycles:</th>
<th>Denaturation:</th>
<th>Annealing:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EVX1/TSS</strong></td>
<td>TCA GTC TCT TCG ATT</td>
<td>ACC GTC ACC GAG ATC CTC AT</td>
<td>40</td>
<td>95°C / 10m</td>
<td>60°C / 1m</td>
</tr>
<tr>
<td></td>
<td><strong>Forward:</strong> ACC GTC ACC GAG ATC CTC AT</td>
<td><strong>Reverse:</strong> GCA GGA CGC TGT TCT TGT TG</td>
<td>118</td>
<td>UNG-activation: 50°C / 2 m</td>
<td></td>
</tr>
<tr>
<td><strong>ZNF333-3'</strong></td>
<td><strong>Forward:</strong> GAA GAT GAA CGA GCC CGT GA</td>
<td><strong>Reverse:</strong> CTT CTC CTG CAC TTC GGG AC</td>
<td>139</td>
<td>UNG-activation: 50°C / 2 m</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Initial denaturation:</strong> 95°C / 10m</td>
<td><strong>Number of cycles:</strong> 40</td>
<td><strong>Denaturation:</strong> 95°C / 15s</td>
<td><strong>Annealing:</strong> 60°C / 1m</td>
<td></td>
</tr>
</tbody>
</table>