INTEGRATIVE SYSTEMS GENOMIC ANALYSIS OF VARIANTS BY WHOLE EXOME SEQUENCING IN CERVICAL CANCER

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Microbiology
By

Santosh Kumari Duppala
Registration Number: 41800824

Supervised By

Dr. Ashish Vyas (12386)
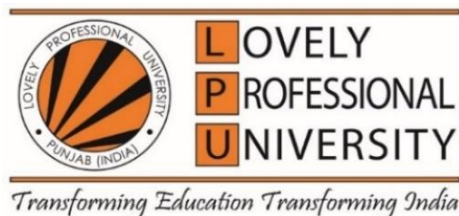**Department of Microbiology (Professor & HOD)**
Lovely Professional University

<table>
<tr><td>Co-Supervised by</td><td>Co – Supervised by</td></tr>
<tr><td>Dr. Sugunakar Vuree</td><td>Dr. Smita C. Pawar</td></tr>
<tr><td>Department of Foundation for Research (Scientist)</td><td>Department of Genetics (Prof.& HOD)</td></tr>
<tr><td>Indo-American Cancer Hospital, HYD</td><td>Osmania University, HYD</td></tr>
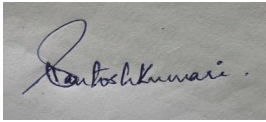</table>

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB
2024

DECLARATION

I hereby declare that the thesis entitled, **"Integrative Systems Genomic Analysis of Variants by Whole Exome Sequencing in Cervical Cancer"** in fulfilment of degree of Doctor of Philosophy (Ph. D.) is the outcome of research work carried out by me under the supervision of Dr. Ashish Vyas, Professor and Head, Department of Microbiology and Biochemistry, of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigators. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: Santosh Kumari Duppala

Registration No.: 41800824

Department/school: Microbiology (School of Bioengineering and Biosciences)

Lovely Professional University, Punjab, India

CERTIFICATE

This is to certify that Ms. SANTOSH KUMARI DUPPALA has completed the thesis entitled **"Integrative Systems Genomic Analysis of Variants by Whole Exome Sequencing in Cervical Cancer"** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original investigation and study. No part of the thesis has ever been submitted for any other degree or diploma at any university. This thesis is fit for the submission and the partial fulfillment of the condition for the award of **DOCTOR OF PHILOSOPHY IN MICROBIOLOGY**.

Name of the Supervisor: Dr. Ashish Vyas

Designation: Prof.& HOD, Department of Microbiology and Biochemistry

Affiliation: Lovely Professional University, Jalandhar, Punjab

Signature of Supervisor:

Name of the Co- Supervisor: Dr. Sugunakar Vuree

Designation: Senior Scientist

Affiliation: Indo-American Cancer Hospital, Telangana

Signature of the Co -Supervisor:

Name of the Co- Supervisor: Dr. Smita C. Pawar

Designation: Prof.& HOD, Department of Genetics

Affiliation: Osmania University, Telangana

Signature of Co- Supervisor:

ABSTRACT

Cervical cancer (CC) is one of the most frequent cancers among women and the fourth leading cancer for mortality worldwide. Most of the death cases are reported in developing countries like Africa and Southeast Asia. As the incidence and mortality rates increase globally, women with advanced and recurrent cancers are showing less response toward chemoradiotherapy, radiation etc., with early detection of CC tumors resulting in a better prognosis. CC is the most common malignant gynecological tumor, with few treatment options. Persistent infections caused by Human papillomavirus (HPV) can give rise to covert infections or benign hyperplasia, eventually progressing to malignancy. The pathogenesis of HPV-associated diseases involves intricate interactions between viral proteins and the host proteome. This study primarily focuses on the early oncoproteins E6 and E7, produced by high-risk HPV (HR-HPV) genotypes known for their propensity to induce malignancies, making them significant public health concerns. The investigation performing HPV genotyping using Restriction Fragment Length Polymorphism (RFLP) methods on liquid cytology samples obtained from patients with cervical lesions explicitly evaluates HPV types 6, 16, 18, and 33. In our study, HPV genotyping of cervical cancer samples through RFLP M09/M11 primers results in HPV 6,11, 16, 18 and 33. The combination of both low and high-risk HPV is predicted. System biology-based analysis is scarce in Human papillomavirus (HPV), which is associated with the expression of viral oncogenic proteins (E6 & E7) that are inactivating tumor suppressors pRb and p53 in the host genome. Phylogenetic analysis of selective high- and low-risk HPV types of cervical cancer using sequencing data generated from PAVE database provides a better understanding of evolutionary relationships and interactions between HPV and humans in causing cervical cancer. The data of HPV-26, HPV-53, HPV-73, and 82 in low risk for anal, vulvar and vaginal cancers show high risk in cervical cancer. The HPV types HPV-82, HPV-53, and HPV-26 are predicted in a single branch and are closely related. As HPV, Oncoproteins / Oncogenes play an important role in the progression and expression of the disease. Further, our study aimed to understand through system biology approaches, the HPV and host genome genes and their interactions can be studied by performing protein-

protein interaction (PPI) studies using the tools like String and Gene mania, develop the interactions that would aid in understanding the expression patterns of the diseased genes. Functional analysis of the HPV types showed that they were involved in signaling processes related to cervical cancer and its cell receptor and chemokine pathways. The EGFR signaling pathway for HPV shows interactions with Interferon 1 (IFNAR1), Laminin 5 (TMEM5), Estrogen (ESR2), Endothelin A (ECE1), ATM (ATM), VEGF (VEGFA), EGFR (EGFR), and GRP78 (HSPA5) proteins. EGFR has found the best interaction results with all inactive L1 capsid proteins using Gardasil 9 followed by Heparen sulfate. To improve patient survival and quality of life, new diagnostic and therapeutic agents are required. If CC tumors can be found at an early stage, the prognosis is much brighter. In this work, further we have performed whole-exome sequencing utilizing V5 (Illumina platform) on 5 paired-end CC samples and normal control tissue samples, and we compared the results with transcriptome studies. Data from PAN cancer research of 277 cBioportal databases, Cervical squamous cell carcinoma (CSCC), and Cervical adenocarcinoma samples were compared and validated. The top ten variants of our study include KMT2C, EPHB1, C1QTNF 9, OR4M1, MGST1, FAS, OPCML, SFXN1etc. KMT2C variations were shown to be the most vicious in this analysis. From an Indian viewpoint, we found a plethora of SNVs and mutations, including those with known, unknown, and possible effects on health. Based on our findings, the KMT2C gene is on chromosome 7 and in exon 8, and all three of the identified variants are missense, coding sequence transcripts, synonymous, non-coding transcript variants, and GnoMAD minor allele frequency <0.05. The variation at position (7:152265083, C>A, T, SNV rs1389082625, 7:152265091, rs62478356) in KMT2C is unique, potent, and pathogenic. The missense, coding transcript CIQTNF maps to chromosome 7 and displays T>C SNV rs. In addition, we performed SSCP (Single strand Conformational polymorphism) analysis on 64 samples of which 48 samples are tumor tissue samples and remaining are control samples and further confirmed them using Sanger sequencing to understand and verify the mutations. KMT2C is downregulated with a log FC value of -1.16 according to the transcriptome data analysis performed and heatmap created on the human Clariom D

Affymetrix platform. Further, we have also analysed the data using various databases such as TCGA, GEO, NCBI, GEPIA, Ualcan, dbSNP, GTex, SNP nexus, Clinvar, GEO methylation, etc. Understanding emerging harmful mutations from an Indian viewpoint is facilitated by our bioinformatics-based, extensive correlation studies of WES analysis and transcriptome. Potentially harmful and new mutations were found in our preliminary analysis. Among these is the gene KMT2C, also known as CIQTNF, which was found to be altered in ten different cases of CC with an Indian phenotype.

Further, *in silico* analysis of KMT2C for protein modeling and SNV variation for KMT2C rs ID 138908625 exon 8 regions Chr 7:152265083 variation C>A, T, protein structure prediction, c score, and TM value evaluated for wild type and also 5 top models of KMT2C by I TASSER docking for homology modeling of the KMT2C variant to identify the 3D structure and binding affinities to get the best structure. The predicted values of the models of KMT2C Query 1 show better binding affinities when compared to and wild type. Further, protein-protein docking studies were performed using Cluspro 2.0 with the compounds Artemisinin, Shikonin, Sitoindosides IX, Bucidarasin A, and Betulin with KMT2C. Betulin shows better binding energy (-12.5 Kcal/mol) and is followed by Bucidarasin (-12.3Kcal/mol) with KMT2C. The present study is the combination of *in silico* work with the whole exome sequencing variants, that can be used in the prognosis and diagnosis of cervical cancer. The docking studies predict the molecular binding affinities of the ligand and the protein fold conformations.

Our study on WES and transcriptome lays the foundation to comprehend the underlying cause of disease and the Genes/ mutations responsible for the cause. Early identification and biomarkers are crucial in determining the cancer stage, enabling clinicians to determine the specific single nucleotide variants (SNVs) responsible and make therapy recommendations based on the patient's clinical data. This study represents the first known application of Whole Exome Sequencing (WES) and transcriptome study to identify mutations or variants in CSCC and adenocarcinoma among patients in South India. The findings revealed the presence of multiple genes and pathways that are commonly mutated

in various other types of tumors as well. These findings will assist in directing future research and developing specific treatments for this malignancy on a global scale.

Acknowledgement

वक्रतुण्ड महाकाय सूर्यकोटि समप्रभ।
निर्विघ्नं कुरु मे देव सर्वकार्येषु सर्वदा ॥

Completing a PhD entails a protracted and arduous journey, including years of concentrated research, experimental work, writing, and rigorous evaluation. It was not an easy task as there are several ups and downs throughout the journey of my PhD.

I expand my deepest gratitude to God almighty, the source of all wisdom and strength, for guiding me through this journey. In moments of challenge, your grace sustained me and in moments of doubt, your faithfulness reassured me, in moments of inspiration your light illuminated my path. I am profoundly thankful for the countless blessings bestowed upon me, enabling me to persevere and reach this milestone. This achievement is a testament to your boundless mercy and the strength you provided when faced with challenges. I offer my heartfelt thanks for the blessings, resilience, and the opportunity to pursue knowledge. May this work be a reflection of your grace and may it contribute positively to the world.

I would like to express my sincere gratitude to **Dr. Ashish Vyas**, my esteemed supervisor Professor and Head, Department of Microbiology and Biochemistry, School of Bioengineering and Biosciences, Lovely Professional University, Jalandhar Punjab. His unwavering support, invaluable guidance, and insightful feedback have been instrumental in completing my PhD thesis. In a very short time, I have learned a lot from you and would like to work further, you have been a source of inspiration in every sense and also an encouragement during my challenging times. You were always there at any time and were very quick to respond to my queries and provide me with immediate guidance and solutions. My heartfelt thanks to my supervisor, the final submission would not be possible without your constant support and guidance.

I would like to acknowledge my indebtedness, render my warmest thanks, and express my sincere gratitude to my esteemed Co-supervisor **Dr. Sugunakar Vure,** Senior Research

Constant support, patience, motivation, and belief in my capabilities have been the bedrock of my success. Your sacrifices, both big and small, have allowed me the time and space to pursue my academic goals. Your love has been my anchor and I am so grateful to have you as a partner.

My Son **Himaansh Sanapala,** you bring boundless joy and inspiration into my life. Your presence and unconditional love have been a source of motivation during both challenging and triumphant moments and infused my academic pursuits with purpose reminding me of the importance of leading an example. Your patience, and support have been a testament to your maturity beyond your age, I appreciate the sacrifices you made when my attention was occupied with my work and I want to know that your sentiments were always felt deeply in my heart.

This thesis is dedicated to my family, my father, mother, husband, and son for being my pillars of strength and for unwavering support and understanding. Your encouragement has been instrumental in my ability to balance the demands of academia and family life. Your love has been the driving force behind my accomplishments.

Thank you, **Uma Shankar Rao** and **Himaansh** for being my greatest cheerleaders and for sharing in the triumphs and challenges of this journey

Lastly, I would like to thank all whom I have not mentioned by name, but directly or indirectly helped, supported, and encouraged me to complete my work and thesis.

<div align="right">

Santosh Kumari Duppala

</div>

DEDICATED TO
MY FATHER,
MOTHER,
HUSBAND AND
SON

Table of Contents

List of Tables

List of Figures

\

Abbreviations

Symbol	Abbreviations
CC	Cervical Cancer
HPV	Human papillomavirus
HR- HPV	High Risk Human papillomavirus
LR- HPV	Low Risk Human papillomavirus
WES	Whole Exome Sequencing
WHO	World Health Organisation
GLOBOCON	Global Cancer Observatory
IARC	International Agency for Research on Cancer
RFLP	Restriction fragment ligand polymorphism
TZ	Transformation Zone
CIN	Cervical Intraepithelial Neoplasia
ASC	Atypical Squamous Cells
ACSUS	Atypical Squamous cells of undermined significance
LSIL	Low-grade squamous intraepithelial lesion
HSIL	High-grade squamous intraepithelial lesion
LEEP	Loop electrosurgical excision procedure
AIS	Adenocarcinoma in-situ
SCC	Squamous cell carcinoma
FIGO	International Federation of Gynaecology and Obstetrics
HSV	Herpes Simplex Virus
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
DES	Diethylstilbesterol
PAP	Papanicolaou test
VIA	Visual inspection with acetic acid

CT	Computed Tomography
MRI	Magnetic resonance imaging
PCR	Polymerase chain reaction
FDA	Food and Drug Administration
CDC	Centre for Disease Control
HC	Hybrid Capture
VLP	Virus like particles
MYB	Mycobilin globulin 09, 11 Primers
LCR	Long control region
NCR	Non-coding region
ORF	Open reading frames
CDK	Cyclin-Dependent Kinases
hTERT	Human Telomerase Reverse Transcriptase
CpG	5'- C- Phosphate -G-3'
miRNA	Micro Ribonucleic acid
HOTAIR	HOX antisense intragenic RNA
lncRNA	Long non coding Ribonucleic acid
VEGF	Vascular endothelial growth factor
MET	Mesenchymal Epithelial Transition
TCDP	T Cell Differentiation Protein
GAS5	Growth arrest specific 5
NGS	Next Generation Sequencing
FISH	Fluorescence in situ hybridization
HIVID	High throughput viral integration detection
WGS	Whole Genome Sequencing
RNA Seq	Ribonucleic acid Sequencing
SMRT - Seq	Single-molecule real-time sequencing technology
SNVs	Single Nucleotide Variants

SNP	**Single Nucleotide Polymorphism**
KEGG	**Kyoto Encyclopedia of Genes and Genomes**
ML	**Machine Learning**
ASR-I	**Age-Standardized rate of Incidence**
ASR-M	**Age-Standardized rate of Mortality**
CADM1	**Cell Adhesion Molecule 1**
TWIST	**Twist family BHLH transcription factor 1**
RASSF1	**Ras association domain family member 1**
TERT	**Telomerase reverse transcritase**
MGMT	**O-6- Methylguanine DNA methyltransferase**
FHIT	**Fragile Histidine Triad**
DAPK1	**Death-associated protein kinase 1**
RARB	**Retinoic acid receptor beta**
CDH1	**E- Cadherin**
P53	**Tumor Protein 53**
PIK3CA	**Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha**
PTEN	**Phosphatase and TENsin homolog deleted on chromosome 10**
STK11	**Serine/Threonine Kinase 11**
KRAS	**Kristen Rat Sarcoma Viral oncogene**
NAAT	**Nucleic acid amplification test**
LBC	**Liquid-based cytology**
RAB3C	**RAS Oncogene family**
GABRA2	**Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha2**
ZNF257	**Zinc Finger protein 257**
SOX 14	**SRY – Box Transcription Factor 14**

ROC	Receiver Operating Characteristic
AUC	Area under curve
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
LLETZ	Large loop Excision Transformation Zone
CKC	Centromere Kinetochore complex
CAR- T	Chimeric Antigen receptor T cell therapy
CSFs	Colony stimulating factor
NSI	Non-Specific immunotherapies
HIPEC	Hyperthermic intraperitoneal chemotherapy
APC	Antigen presenting cell
BRAC1	Breast cancer gene
ER α	Estrogen receptor alpha
HIC1	Hypermethylated in cancer 1
TIMP2/TIMP3	Tissue Inhibitor of Metalloproteinase
CAV1	Caveolin 1
hMLT1/2/3/6	Human mutL 1, 2, 3, 6 etc
CNVs	Copy number variations
MYOD1	Myogenic Differentiation 1
PR	Progesterone Receptor
ERBB2	Erthroblastic leukemia viral oncogene
ADGRB1	Adhesion G protein-coupled receptor B1
ARHGAP5	Adhesion Rho GTPase Activating Protein
FastQc	Quality check
BAM	Binary Alignment membrane L
BGLAP	Bone gamma carboxyglutamate protein
GSK3β	Glycogen synthase kinase 3 beta
PCNA	Proliferating cell nuclear antigen

HPGD	15-hydroxyprostaglandin dehydrogenase
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
PPI	Protein- Protein Interaction
PAGE	Polyacrylamide gel electrophoresis
MAF	Minor Allele Frequency
NCBI	National centre for biotechnology information
SSCP	Single stranded conformation polymorphism
dbSNP	Single nucleotide polymorphism database
ANNOVAR	Annotate Variation
I TASSER	Iterative Threading ASSEmbly Refinement
EBRT	External-beam radiation therapy
NRHD	Nucleosome remodelling and histone deacetylation
CADD	Computer-aided design
UCSC	University of California, Santa Cruz
SAM	Sequence alignment map
BAM	Binary alignment map
SIFT	Scale-invariant feature transform
DEGs	Diethylene Glycol
HPGD	15- Hydroxyprostaglandin dehydrogenase
IGV	Integrative genome visualization browser
TBS buffer	Tris Buffer saline
RSaI	Restriction Endonucleases
PAVE	PapillomaVirus Episteme
MEGA X	Molecular Evolutionary Genetics Analysis X
FASTA	Fast Adaptive Shrinkage Threshold Algorithm
PDB	Protein Data Bank
TNM	Tumor, node and Metastasis

FFPE	**Formalin fixed parfiffin embedded**
PBS	**Phosphate buffere saline**
qPCR	**Quantitative polymerase chain reaction**
TES	**Target enrichment system**
VCF	**Variant calling format**
GERP	**Genomic Evolutionary rate profiling**
EDTA	**Ethylenediamine tetraacetic acid**
TEMED	**Tetramethylethylenediamine**
KMT2C	**L lysine methyl transferase**
C1QTNF 9	**C1Q Tumor necrosis factor protein 9**
TIMER	**Tumor Immune estimation resource**
PARP1/2	**Poly (Adi – ribose) polymerase**
OR4M1	**Olfactory receptor 4M1**
ARVCF	**Armadillo Repeat gene deleted in Velo-Cardio facial syndrome**
NAMPT	**Nicotinamide phosphoribosyltransferase**
EPHB1	**EPH receptor B1**
FAS	**Fetal alcohol syndrome**
MGST1	**Microsomal Gluathione S transferase 1**
HS6ST3	**Heparan Sulfate 6-O-Sulfotransferase 3**
SFXN1	**Sideroflexin 1**
OPCML	**Opioid binding protein/cell adhesion molecule**
FIGNL1	**Fidgetin 1**
OR4K2	**Olfactory receptor 4K2**
KCNJ12	**Potassium Inwardly Rectifying Channel Subfamily J Member 12**
GEPIA	**Gene Expression Profiling Interactive Analysis**

COSMIC	COmmon Software Measurement International Consortium
STRING	Search Tool for Retrieval of Interacting Genes/Proteins
Gene MANIA	Multiple Association Network Integration Algorithm
KMT2D	Lysine Methyltransferase 2D
ASH2L	Set1/Ash2 histone methyltransferase complex subunit
PAXIP1	Pax interacting protein1
NCOA6	Nuclear receptor coactivator 6
KDM6A	Lysine Demethylase 6A
TRANSFEC	Transcription factor database
UALCAN	University of ALabama at Birmingham CANcer data analysis Portal
CESC	Cervical squamous cell carcinoma and endocervical adenocarcinoma
LOMETS	Local Meta-Threading Server
TM score	Template Modelling score
RMSD	Root Mean Square Deviation
Cov	Coverage
GO	Gene Ontology
MTA1	Metastasis associated protein 1
AKT	Ak strain transforming
TGFβ1	Transforming growth factor-beta

Chapter 1

Introduction

In today's world, one of the major concerns is health, and in the current scenario, there are several communicable diseases and non-communicable diseases. Diseases in their myriad forms, have been the focal point of medical inquiry, prompting researchers to delve into the intricacies of pathophysiology and molecular aberrations that underscore various health challenges. Among these diseases, Cancer emerges as a complex constellation of conditions characterized by uncontrolled cell growth, and invasion into surrounding tissues. Cancer is a multi-stage process, which includes the transformation of a normal cell into a precancerous lesion, and later to a malignant tumor. The actual cause of cancer lies between the interactions of an individual's genetic factors and carcinogens. These carcinogens are mainly categorized as physical, chemical and biological carcinogens, also listed by the International Agency for Research on Cancer (IARC), and World Health Organization (WHO). Genetic disorders caused by herited or inherited factors play a major role in cell growth to increase. It is the second-largest mortality factor worldwide. According to WHO, cancers contribute to the leading cause of mortality. There are various risk factors associated with cancer, which also include chronic infections from biological carcinogens, such as human papillomavirus (HPV), Herpes virus, hepatitis B virus, and hepatitis C virus (Sung et al., 2021). HPV infection can increase the risk of cervical cancer and a few other cancers. CC is reported as the second leading cause of death among women worldwide, which is the fourth most prevalent malignancy among all the cancers and is brought on by recurrent infections with high-risk such as HPV (Drolet et al., 2021). Further, 200 different forms of HPV are categorized into low-risk and high-risk HPV groups according to their etiology. HPV 16 and 18 are the primary causes of 65-70% of cervical malignancies worldwide. The lack of general public awareness contributes to cervical cancer. According to the WHO, a woman passes away from cervical cancer every eight minutes worldwide (Aggarwal, R. et al.,2023 and Kloku, C. A. 2015).

According to Global Cancer Observatory (GLOBCON) 2020, there were 604,127 cases of CC overall (3.1%), and there were 341,831 deaths (3.4%) across the world. In India incidence rates are 97,000 and mortality rates are nearly 60,000 per year (Poondla et al. 2021)). The average estimated rate of age standardization incidences are nearly 13.3/one lakh women. The risk factors include Chlamydia, Herpes Virus, HIV etc., and other factors like smoking, parity, alcohol, multiple sex partners, early age marriages, oral contraceptive pills, etc. 90% of the cases are due to high-risk HPV. The low-risk HPV such 6,11, causes genital warts, and benign tumors. Due to improper hygiene practices and a lack of awareness about HPV immunization, the incidence rate is more in number in low- and underdeveloped-income nations than in high-income countries. It develops into mild dysplasia and then progresses to invasive carcinoma and TNM stages I through III (Sobin. et al 2011 & Acharya et al.2023). HPV vaccination with Gardasil Quadrivalent, which is related with HPV 6, 11, 16, and 18 types, is the initial step in preventing CC (IARC, *WHO*).

Despite the advent of preventive measures, intricate molecular mechanisms driving the transition from HPV infection to CC, remain incompletely understood. In this study it embarks on an ambitious exploration utilizing a multi-faceted approach that combines HPV genotyping, whole exome sequencing (WES) and systems level of analysis to unveil the comprehensive genomic landscape in CC. However, HPV particularly HPV 16 and HPV 18 has been identified as the primary etiological factor in the development of CC. However, the precise interplay between viral genomic integration and host genomic alterations happens and by incorporating HPV genotyping through Restriction Fragment Ligand Polymorphism (RFLP), this research aims to unravel the diversity of HPV types present in CC samples, providing insights into the intricate relationship between viral genomics and the evolving tumor landscape.

Next Generation Sequencing methods like whole-exome sequencing (WES) is a modern and integrative approach for understanding several cancers like breast cancer (Huang et al., 2018), cervical cancer (Wang et al., 2023). WES, a powerful tool capturing the protein

coding regions of the genome, offers an unprecedented opportunity to decipher the somatic mutations driving cervical carcinogenesis. This study integrates WES data to identify and characterize genetic alterations and mutations to understand the mutational burden and pinpoint potential driver events in the context of HPV associated CC. WES approaches with correlation with transcriptome data and system level provides a broader perspective considering the intricate network of molecular interactions within the cellular level. Incorporating methodologies from system biology, this research seeks to unravel the complex dynamics of signaling pathways, gene regulatory networks, and host virus interactome, offering a more wide-range understanding of the biological processes underlying CC progression.

The present work on integrative system genomic analysis of variants by (WES) in cervical cancer may identify variants that are potential targets for cervical cancer. Next Generation Sequencing methods like (whole-exome sequencing) is a modern and integrative approach for understanding several cancers like breast cancer (Huang et al., 2018), cervical cancer (Wang et al., 2023). The work being done now on an integrative system genomic analysis of variants by WES in cervical cancer may help find HPV targets that are good for cervical cancer with specific integration sites.

1.1 Cervix and anatomy of the uterus

The cervix, measuring roughly 3 cm in length and 2.60 cm in width, is a cylindrical structure consisting of stroma and epithelium (Prendiville & Sankaranarayanan, 2017). It is also referred to as the birth canal. The cervix is the distal and small portion of the uterus that connects the uterus to the vagina, serving as a passage between them (**Figure 1**). The cervix is situated anteriorly among the bladder and the gut, and posteriorly between the bowel (Stolnicu et al 2021). The size and shape of the uterus, however, vary depending on factors such as a woman's age, number of pregnancies, and menstruation status. The lower portion or intravaginal segment of the structure is located at the superior part of the vagina, while the higher portion is situated in the pelvic/abdominal cavity, positioned above the

vagina. Cervical carcinoma is characterized by the existence of malignant cells in the epithelial lining of the cervix. Before the formation of cancer, these cells experience dysplasia, leading to the emergence of abnormal cells in the cervical tissue (Prendiville & Sankaranarayanan, 2017).
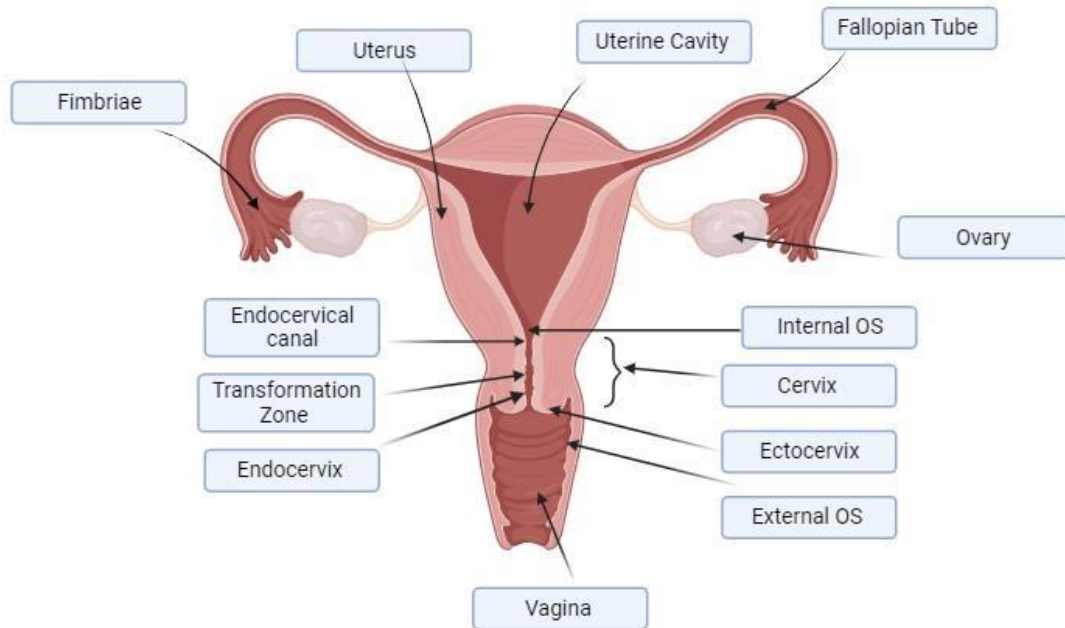
Figure 1: Structure and parts of the Reproductive system and Cervix

(The image was prepared in Biorender Software)

The cervix is divided into three parts and made-up fibromuscular tissue, they are a) Ectocervix b) Endocervix, and c) Transformation zone

a) Ectocervix

It is the outermost part of cervix is ectocervix and it is lined by epithelium which is soft and called as non-keratinized squamous epithelium (Stolnicu, et al 2021). As it is the outer part and it is visible during the cervix examination by doctors, while performing pap test.

b) Endocervix

It is the inward side of the cervix and it is lined by columnar epithelial cells and keratinized cells that secrete mucus. These cells connect to the uterus (Khorsandi, et al 2023). The cervical canal traverses the central region of the cervix, extending between the internal os to the external os (os denotes the aperture located in the middle of the cervix). The junction between the endocervix and ectocervix is the primary site of origin for most cases of cervical cancer. During a pap smear exam, both squamous and columnar epithelial cells are taken in order to identify any anomalous cellular growth.

c) Transformation Zone (TZ)

The (TZ) is located at the junction between the endocervical canals of the normal and unaffected columnar epithelium, where the normal squamous epithelium derived via the vaginal and ectocervical squamous epithelium is present (Stolnicu et al 2021). HPV infects the deepest layers of the epithelium in the (TZ), triggering the emergence of cervical intraepithelial neoplasia (CIN). The columnar epithelium of the ectocervix, which is a location where abnormalities can be discovered, is exposed to the somewhat acidic vaginal environment (Stolnicu, et al 2021) (De Tomasi, et al 2019). This exposure leads to squamous metaplasia, causing physiological changes in the (TZ).

1.2 Functions of the cervix

The cervix acts as a barrier between the vaginal canal and the uterus. The cells lining the endocervical canal create a colloidal combination of electrolytes (mainly sodium chloride) and simple carbohydrates (glycogen), known as acid and neutral mucin (Lacroix, et al 2020). The cervical mucosal secretions that fill the entrance of the cervical canal have a significant role in both limiting the ascent of pathogens, thereby protecting the developing fetus from infections, and facilitating the ascent of sperm to the fallopian tubes (Suarez & Pacey, 2006). During the ovulatory phase of the menstrual cycle, the release of cervical mucus facilitates the movement of sperm from the vagina to the uterus (Suarez & Pacey, 2006). In the event of unsuccessful fertilization, the cervix undergoes a small dilation, enabling the passage of blood flow from the uterus. During childbirth, the cervix dilates to facilitate the passage of the newborn.

1.2.1 Cervical lesions classification

In various medical centers, cervical smear samples are categorized based on the Bethesda III Classification System 2001 (Apgar et al., 2001). This system defines that the lack of cellular abnormalities is classed as negative for intraepithelial lesions. Aberrant cells are categorized based on the extent of deviation from normalcy (Table 1).

a) ASCs (Atypical Squamous Cells)

ASC implies the presence of slightly aberrant cells in the cervix, which could indicate an area of inflammation. More than 80% of these cells are in a normalized state, while 5-10% of these women have dysplasia in the form of CIN II or CIN III. It is classified into two distinct categories.

b) Cells displaying unusual characteristics that cannot be definitively determined **(ACS-US)**

Individuals with atypical squamous cells of undetermined significance (ACSUS) should be screened for HPV infection due to the substantial differences in morphology between these cells and normal cells. ACS-US cases that test positive for HPV are directed to undergo colposcopy.

c) ASC-H (Atypical Squamous Cells)

Atypical squamous cell lesion is indicative of an increased risk for the precancerous stages, specifically CIN II or CIN III. These samples are sent for colposcopic evaluation.

d) LSIL refers to a mild abnormality in the squamous cells of the epithelium.

LSIL is a minor abnormality found in a Pap test that is linked to a human papilloma infection. Pathologically, it is classified as early dysplasia cervical intraepithelial neoplasia (CIN1). The diagnosis is typically made with the use of a Pap smear or colposcopy. The excisional therapy approach involves the removal of aberrant cells using the loop electrosurgical excision procedure (LEEP) and conization.

LEEP: LSIL can be eradicated with the application of electric current or heat.

Conization: Excise the conical segment from the atypical cervix.

e) HSIL refers to a severe form of abnormal cell growth in the squamous epithelium.

HSIL positive samples are classified as moderate or severe dysplasia, sometimes referred to as CIN II, CIN2/3, CIN III. The diagnosis of this condition typically involves the use of a Pap smear, colposcopy, and hysterectomy (Swain M. 2023) which is the standard therapeutic approach for removing CIN II and CIN III lesions. For women who do not plan to have children in the future, LEEP and conization excisional treatment procedures were favored.

f) Adenocarcinoma *in-situ* (AIS)

Glandular cell abnormalities are present in the cervix. Untreated, these lesions might develop into malignancy. Colposcopy and biopsy are the recommended subsequent examinations for endocervical screening.

	AGE 21	AGE 21-29	AGE 30-65 and older (>65)	
Normal pap test	Women do their first pap test at this stage	Pap tests every 3 years	HPV negative	HPV positive
			Every 5 years, a Pap and HPV test is performed	HPV typing and co-testing are performed every 1 year
ACS-US	HPV and Pap test will be redone in every 12 months	HPV and Pap smear test repeated every12 months	Repeat HPV and pap test every 3 years	Colposcopy
LSIL	Pap test will be redone in every 12 months	Colposcopy	Pap test repeated in 12 months preferable colposcopy	Colposcopy
ACS-H	Colposcopy	Colposcopy	Colposcopy	Colposcopy

HSIL	Colposcopy	Rapidly done excision diagnosis or colposcopy	Rapidly done excision diagnosis or colposcopy	Rapidly done excision diagnosis or colposcopy and biopsy for HPV testing

Table 1: Cervical cancer screening test and recommended future follow-up strategies

1.2.2 Cervical Intraepithelial Neoplasia (CIN)

It refers to abnormal cell growth in the cervical epithelium.

Pap smear detects dysplastic or aberrant cells, which are referred to as pre-malignant or precancerous cells. The term used to refer to these cells that have the potential to become cancerous is cervical intraepithelial neoplasia (CIN) (Mutuku, 2020)

Cervical intraepithelial neoplasia (CIN) is classified into three categories according to the extent of aberrant cell growth in the epithelial tissue:

A) Low-grade neoplasia, often known as CIN 1, is a condition characterized by modest dysplasia. It is not cancerous and typically resolves spontaneously without requiring therapy. At this stage, just a few numbers of abnormal cells can be seen on the outer layer of the cervix (Markovic & Markovic 2008). The condition is classified as low-grade squamous intraepithelial lesions (LSIL).

B) In cases of **mild dysplasia (CIN 2)**, aberrant alterations affect around two-thirds of the epithelium and can develop into cancer if left untreated. Possible treatments for eradicating aberrant cells in CIN2 may encompass laser therapy, cryotherapy, and cone biopsy techniques.

C) Severe dysplasia (CIN 3) is the most highly aggressive type of dysplasia, impacting over two-thirds of the epithelium. Persistent infections of HPV cause CIN3 (Byun, Jung

Mi, et al. 2018). The likelihood of advancing to invasive cancer is greater, and if left untreated, laser therapy, cryotherapy, and cone biopsy are alternative therapeutic options for CIN3 lesions. Both CIN2 and CIN3 lesions are categorized as HSIL.

D) Squamous cell carcinoma

Carcinoma in situ cells are non-malignant cells that have the potential to transform into cancerous cells over an extended period and spread to adjacent areas. It is classified as a CIN3 category of precursor lesions and requires prompt treatment.

1.2.3 Cervical cancer that has spread beyond its original site – invasive cancer

In invasive cervical cancer, malignant epithelial cells detach from the basal layer and spread to the cervical stroma by metastasis. Cancerous growth gradually forms and disseminates to adjacent organs, such as the bladder and rectum, through the bloodstream or lymphatic system.

1. Normal

2.CIN1-Low risk

3.CIN2-moderate risk

4.CIN3-high risk

5.SCCA- High risk
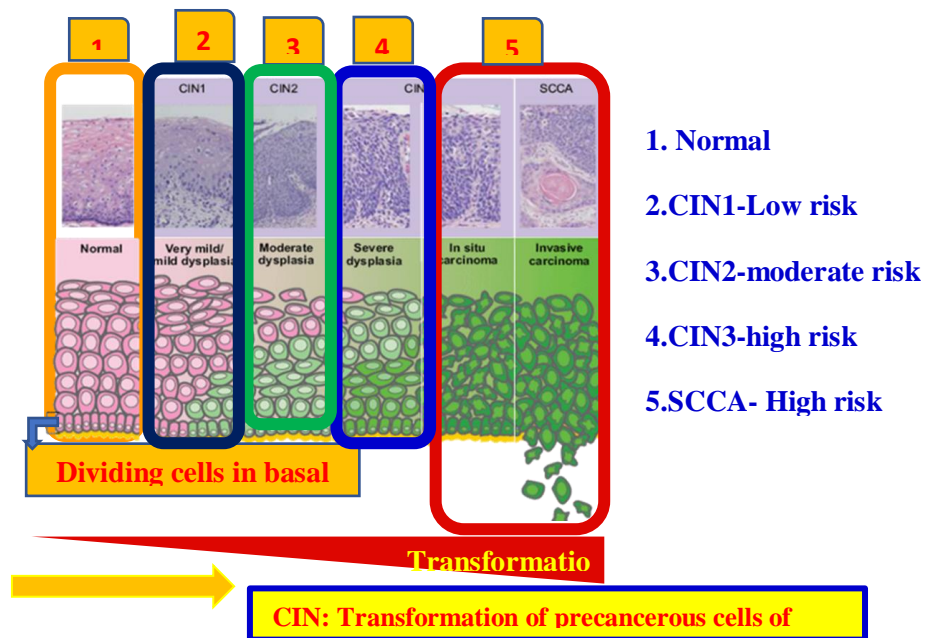
Figure 2: Transformation of precancerous cells of cervix to cancerous and their stages CIN1, CIN2, CIN3 and *in situ* carcinoma, Invasive cancer

Courtesy to Wright TC & Schiffman, *N EnglJ Med* 2003 Feb 6;348(6):489–90

1.3 The FIGO (International Federation of Gynecology and Obstetrics) staging system for cervical cancer

The FIGO staging system (2018) is a valuable tool for determining the stage of cervical cancer through clinical examination (Bhatla et al., 2019). Several diagnostic techniques employed to ascertain the stage of the disease include tissue biopsies, palpation, examination, imaging, colposcopy, intravenous urography, proctoscopy, and cystoscopy. Staging assists the physician in identifying the appropriate treatment for the patient, enhancing the chances of recovery.

Stage 0: This is classified as an in-situ carcinoma, when there are irregular epithelial cells seen in the inner membrane of the cervix.

Stage I: The cervix shows the presence of a malignant tumor on its surface, but the cancer has not metastasized to distant areas or surrounding lymph nodes.

Stage II: Malignant cells infiltrate outside the uterus, but have not metastasized to the interior of the vagina or pelvic walls. (Parikh et al. 2008).

Stage III: In this stage, the cancer may metastasize to adjacent lymph nodes, impairing kidney function and lymph nodes.

Stage IV: Cancer metastasizes beyond the pelvis and infiltrates other organs such as the rectum, bladder, and lungs.

1.4 HPV-associated Cervical Cancer Progression

HPV is the most frequent causative agent of viral infection in the reproductive tract of men and women, which mainly involves penis, vagina, cervix, and anus. Although most HPV infections are asymptomatic and non-oncogenic and resolve instinctively with a normal immune system. Based on the associated risk, sexually transmitted HPV is divided into three groups, namely LR HPV (non- oncogenic), moderate risk and high-risk HPV (oncogenic). There are approximately 14 HR HPV types (Burd, 2003).

Cervical cancer is mainly caused by human papillomavirus (HPV) infection, particularly HPV 16 and 18, which together account for about 70-73% of all cervical cancer cases worldwide. The frequency of high-grade cervical cancer ranging from 41% to 67% and 16

to 32% of low-grade lesions occur in the cervix. Additionally, 20% of CC cases are caused by other HPV types, namely 31, 33, 35, 45, 52, and 58.

HPV strains are categorized into **low-risk** and **high-risk** groups based on their potential to cause cancer or other serious conditions. The primary factor distinguishing **low-risk** from **high-risk** strains is the ability of high-risk types to cause persistent infections and significant cell changes, leading to cancer over time (Villalobos-Oyarce, R.,et al 2023).
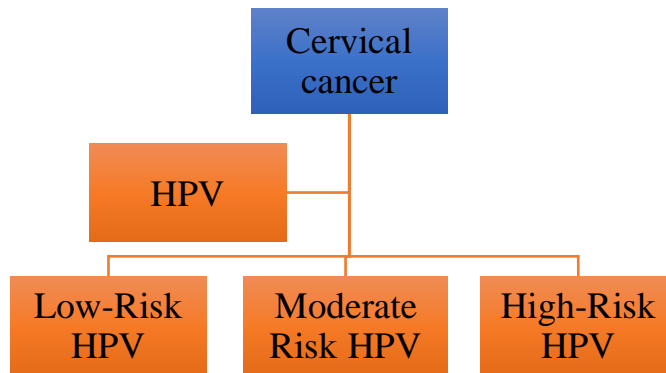
Figure 3: Schematic view of HPV categorized based on their pathogenicity

- Low risk HPV they don't cause Cervical cancer 6,11,40,42,43,44,53,54,61,72,73,81 but cause genital warts
- Moderate risk HPV 45,33,58,31,52 cause genital warts
- High risk HPV 16,18,31,45,33,35, 39,51,52,56,58,66,68,70 cause cancer

1.5 Causes of Cervical cancer

It is recognized that CC can be sexually transmitted, most commonly through vaginal, anal, or oral sex. Women are more likely to contract HPV because of a number of genetic and lifestyle factors (Arbyn et al., 2011). Multiple sexual partners, underage sex, personal history of cervical, vaginal, or vulvar dysplasia, family history of gonorrhea, chlamydia, and other STDs constitute the majority of these risk factors.

The presence of specific herpes viruses such as HSV-2, HSV-6, and HSV-7 was identified in samples taken from patients with aggressive cervical cancer (Li & Wen, 2017). Research has shown that the presence of chlamydia bacteria increases the likelihood of developing CC (Zhu et al., 2016).

Smoking increases the likelihood of developing CC due to the carcinogenic effects of cigarettes, which damage the DNA of cervical cells and weaken the immune system's ability to combat HPV infection (Sugawara et al., 2019). A weakened immune system, use of the hormone diethylstilbestrol (DES) during pregnancy, pregnancy before the age of 17, usage of oral contraceptives for more than five years, and three were carried to term (Siracusano et al., 2014).

1.6 Symptoms of cervical cancer

The majority of individuals with dysplasia or in situ cancer typically do not exhibit any initial indications or symptoms. Once the abnormal epithelial cells transform into cancerous cells, they have the potential to infiltrate the neighboring tissue and give rise to symptoms.

The primary sign of concern is abnormal or atypical bleeding, which may manifest as: Menstrual bleeding that is heavier than usual - Spotting after engaging in sexual intercourse - Increased vaginal discharge - Spotting following a pelvic exam

Additional potential risk factors encompass smoking, polyamory, multiple childbirths, and the use of oral contraceptive pills.

1.7 Assessment or diagnosis

The diagnosis of cervical cancer can be accomplished through the utilization of the following methods, as explained below:

a) Pap Test

A Pap test is a conventional screening method that can detect both precancerous and cancerous cells in the cervix. The Pap (Papanicolaou) smear, which was first established

in 1941, is a method used to spot cells from the cervix (Michalas S.P 2000). When atypical cells are detected, they are classified based on their level of abnormalities. After a Pap smear detects squamous cells or abnormal growths in the cervix, various supplementary procedures can be conducted to confirm the findings. These tests may include colposcopy, cone biopsy, imaging techniques such as CT scan and MRI, as well as DNA hybridization and polymerase chain reaction (Ramzan et al.2015).

b) Visual inspection with acetic acid (VIA) Test

VIA, a "screen-and-treat" called visually performed test that was suggested for less developed countries. The worth of VIA is that most LSIL or HSIL erosions are acetowhite (i.e. they show a white color while vinegar or 5% acetic acid is added to the cervical epithelium which might show pre-invasive lesions, the normal cervical epithelium maintains its pink color after application of 5% acetic acid (Prabhu, 2015)). VIA allows quick results and treatment can be given at proper time, eg.by Cryocoagulation. Most importantly, VIA does not need a cytopathology lab and it is available at low cost for patients and little infrastructure for doctors. However, VIA is more complicated test but it looks generally very simple aceto-whitening used for cervical finding of cells, but several women may find aceto-whitening of cells but it is not due to the CIN2+ lesions but might be due to immature squamous epithelial cells or other conditions etc. In case of insufficient quality control, VIA will lead to over treatment subjected patients who do not have CIN 2+ (Rachana L. Y. 2018).

c) HPV DNA Testing

HPV DNA testing is superior to cytology and its capacity to detect high-grade CIN with greater sensitivity. Currently, in the majority of nations that conduct screenings for cervical cancer, such as the United States, HPV DNA testing is employed as a triage test for women who have received a cytological diagnosis of atypical squamous cells of ambiguous significance (ASCUS) (Naucler et al 2009).

1.8 Colposcopy

Colposcopy is a diagnostic procedure that utilizes magnifying lenses to closely and clearly inspect the cervix. The objective is to analyze the transition zone and determine the area of irregularity and grade it based on morphological parameters (Gaddi Mahalakshmi S. 2018). Colposcopy is the parameter to detect the irregular changes in the cervix (**Figure 4).**
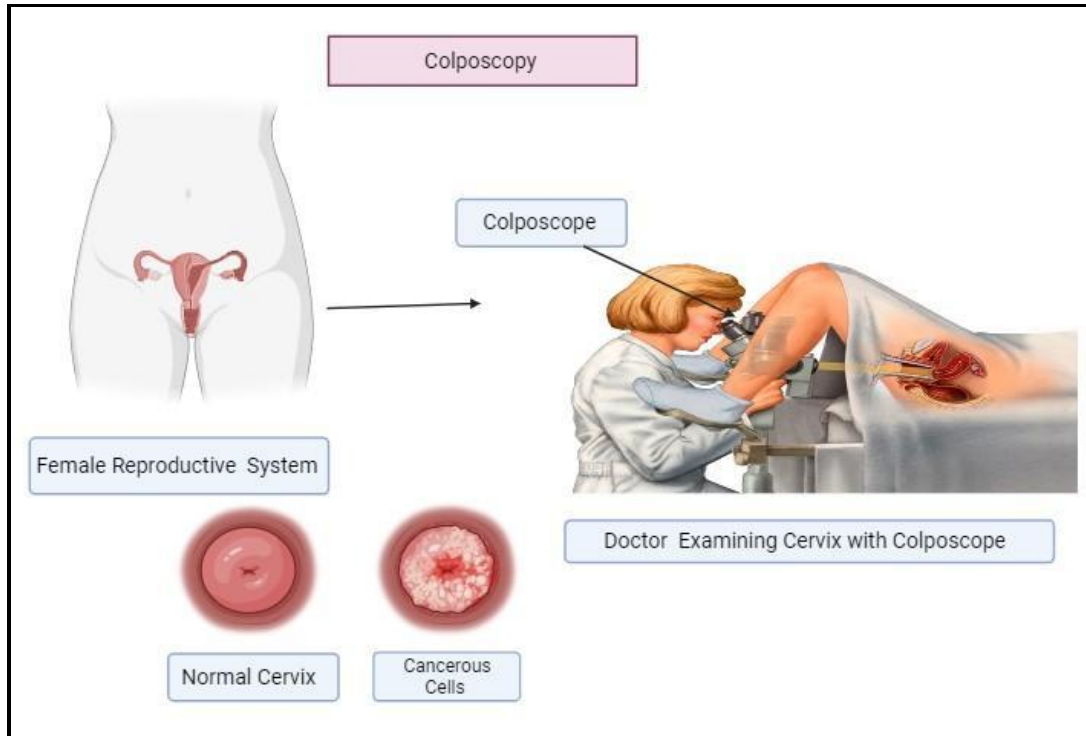
Figure 4: Colposcopy showing normal and cancerous cells. (Image self -made through Bio render)

1.9 Molecular techniques for HPV infection detection
1.9.1 Hybridization Assays for Detection of HPV

Multiple signal detecting approaches are accessible that can improve the specificity of these assays. Hybrid Capture Tm-2, Cervista, and HPV PCR using generic primers were frequently employed for the identification of genital HPV types.

The hybrid capture-2 method is the initial approach sanctioned by the FDA (Food and Drug Administration) for identifying 13 high-risk HPV varieties. This approach adheres to the subsequent steps: 1) RNA-DNA hybrids are formed by combining target HPV DNA with

RNA probes. 2) These hybrids are then immobilized onto a solid phase that is glazed with antibodies specifically designed to bind to these hybrids. 3) The signal from the captured

Schematic view of signal amplification using invader technology by Hybrid Capture or Cervista
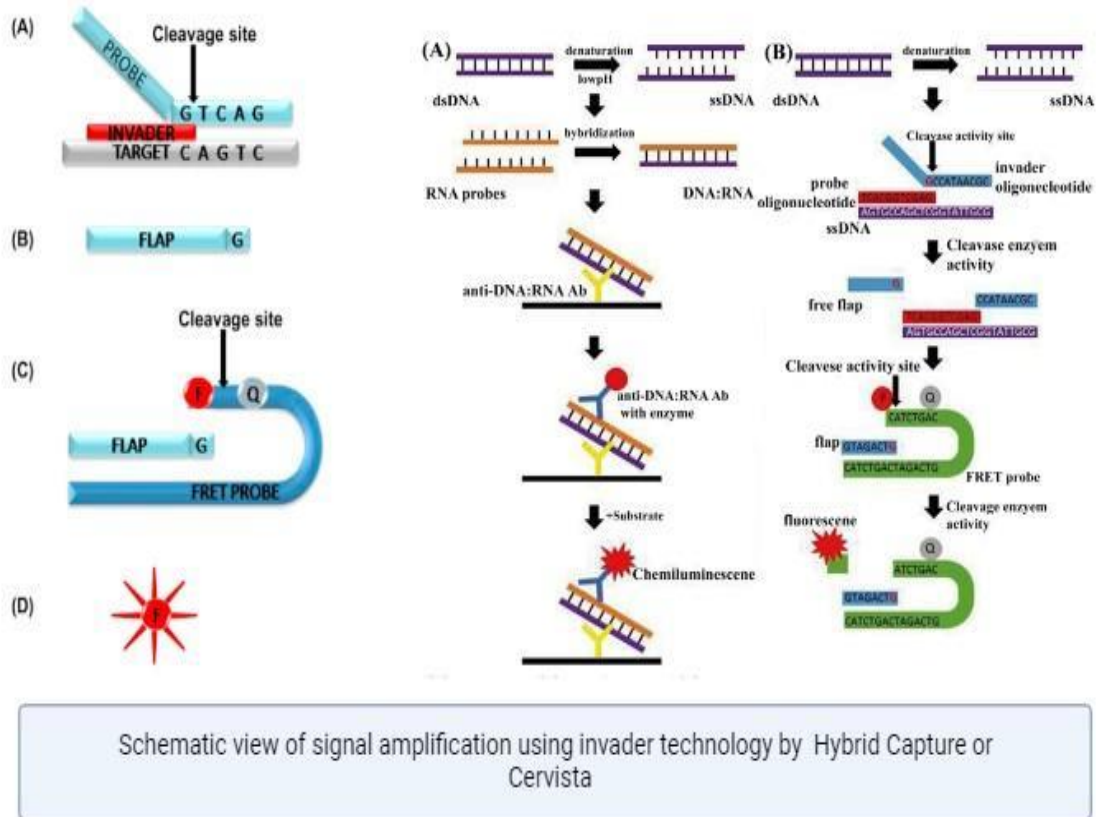
hybrids is greatly amplified, by a minimum of 3000 times, using alkaline phosphatase. The second technique, known as CervistaR, utilizes oligonucleotide binding and fluorescence production to identify 14 high-risk HPV types. Both approaches, including CervistaR, demonstrated great specificity and sensitivity in detecting the HR HPV 16 and HPV 18 genotypes (**Figure 5**) (Viana et al. 2020; Sitarz & Szostek, 2019).

Figure 5: Schematic image of Signal Amplification using Invader technology by Hybrid Capture and Cervista (Arney et al 2010).

1.9.2 HPV Genotyping and Nucleic acid amplification

Various amplification techniques can be utilized for the detection of HPV nucleic acid, including Microarray, Papillocheck, PCR-RFLP, PCR, COBAS, and linear array. PCR-

based amplification is a widely utilized and convenient approach for routine detection using gene primers.

The PCR-RFLP genotyping based approach utilizes consensus primers, including SPF1/SPF2, GP5+/GP6+, and MYB09/MYB11, to amplify a conserved section of the HPV genome, specifically the L1 capsid protein. These primers are designed to target a broad range of genital HPV strains. Subsequently, the amplified products underwent direct sequencing or line probe assay to identify HPV genotypes. An inherent limitation of this approach is its potential inability to identify all HPV genotypes present in the samples (Gravitt et al., 1998; Abreu et al., 2012).

1.9.3 Current Public Health Initiatives in Cervical cancer

1.9.3.1 HPV Vaccination

WHO and CDC recommend HPV vaccination to prevent cervical cancer. The age group of children and adolescents from 10 to 16 years it is recommended before the sexual activity (Meites, E. 2019 and Curtis et al. 2014). Catch up vaccination who didn't receive the vaccine should be taken before 26 years of age and after 27 – 45 years also can take the vaccine but the benefit will be less when compared to other groups. Administering HPV vaccinations is a significant preventive measure against cervical cancer (CC). The primary idea is that the vaccines consist of non-infectious virus-like particles (VLPs), which mimic the immunogenicity of the actual virus and can be administered with or without adjuvants. Then, stimulate the production of neutralizing antibodies specific to the relevant HPV genotypes. Prophylactic HPV vaccinations have been created to reduce the incidence of CC and related fatalities by preventing infection from high-risk HPV strains. These vaccines have the potential to prevent up to 75-80% of CC cases. The World Health Organization has approved three commercially available preventive HPV vaccinations Cervarix, Gardasil and Gardasil 9 vaccines or HPV vaccination (Roden & Stern, 2018).

Cervical cancer screening test, Pap test and HPV DNA test should be undergone every 2-3 years to all the women aged above 25 years (Jayasinghe et al. 2016 and Lees et al. 2016).

Raising awareness and education among students, schools, colleges and government institutions. Mobile health services should be incorporated in health communities for screening and treatment purpose. These efforts are aimed at reducing the incidence of cervical cancer globally, with an emphasis on prevention through vaccination and early detection through screening

1.9.4 Integration of HPV associated cervical cancer

HPV double-stranded non -enveloped DNA virus. The size of the genome is approximately 8kb and present in circular form (Marlow et al., 2014). The integration of the HPV genome in the host genome is the most critical step for the development of CC. The HPV genome typically contains 8 ORFs, divided into three functional segments: Early gene coding region (E), late gene coding region (L), and long control region (LCR), also known as non-coding region (NCR). The preliminary establishment and subsequent disease progression of this form of cancer is reliant on E6 and E7 oncogenes, which lead to uncontrolled proliferation of cervical mucosal epithelium cells (Rajasekaran et al., 2015).

Together E6 and E7 proteins are considered as biomarkers of CC and regulate all the hallmark genes responsible for disease progression. In the (**Figure 6A,6B**) as shown these proteins target growth suppressor genes of the cell, primarily tumor protein P53 (p53) through E6 and retinoblastoma protein pRb through E7. Both E6 and E7 oncoproteins engage in maintenance of viral episomes during the viral life cycle and promote cell proliferation by activating S phase controllers. E6 is a transcription factor that breaks down p53, which is necessary for the DNA restoration system to control cell cycle. E7 is linked to several molecular processes that promote cell growth, including histone deacetylases and the CDK inhibitors p21 and p27. The early proteins E6 and E7's chromosomal instability can induce telomerase hTERT's enzymatic activity, which inserts hexamers to the end of telomere.

Perseverance of HPV infection is associated with cellular transformation, disease progression, precancerous lesions, and invasive CC (Marlow et al., 2014). Epigenetics and histone modifications have a substantial impact on the development of CC through the

silencing of CpG islands, hypo- and hypermethylation of tumor suppressor genes, and the advancement of both HPV and cellular alterations. In addition, host genetic variations, aberrant long non-coding RNA, miRNA, and methylation silencing gene expression all contribute to mutations that develop, regulate cell cycle, and cause cervical cancer (Poondla et al 2021).

Interaction with Cellular Pathways

PI3K/AKT/mTOR: HPV can activate this pathway, often through mutations in the PIK3CA gene, promoting cell survival, proliferation and resistance to apoptosis (Bossler et al 2023).

Wnt signaling pathway: Disruption of genes FAT1can dysregulate the Wnt signaling pathway, which is involved in cell proliferation and further contributing to the cancerous phenotype (Peng, Z et al 2021).
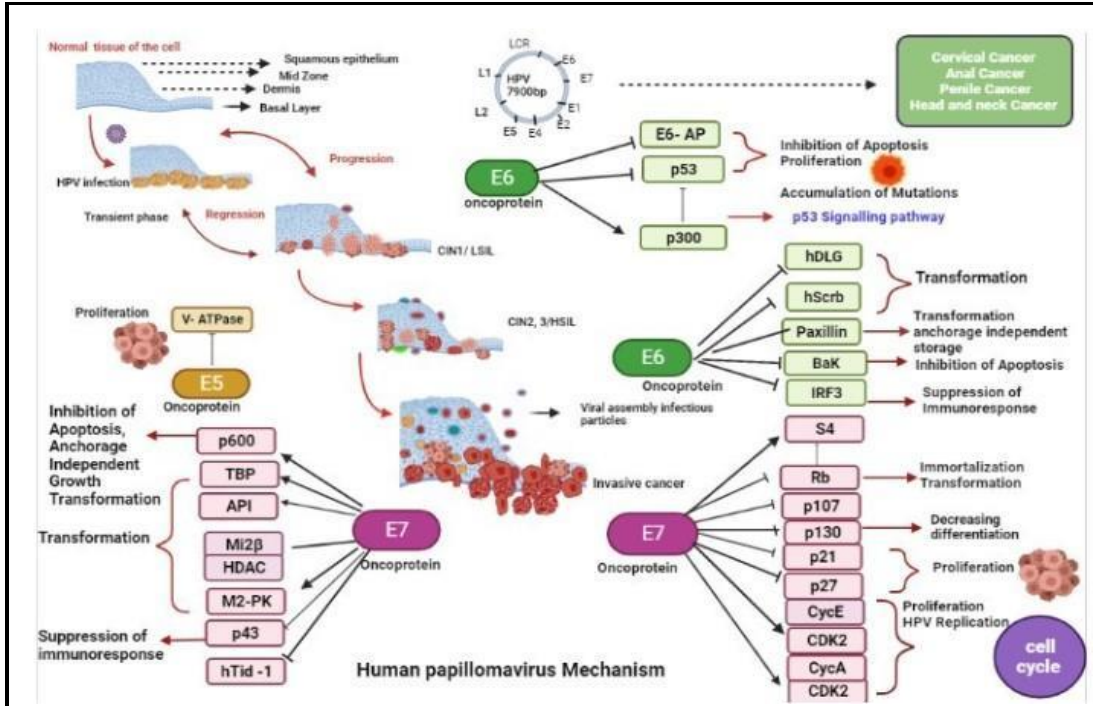
Figure 6A: Mechanism of HPV and Viral oncoproteins (The image was performed in Biorender Software).
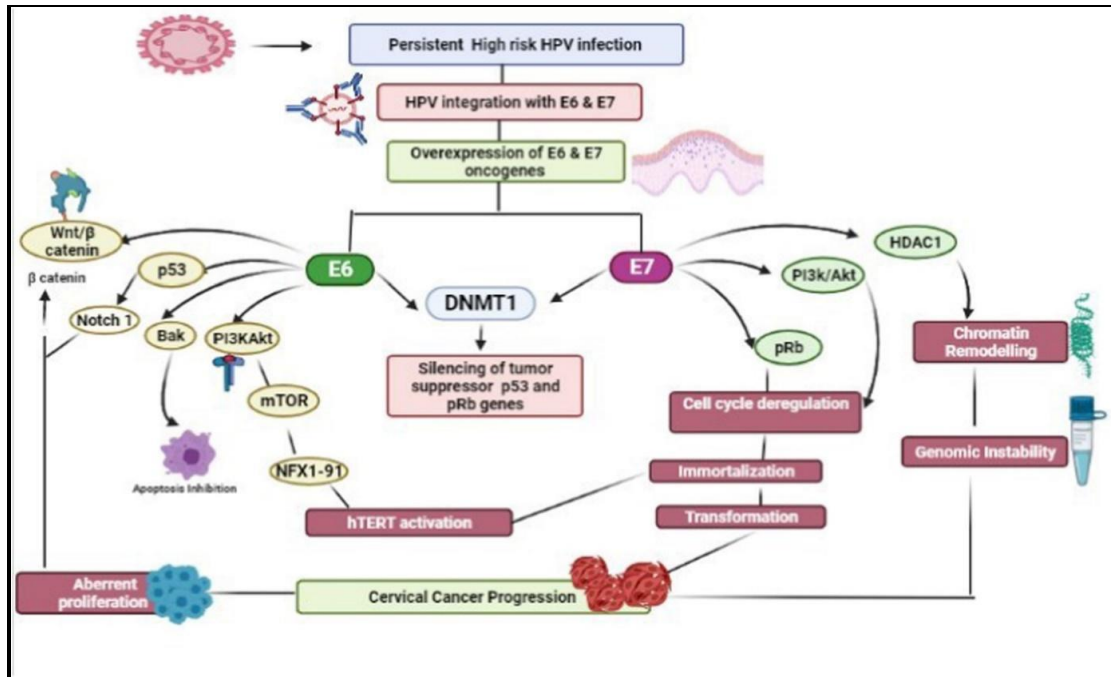
Figure 6B: Cervical cancer progression and integration of E6 and E7 oncoproteins (The image was prepared in Biorender Software).

Several epigenetic alterations, including DNA methylation, histone modification, and gene silencing by non-coding RNAs, have been detected in both the viral DNA and the genome of the infected cells. These modifications both start and maintain epigenetic changes, which are directly linked to a chronic high-risk HPV infection and cervical cancer.

According to FIGO staging, the tumor size and lymphatic metastasis were enlarged in cervical tissues due to long non-coding RNAs such as HOTAIR and lncRNA. In addition, the upregulation of VEGF, epithelial mesenchymal transition decreases the expression of catenin. Conversely, novel identified long non-coding RNAs like MET, TUCS8, and GAS5 show downregulated genes in cervical cancer (Jiang et al., 2015; Qiu et al., 2014). The expression of intracellular and intercellular miRNAs associated with viral oncogenes E6 and E7 in HPV-mediated transcription genes Decrease in intracellular concentrations of miR-27b, miR-143, miR-23 are linked to tumorigenic activities (Honegger et al., 2015).

1.10 Next-generation Sequencing (NGS) in CC

The integration of the HPV genome in the host genome may act as a biomarker for therapy, prognosis, and clinical analysis of CC patients. The configuration and physical state of the HPV genome could be integrated, episomal, or both. According to a prior study, patients with episomal HPV in cancer cells far better than those with an integrated HPV genome in terms of survival. The physical status of the HPV genome has been discovered using a variety of techniques, including), whole genome DNA sequencing (WGS), PCR-based approaches, RNA sequencing (RNA-seq), (HIVID), fluorescence in situ hybridization (FISH), and single-molecule real-time sequencing technology (SMRT-seq) (Yu et al., 2021). Each technique, though, has its advantages and disadvantages. To analyze HPV integrations in the human genome during whole-genome sequencing using NGS methods, a high coverage (>30 x coverage) of the HG- human genome and a complex algorithm are necessary. NGS short-reads (150–500 bp) can produce ambiguous and inaccurate maps of viral integration and difficult-to-assemble repetitive sequences.

1.10.1 Whole Exome Sequencing (WES)

NGS technologies have facilitated expedited sequencing endeavors to examine a diverse range of samples. From the entire set of genes in an organism to the specific genes that are transcribed into RNA to the subset of genes that code for proteins, these advancements have revolutionized our understanding of non-specific genetic variations in reproductive cells, mutations that occur in body cells, and changes in the structure of genes. In addition to establishing connections between a variant and human genetic disease, as described by (Zhang et al.2011). This can facilitate comprehension of intricate genetic illnesses, leading to improved diagnostic capabilities and the evaluation of disease susceptibility. However, the processing of exome sequencing data to identify variations still presents numerous hurdles (Meena et al.2018). Using NGS methods, such as (WES), it is possible to identify genetic variants or mutations of HPV that contribute to cancer or other infectious diseases. Exons are captured for enrichment using different library preparation procedures in the WES techniques. The CC patients with progressive and recurrent CC has a bad prognosis. Therefore, we can identify and analyze several genes that code for CC and repetitive CC,

as well as SNVs and mutations that are both responsible for CC and adenocarcinoma, using whole genome sequencing.

1.10.2 PCR Single Strand Conformation Polymorphism (SSCP)

Single strand conformation polymorphism is one of the techniques in molecular biology used to detect variations in DNA sequences, particularly single nucleotide polymorphisms (SNPs) or small insertions/deletions (indels). In the context of cancer research, SSCP can be employed to identify genetic mutations or variations associated with cancer susceptibility, progression and drug response.

The principle behind SSCP lies in the fact that single-stranded DNA molecules with different sequences will fold into unique conformations based on their nucleotide composition and sequence variations (Leung, K. H et al.2008). These conformational differences can be detected by electrophoresis, where the mobility of the DNA fragments through a gel matrix is influenced by their shape and size. The bands on the gel represent different conformations of DNA fragments. Each band may correspond to a different DNA sequence variant or mutation.

1.10.3 Transcriptomics Analysis using Affimetrix platform

Transcriptomic analysis is useful for identifying genes that are not functioning properly in cancer cells. This is done by comparing the gene expression profiles of tumor and non-tumor tissues. This technique provides crucial information about genes that are not working correctly in tumors, making it a valuable method for pinpointing biological markers of cancer progression and disease severity. In addition to gene expression and regulation research, transcriptomics is crucial for detecting gene alterations and mutations.

1.10.4 Systems Biology approaches in CC

The ability to predict CC using a collection of patient data on risk factors and clinical features is necessary for an early diagnosis of CC and subsequent clinical treatment. Although early stage CC can be identified with routine cervical Pap tests and cytological

screening, that needs significant amounts of resources and patient involvement. This could be the reason of higher mortality rates in many developing nations. Cervical screening programs are evidently only available to a small number of women, as low as 5% in some developing nations (Bray et al., 2018). Therapeutic procedures such as radiotherapy and surgery are effective measures, but they can also have a number of negative impacts on individuals who are not diagnosed at an early stage because these patients typically acquire more severe types in CC (Sowjanya 2014; Mandelblatt et al., 1991; Olusola et al., 2019; Workowski & Bolan, 2015). The result is that many cases of CC are identified late, when therapy is more difficult and survival is poorer. There are various causes for this, including lack of awareness and the cost of screening and diagnosis. However, as machine learning and system biological approaches have been demonstrated to have strong predictive value for a variety of diseases, this problem may possibly be improved utilizing ML analysis of patient clinical data that is easier to get. System biology approaches from KEGG databases, protein-protein interactions, pathway interactome and biological interactome network.

However, as *in silico* bioinformatic tools have been demonstrated to have strong predictive value for a variety of diseases, this problem may be improved by utilizing computational biology analysis of patient clinical data that is easier to get.

Summary:

Based on the above information and introduction this study has been designed to perform Next generation sequencing technology of Whole Exome Sequencing studies on cervical cancer samples on Illumina platform by Integrative system biology approaches and *in silico* bioinformatic tools to make the study more constructive in predicting the results. Combining **HPV genotyping**, **Single-Strand Conformation Polymorphism (SSCP)**, **Whole Exome Sequencing (WES)**, **transcriptomics** and **bioinformatics** in the study of cervical cancer can lead to a deeper understanding of the disease's molecular mechanisms and potentially guide personalized medicine. WES can reveal novel or previously unknown genetic mutations that could become targets for new treatments, like immunotherapy or

small-molecule inhibitors. So, this comprehensive, multi-omics approach provides a holistic understanding of the molecular underpinnings of cervical cancer. It helps in discovering new biomarkers, identifying therapeutic targets, and tailoring personalized treatments, which can lead to improved outcomes and more effective prevention strategies.

Chapter 2
Review of Literature

2.1 Cervical cancer

The cells of the cervix undergo transformation, display abnormalities, and develop precancerous changes or moderate dysplasia (i.e. presence of deviant cells within a tissue). Generally, precancerous lesions are not detected in time, which later go through alterations, gradually differentiate and display (Cervical Intraepithelial Neoplasia) CIN1, then advance to CIN2 and CIN3, and finally develop invasive cancer. CC first appears in the cervix, which is a long and narrow opening of vagina that leads to the lower end of the uterus. The area where the exocervix (the part of the uterus that extends into the vagina) is meet the endocervix (the entrance of the cervix leading into the uterus) is termed as the "transformation zone" (T-zone), which appear to be most favorable spot for the development of precancerous cellular alterations (Herfs et al., 2012; Höckel et al., 2014) (**Figure 7).**
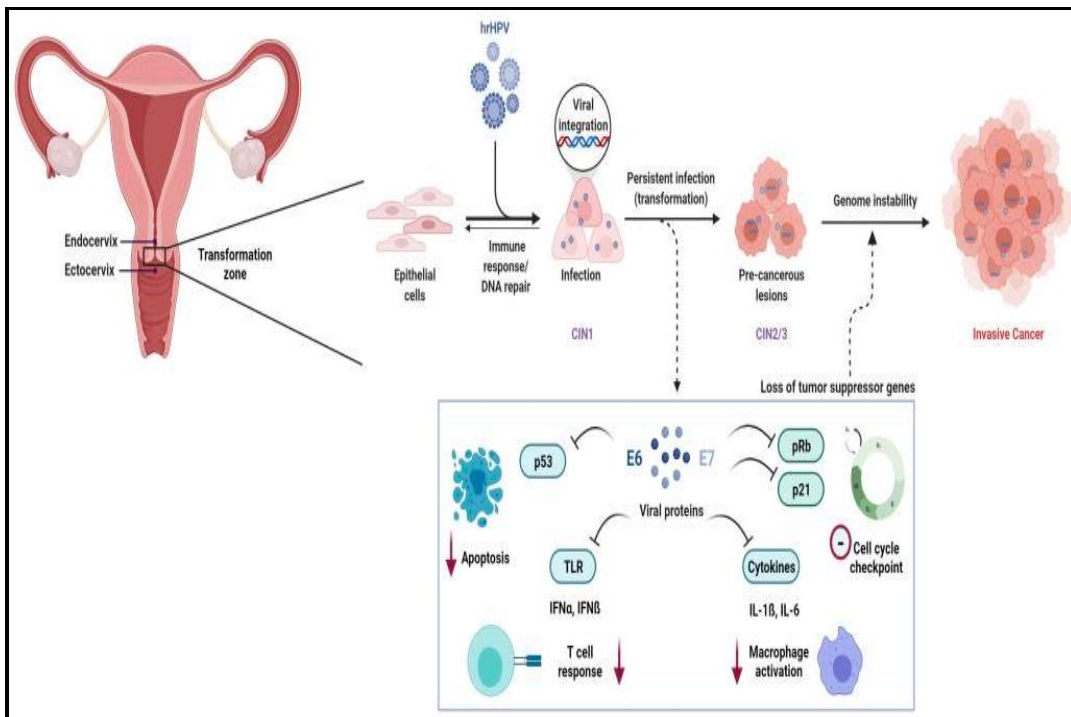
Figure 7: Schematic representation of HPV pathogenesis in Cervical cancer (The image performed in Biorender Software).

Released viral proteins E6 and E7 prevent TP53-mediated apoptosis, p21-mediated cell cycle checkpoint, toll-like receptor (TLR)-mediated T-cell response, and cytokine-mediated macrophage activation. (Ramachandran & Dörk, 2021)

The HPV infected cells slowly transformed into precancerous cells such as CIN, dysplasia, and squamous intraepithelial lesion (SIL), which can be typically diagnosed by a Papanicolaou (Pap) smear test (Herfs et al., 2012).

Even after the development of precancerous cells, reversions are frequently feasible; otherwise, invasive cancer may develop in some circumstances. Recurrent cancer refers to the return of CC following therapy; it can be limited (recurrence at the same site), regional (recurrence in nearby tissues), or distal (recurrence in distant organs) (Peiretti et al., 2012). Cancers of the mouth, throat, larynx, anus, vulva, vagina, lung, bladder, colorectal, or pancreatic are more common in CC survivors (Burd, 2003).

However, most HPV infections (about 87%) are temporary, remain undetectable and do not cause any severe health issues. Although, about 12.9% of HPV infections are persistent and can further lead to development of low-grade lesions to high grade cancerous lesions. The occurrence of CIN depends upon various factors, such as age, immunity and lifestyle of a particular person as well as on the type of HPV (Bruno et al., 2021).

Premalignant transformations often appear as abnormal growth or lesions. CIN is the term used to characterize the premalignant lesion of squamous cell carcinoma (SCC). The kind, intensity, and prevalence of high-risk HPV infections are key determinants of CIN infection (Bruno et al., 2021). CIN is categorized into three grades, CIN1, CIN2, and CIN3, which correspond to 1/3, 2/3, and severity of the alterations, specifically the percentage of the epithelial layer with neoplastic changes.

As a result, atypical cells are present in almost all of the epithelium's layers (Solomon et al., 2002). The Bethesda and Pap categorization systems are used often in cytology (Table 2). The European Pap classification is classified into 5 grades, PAP1 to PAP5, based on the severity of anomalous smears (Papanicolaou & Traut, 1997). According to the American Bethesda classification, HSIL, which are more likely to proceed to CC, and LSIL, are less likely to progress (Solomon et al., 2002; Poondla N. et.al.2021).

Histology		Cytology	
Dysplasia	CIN	Bethesda	Papanicolaou
Normal	Normal	Within normal limit	Pap1
Benign atypia	Inflammatory atypia	Benign cellular changes	
Atypical	Squamous atypia	ASCUS	Pap2
Mild Dysplasia	CIN1	Low grade SIL	Pap 3A1
Moderate Dysplasia	CIN2	High Grade	Pap 3A2
Severe Dysplasia	CIN3		Pap 3B
Carcinoma in situ			Pap 4
Micro invasive cancer	Micro invasive cancer	Micro invasive cancer	Pap 5

Table 2: Overview of several nomenclatures and classification for histologic and cytologic cervical abnormalities.

2.2 Incidence and mortality of cervical cancer (CC)

According to The Global Cancer Observatory (GLOBOCAN) 2020 database, CC is the fourth most common cancer in the world, which accounts for roughly 604,127 cases and 341,831 deaths in women (Sung et al., 2021). Globally, cervical cancer is projected to have an age-standardized incidence rate of 13.3 per 100,000 women (**Figure 8**). However, the average age-standardized mortality rate is 7.3 per 100,000 women. The highest prevalence of cervical cancer and subsequent mortality rate are reported in low- and middle-income

nations (LMICs) such as eastern Africa, middle Africa and southern Africa. In contrast, in Australia, New Zealand, and the United States, it has an incidence that is 7–10 times lower. However, low- and middle-income nations had higher incidence and mortality rates (18.8/100000 vs. 11.3/100000 incidence and 12.4/100000 vs. 5.2/100000 for mortality) than developed countries (Drolet et al., 2021; Sung et al., 2021), as shown in the **Figure 8 and 9.**

About 87% of CC-related mortality takes place in underdeveloped and developing nations due to the absence of efficient screening programs intended to find and treat precancerous abnormalities. The conditions, coupled with other sociocultural norms including early marriage, high parity, and to some level polygamy, were found to exacerbate the susceptibility of the community studied (Lorenzi et al., 2015).

The American Cancer Society predicts that 13,960 new cases of CC will be identified in the United States in 2023-24. There will be 4,320 deaths from CC in women. [http://www.cancer.org]. Many elderly women are unaware that they continue to have a risk of acquiring cervical cancer as they become older. Over 65-year-old elderly women make up more than 20% of incidences of cervical cancer. A recent study by the National Cancer Institute reported that the average age range for cervical cancer diagnosis is between 35 and 44. The age-standardized rates (ASRs) per million people in the greatest risk are in the regions of eastern Africa (40.1), western Africa (29.6), southern Africa (43.1), Central Africa (26.8), and south Pacific Ocean region Melanesia (27.7) (http://gco.iarc.fr). A recent Lancet study estimated the total number of cervical cancer cases in Asia, in which India accounts for 21% of CC cases followed by China with 18% of CC cases.

In India, a total of 60,078 of the 96,922 women who receive a CC diagnosis per year pass away from the illness (Bruni et al., 2023). The risk of having CC is considerable among the 429 million Indian women between the ages of 15 and 44, with the peak risk occurring between the ages of 40 and 64 (Bruni et al., 2023). The histological kind of cancer that affects the ectocervix most frequently is squamous cell carcinoma, whereas the endocervix

typically has adenocarcinoma. More than 70% of Indians live in rural areas, where women with poor socio-economic level are most likely to get CC, which tends to be the most common malignancy among women (Nandakumar et al., 2009; Sarathi et al., 2015). It was observed that compared to 41.2% of patients who were aged, 79% of younger individuals had early diagnoses, such as stage IA-IIA. Because elderly women typically pass away within three years after being discovered, the situation may be more serious in this case (Gao et al., 2013). Although the overall death rate has been greatly cut by preventative and therapeutic strategies, elderly women's incidence has not significantly decreased. The consistent screening technique could not be used on patients who were elderly, which constrained its ability to treat patients (Jemal et al., 2011).
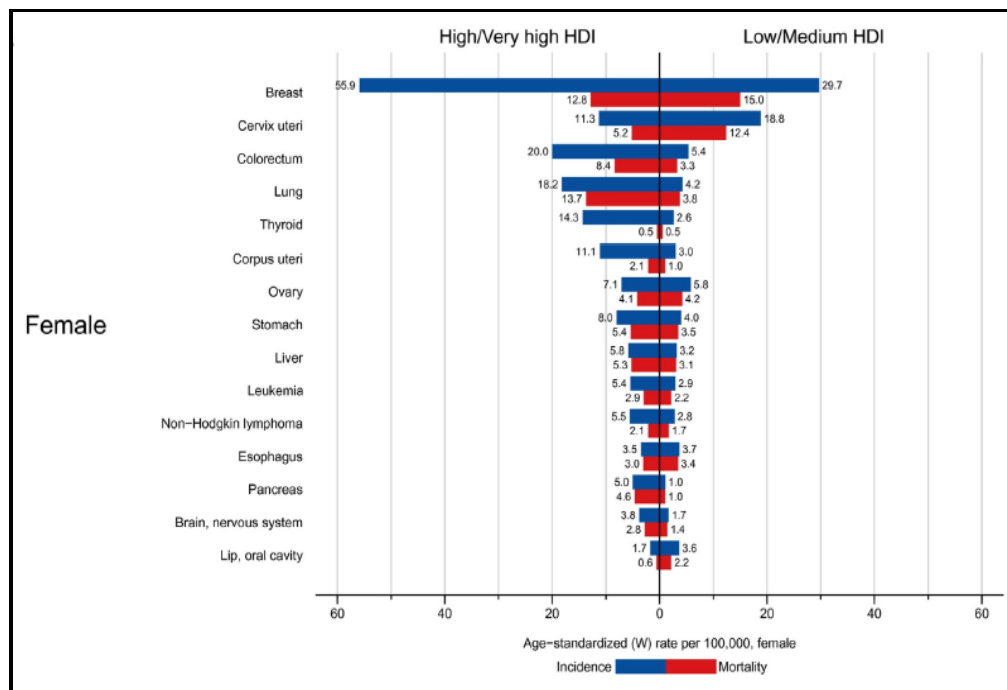
Figure 8: Age Standardized rate with very high-density incidence and mortality (https://gco.iarc.fr/en)
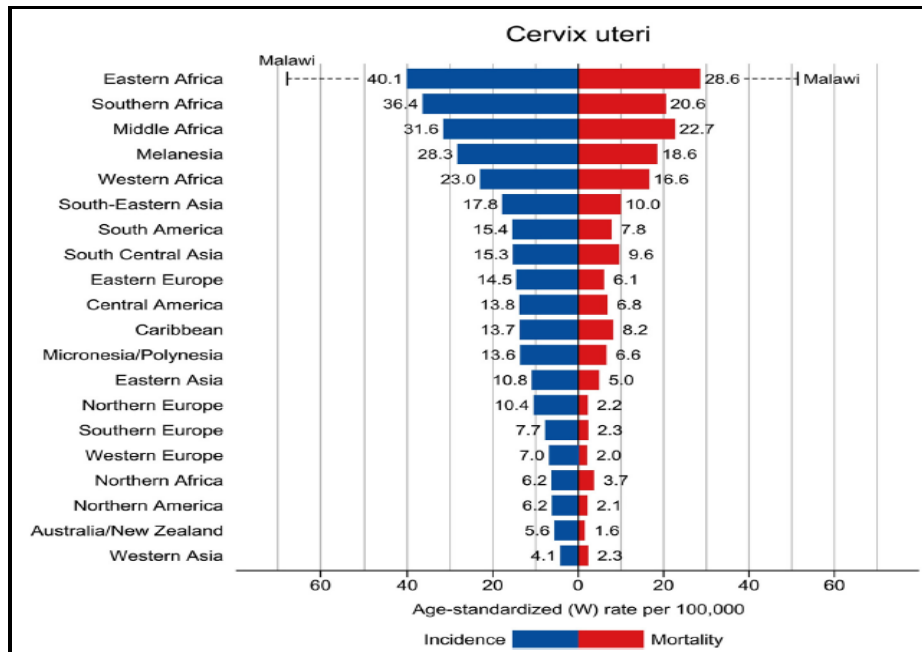
Figure 9: Age Standardized rate (ASR) per 100,000 incidence and mortality rates around all the countries by GLOBOCAN (https://gco.iarc.fr/en)
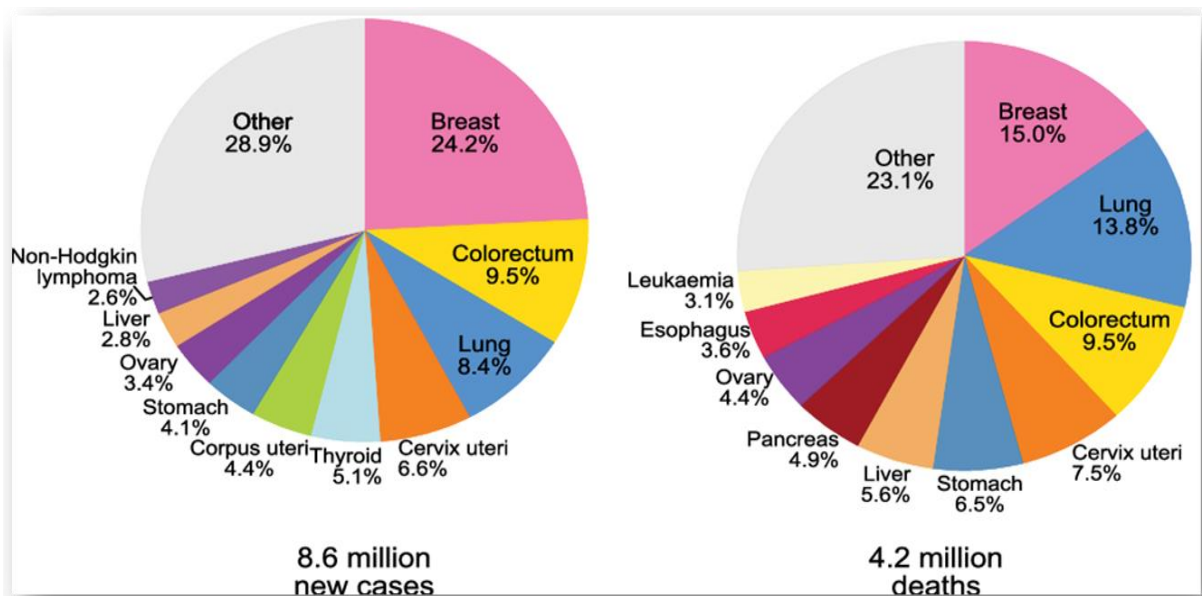
Figure 10: According to WHO mortality and Incidence rates of all cancers around the world (https://gco.iarc.fr/en).

According to Global Cancer Observatory, females aged from 20 to 85 estimated incidence rates from 2020 to 2040 are 124 K to 191 K and deaths would be from 77.3K to 125K on an average in India. The Asian countries estimated mortality rates would be 199K to 391K, from 2020 to 2040 (**Figure 11**).

Figure 11: Estimated incidence and mortality rates for 2040 according to GLOBOCAN

(https://gco.iarc.fr/en)
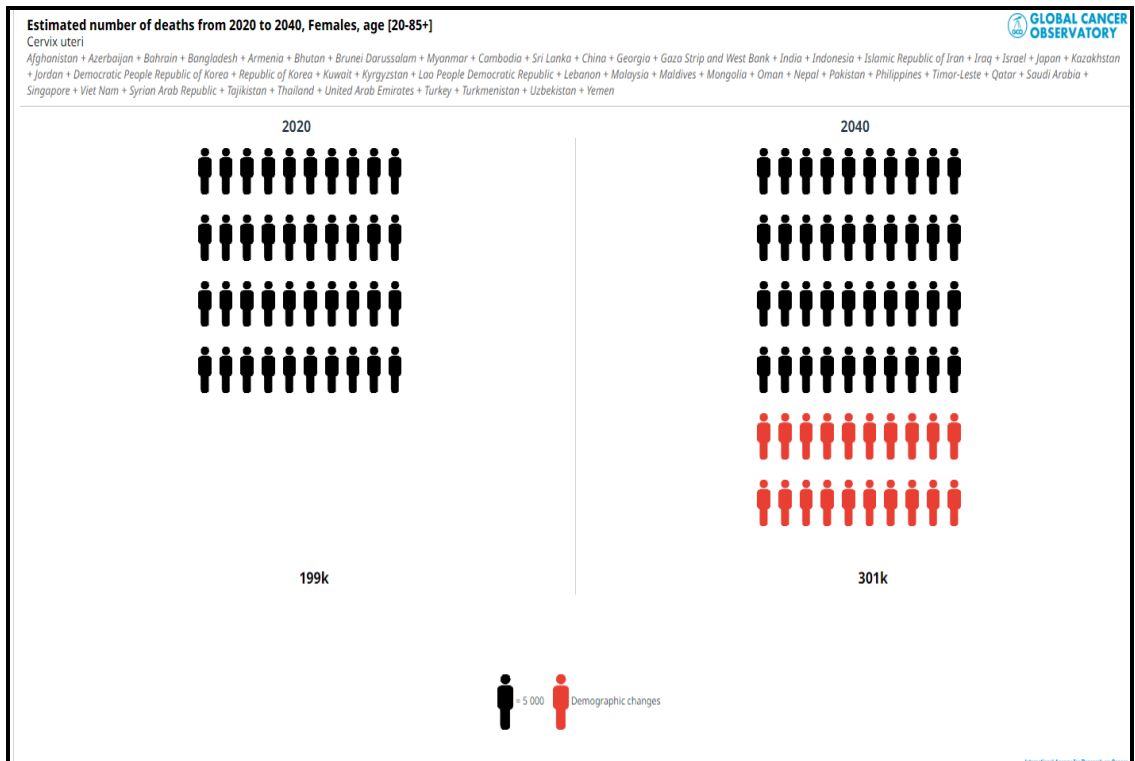
Figure 12: Estimated mortality rates of Asian countries from all over the world according to GLOBOCAN 2040 (https://gco.iarc.fr/en).

The number of incidence and mortality cases in European countries is very less or nil when compared to other countries like Africa, India, US, Latin America and Carribean (**Figure 12**). The European countries' incidence is from 57.3 K decreased to 55.1K and mortality rates might affect from 26.0K to 27.2 K it remains almost the same (**Figure 13**).
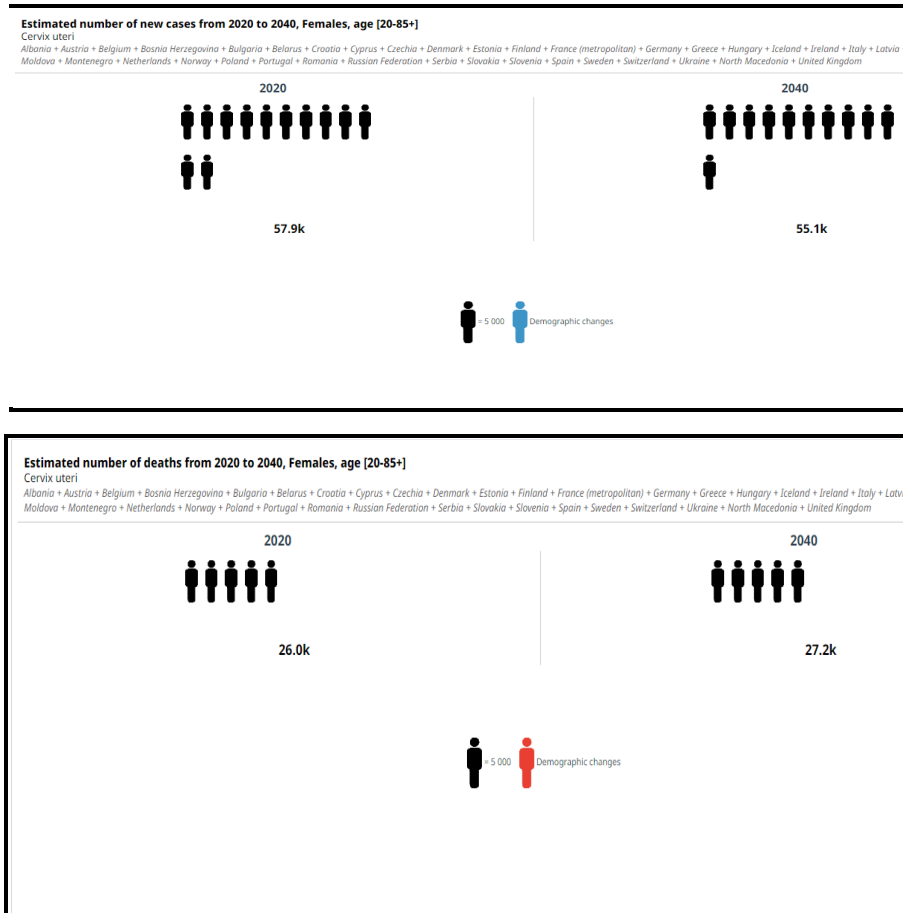
Figure 13: Estimated incidence and Mortality rates of European Countries of GLOBOCAN 2040 (No new Cases of Incidence and Mortality)(https://gco.iarc.fr/en).

The overall world Incidence and mortality cases in the world will be high from 2020 to 2040 from 603K to 847K and mortality rates from 342K to 524K (**Figure 14**). The estimated rates will be increase around the world.
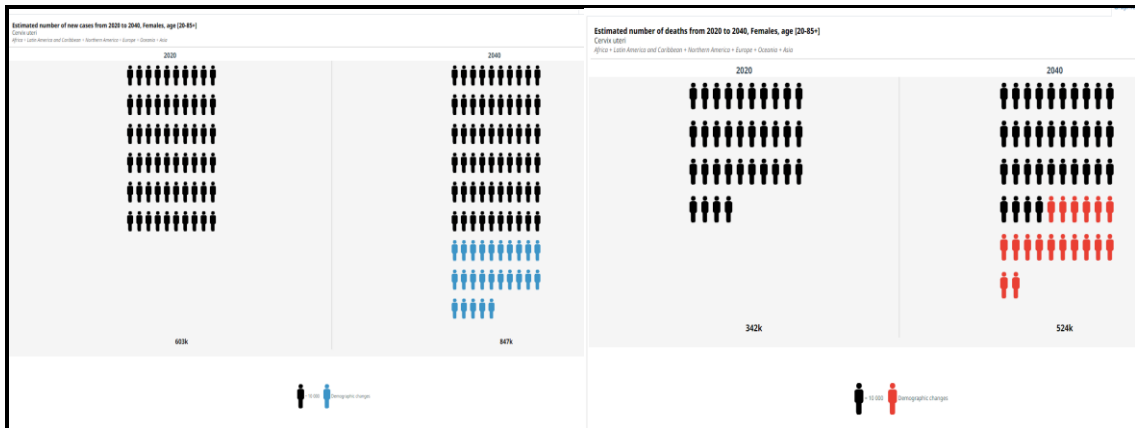
Figure 14: GLOBOCAN the estimated number of incidence and mortality cases of Africa, Latin America, Europe, Asia, North America and all over world (https://gco.iarc.fr/en).

2.2.1 Incidence and Mortality CC cases in India

In India, overall CC rate is ~6-29% when compared to other remaining cancers. AAI (Age adjusted incidence) rate increases to the highest with 23.08 divided by 1100,000 in Mizoram & very lowest 4.91/100,000 in Dibrugarh (Dutta et al.2012). The survival degree rate of cervical cancer in Chennai is about 60% and in Bhopal is 35% (**Table 3**). In developed countries, public awareness & health programs are bringing awareness among the public but in India and developing countries there is a lack of perception about cervical cancer. As in India, it is stated that part of the world is burdened with cervical cancers. Cancer mortality accounts for 17% -18% between the age of 29-60 years. Among 1 in 53, women in India are suffering with cervical cancer when measured with 1 out of 100 women in high-income countries.

State in India	Year	Relative proportion	Incidence Rate (AAR)	HPV Prevalence
Ahmedabad	2012-1013	9.28	6.91	16.2
Aurangabad	2012-2014	19.86	14.30	32.9
Bangalore	2012	12.29	15.33	34.6
Barshi rural	2012-2014	26.74	16.09	36.8
Barshi Expanded	2012	28.56	14.65	33.0
Bhopal	2012-2013	12.54	13.83	29.7
Cachar District	2012-2014	13.86	12.65	34.9
Chennai	2012-2013	12.60	15.88	35.9
Delhi	2012	10.83	15.53	36.9
Dibrugarh district	2012-2014	6.39	4.91	12.8
Kamran urban district	2012-2014	8.57	14.52	34.0
Kolkata	2012	10.09	10.43	23.8
Kollam	2012-2014	6.81	6.69	14.7
Manipur	2012-2014	9.71	6.14	13.9

Meghalaya	2012-2014	11.08	9.55	23.9
Mizoram	2012-2014	25.94	23.07	62.0
Mumbai	2012	7.66	9.03	18.7
Nagaland	2012-2014	16.67	13.24	34.2
Nagpur	2012-2013	13.53	12.88	30.2
Pasighat	2012-2014	23.90	22.54	66.3
Patiala	2012-2014	10.48	11.46	29.0
Pune	2012-2014	10.55	8.95	20.0
Sikkim	2012-2014	10.03	10.05	24.5
Thiruvananthapuram	2012-2014	6.04	7.00	14.6
Tripura	2012-2013	16.84	9.15	23.6

Table 3: HPV occurrence in women with population screening studies in India *HPV prevalence for cervix uteri based on average age adjusted (GLOBOCON and WHO).

2.3 Biomarkers of Cervical Cancer

The ongoing research on biomarkers contributes to better management in terms of screening, early clinical diagnosis, and precise prognosis of CC. The biomarkers of CC are classified under two categories: viral biomarkers (such as hrHPV DNA testing, genotyping, HPV E6/E7 mRNA viral load, transcriptional status, viral DNA methylation, WES of

specific viral proteins), and cellular biomarker (such as CDKN2A/p16ink4a, proliferation indicator MKI67/Ki67, Topoisomerase II alpha (TOP2A), Mini chromosome maintenance complex component 2 (MCM2) (Molina et al., 2020).

It is confirmed that the viral biomarker hrHPV infection is the primary etiological reason for the development of CC. As a result, a lot of research is being done on biomarkers that are directly related to the HPV life cycle and the natural course of HPV-dependent cervical malignancy. They can function as cellular biomarkers for CC and represent aberrant circumstances or cellular transforming processes linked to HPV. Thus, the overall picture of the relevant altered biomolecular pathways would be taken into justification to discover trustworthy functional biomarkers (Nam et al., 2008; Romagosa et al., 2011; Woodman et al., 2007). **p16INK4a** is a biomarker which on overexpression of this protein correlates with HPV-mediated disruption of the pRb pathway and is used as a surrogate marker of high-risk HPV activity (von Knebel Doeberitz, et al 2012).

In addition to persistent hrHPV infection, further epigenetic modifications are needed for transformation from precancerous to cervical carcinoma. DNA methylation is one of the common and early triggers of molecular changes to develop invasive cervical malignancy. The methylation markers include T Cell Differentiation Protein (MAL), Cell adhesion molecule 1 (CADM1), Ras Association Domain Family Member 1 (RASSF1), Telomerase reverse transcriptase (TERT), Death-associated protein kinase 1 (DAPK), Cyclin A1 (CCNA), E-cadherin (CDH1), Twist Family BHLH Transcription Factor 1 (TWIST), O-6-methylguanine-DNA methyltransferase (MGMT), Fragile Histidine Triad (FHIT) Di adenosine Triphosphatase (DAT), and retinoic acid receptor beta (RARB) (Pardini et al., 2018). Recently, it was discovered that the expression of several microRNAs (miR-21, miR-127, and miR-199a) is upregulated in cervical intraepithelial neoplasia while the expression of some microRNAs (miR143, miR214, miR-218, and miR-34a) is decreased in cervical cancer (Das Poulami et al. 2016). Recently, proteomics is the most developed and promising technique which uses identify differentially expressed proteins for the

biomarkers to predict different stages of CC (Kausar Neyaz et al., 2008; Neyaz & Ahmad, 2019). (**Figure 15)** (**Table 4**)
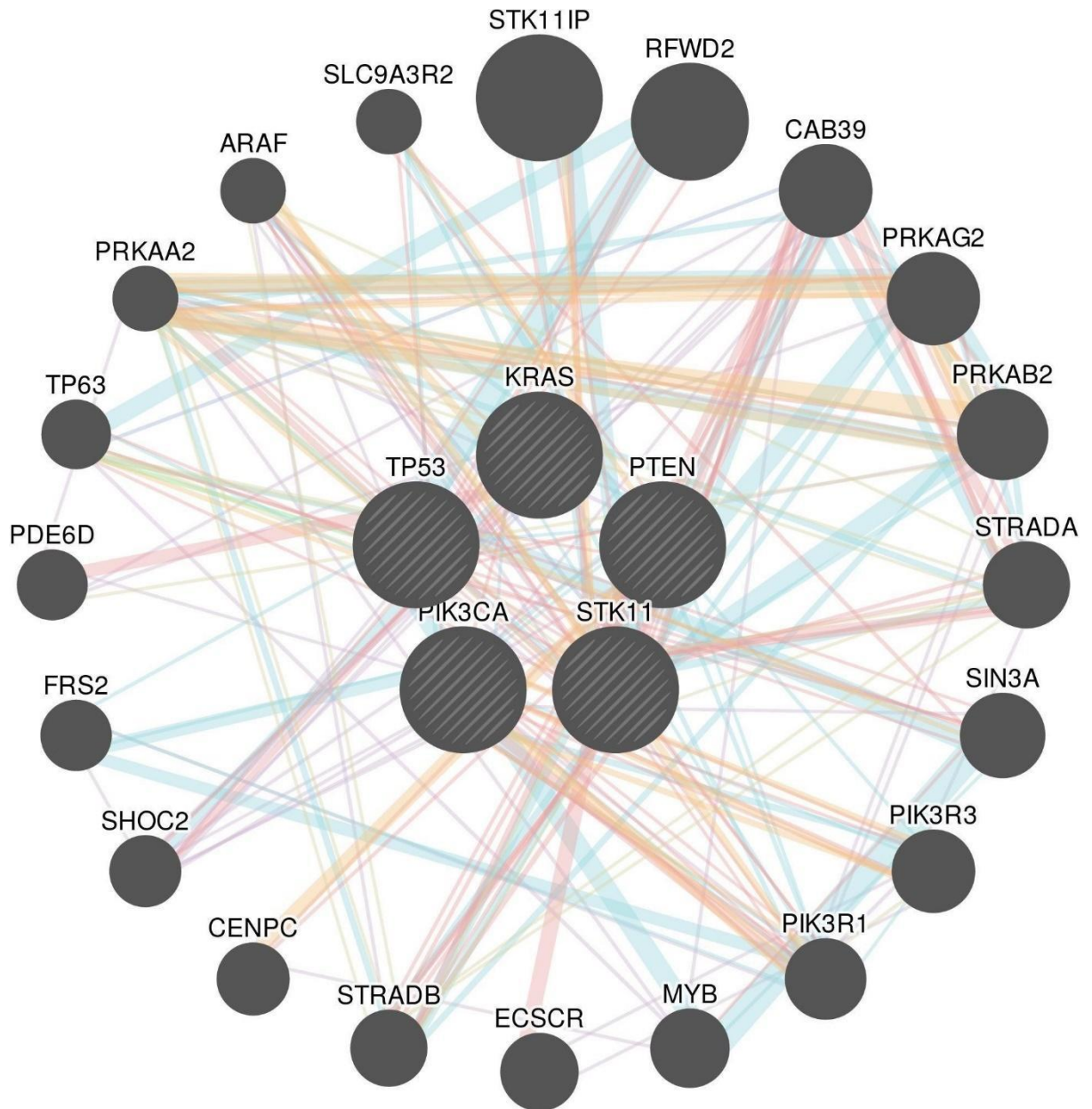
Figure 15: Systems genomic association of PIK3CA, PTEN, TP53, STK11, KRAS with a gamut of pathways.

The image was made using Genemania wherein the dark pin edges between the nodes represent the physical interactions while the other pink edges indicate the mere associations, and the blue edges indicate they are associated with the common pathways.

S. No.	Genetic and epigenetic biomarkers	Sensitivity	Specificity	ROC or AUC analysis	References
1	RAB3C GABRA2 ZNF257 SLC5A8	96.2%	95.2%	94.2% 100% 100% 100%	(Xu et al., 2019)
2	Septin 9	50-80%	81.48	0.80	(Bu et al., 2020)
3	SOX 14	94.12	86.46	0.99	(J. Zhao et al., 2021)
4	SCC-AG	78.6-81.2	74-100	0.79-0.89	(Onyango et al., 2020)
5	M-CSF	68-87.7	64.7-94	0.79-0.86	
6	VEGF	56-83.5	74.6-96	0.83-0.88	
7	MICRORNA 9	52.9-67.3	76.4-94.4	0.71-0.85	
8	P16INKa/KI-67	50-100	39-90.4	0.76-0.90	
9	HPV E6/E7 MRNA	65-100	42.7-90	0.59-0.80	
10	Mir 92 a 5p	95	87	0.93-0.94	(Azimi et al., 2021)
11	Mir 155	89-95	83	0.91-0.98	
12	ASTN1 DLX1 ITGA4 RXFP3 SOX17	64.8	94.6		(El-Zein et al., 2020)

	ZNF671				
13	POU4F3	74	89		
14	FAM19A4	64.9	76.4		

Table 4: Biomarkers for Cervical cancer; comprehensive table on the genetic and epigenetic markers in the context of cervical cancer with clinical information on sensitivity and specificity analysis or ROC analysis.

2.4 Screening and treatment of precancerous lesions for cervical cancer prevention (WHO)

Cervical intraepithelial neoplasia (CIN) is a precancerous condition in which atypical cellular changes are observed on the epithelial cells of the "T zone" of the cervix. CIN can be seen in one of three stages: CIN1, CIN2, or CIN3. Cervical cancer can develop from CIN2 or CIN3 (together referred to as CIN2+) if remains untreated for a significant period. The screening and prognosis of CC can be confirmed by following a standard procedure which includes cytology, colposcopy, biopsy, and histology of cervical dysplasia (Perkins, Rebecca B., et al. 2023).

Alternatively, WHO guidelines provide recommendations to use a "screen-and-treat" approach, which suggests starting screening tests for HPV detection at the age of 30 years for the general population of women with regular screening every 5 to 10 years. The treatment decision is made, ideally, as soon as possible if it comes up with a positive HPV screening test. The standard relevant screening procedures involve (HPV) testing, (VIA), and cytology (Pap test) (Onwuka et al 2017). However, the World Health Organization (WHO) advises utilizing HPV DNA detection as the primary screening test among both the overall population of women and those infected with HPV.

2.4.1 High-risk HPV DNA tests

These tests target a collection of HPV genotypes with a high risk of cancer (http://www.cancer.net/cancer-types/cervical-cancer/diagnosis). The most prevalent and high-risk genotypes of HPV are 16 and 18, respectively. When HPV tests detect HPV16 and 18 apart from other carcinogenic kinds (which may include HPV45 in rare situations),

we refer to those tests as having incomplete genotyping. This recommendation explicitly mentions partial genotyping to recognize women who pose the greatest risk of developing CC among HPV positive communities. HPV DNA testing can be done using provider sampling or self-sampling. An HPV DNA test in this recommendation refers to a HR HPV DNA test, also known as a nucleic acid amplification test (NAAT) (Kundrod et al. 2020)

VIA testing makes use of diluted vinegar (acetic acid) on the cervix without the use of a magnifying glass to find aceto-white lesions that may require either ablation or excision; or additional examination. VIA is not recommended for women after menopause or in case of non-distinctive transformation zone.

Cytology

It comprises the PAP smear test and liquid-based cytology (LBC) that use slides that are prepared and interpreted under a microscope by a qualified professional to find abnormal cells on the cervix.

As per WHO recommendations, early screening with an HPV test or VIA show promising results in reducing CC cases and related fatalities, in comparison to no screening condition. The "screen and treat" strategy screening test may also minimize CIN2+ reappearances (3 in 1000 fewer), CCs (1 in 10,000 fewer), and related fatalities (6 in 1,000,000 fewer), in comparison to the VIA approach and cytology followed by colposcopy. Even though malignancies were found during the initial screening with VIA, additional resources were needed for cytology, followed by colposcopy (Porras, Carolina, et al. 2012). In comparison to HPV screening test, recent studies found that there were more CIN2+ recurrences, VIA (4/1000 more), CCs (1/10000 more), and fatalities (7/100000 more). When screening for the first time, HPV discovered more instances than VIA (7/10000 more), and since this approach has a lower risk of false positives, relatively few women (1/10 fewer) may undergo unnecessary testing (Harlfinger et al., 2023).

There are a few important treatment modalities that mainly aim to eliminate the cervix's transformation zone as well as any cervical regions that have been detected as abnormal

during screening. The type of treatment options includes cryotherapy, (LEEP/LLETZ) (CKC) (Castle et al 2017). These treatment techniques can be categorized in two categories: ablative (removing aberrant tissue surgically using LLETZ or CKC), or excisional (freezing using cryotherapy or heating to cause thermal coagulation).

2.5 Treatment modalities

Treatment modalities for CC differ depending on the patient's age, health status, predicted side effects, and cancer kind and stage. Therefore, a systematic strategy is required to deliver patient-specific comprehensive care for cancer patients. Pregnant women's treatment is also handled delicately to have as little impact on the unborn child as possible, or it may even be postponed until term. The following therapies are available and differ from patient to patient. On the National Comprehensive Cancer Network (NCCN) and (NCI) websites, comprehensive treatment choices, personalized suggestions, and assistance are available [www.nccn.org] **(Table 5).**

S. No.	Type of treatments	Procedure	References
1	Surgery	Killing of cancer cells by keeping a cooled metal probe on vagina. The use of a focused laser beam to destroy cancer cells. Involves removal of uterus and cervix. Involves removal of cervix leaving the uterus intact with pelvic lymph nodes.	Zhang et al. 2017
2	Chemotherapy	Involves the restriction of further spreading of cancer cells by administration of oral drugs	Tewari et al 2014

3	Radiation therapy	Involves utilization of high energy rays or particles to destroy or shrink affected cells	Morris et al. 1999
4	Palliative care	Includes medication, nutritional changes, relaxation techniques for patient	Saslow et al. 2012
5	Follow-up care	Includes implementation of a physically active lifestyle, healthy eating habits and diet, strict vegetarian diet, restriction of alcohol	Elit, L., et al. 2010

Table 5. Treatment modalities of CC

2.5.1. Surgery

Cryosurgery is a technique used to destroy cancer by using a freeze metallic probe after cooling with liquid nitrogen (Santesso et al., 2012). Laser surgery utilizes a highly focused laser beam that is transmitted through vagina to eliminate malignant cells (Mendoza-Nava et al., 2012). According to recent studies, conservative surgery is recommended for young CC patients as it preserves the fertility of women patients. The procedure includes a radical trachelectomy and a cervical cone biopsy (Rema & Ahmed, 2016).

Radiation therapy: This technique implies high-energy X-rays to kill cancer cells. It can be of intracavitary brachytherapy (internal radiation therapy) where the source of radiation is implanted in the vagina, or external-beam radiation therapy ((EBRT) type (Eifel et al., 2014). Brachytherapy may cause vaginal and vulvar irritation, resulting in redness and discharge. Radiation therapy is administered over a set number of sessions. Radiation therapy frequently causes side effects such leukopenia, anemia, an abnormal menstrual cycle, and skin vagina.

2.5.2 Chemotherapy

Chemotherapy involves administration of oral or intravenous medications to cancer patients, works to eradicate cancer cells by preventing them from proliferating, and producing new cells. (Tewari & Monk, 2014). Chemotherapy typically entails a defined number of sessions over a predetermined time frame. As per patient's condition, doctors may prescribe one medication at a time or a variety of medications administered simultaneously. Chemotherapy and radiation therapy are frequently combined for cervical cancer patients, also known as concurrent chemoradiation.

It is recommended to use chemotherapy for the treatment of recurrence cancer. According to studies, people with advanced CC may benefit significantly from adding Avastin to their chemotherapy regimen. Chemotherapy medications that are frequently used include carboplatin, mitomycin, gemcitabine, docetaxel, paclitaxel, isocyanide, 5-fluorouracil, irinotecan, cisplatin, and topotecan (Tewari & Monk, 2014). Chemotherapy side effects can differ depending on the phase of cancer and the dose of the medication, which mainly includes exhaustion, infection risk, nausea, vomiting, loss of hair, appetite decrease, aphthous ulcers, and diarrhea (Siegel et al., 2013). Recent studies have indicated that peripheral neuropathy is experienced by 60% of patients with cervical cancer as a transient side effect. The peripheral nerve intrusion causes lumbosacral plexopathies and full nerve degeneration, possibly being the cause of the pelvic pain (Mitra et al., 2016). Peripheral neuropathy affects peripheral nerves that commonly results in dimness, numbness, and muscle spasms, largely in the hands and feet. Also, it may have an impact on many biological functions like metabolism, urine, and circulatory. The prediction of these medications using chemical interaction models or molecular modeling methods has received a lot of interest.

It is possible to determine whether a medicine is effective based on its molecular description. Chemotherapy interferes with the normal functioning of bone marrow cells, which results in a decreased count of the number of RBCs, WBCs, and blood platelet cells. This result in increased susceptibility to infection, more hemorrhage from superficial

wounds, and enhanced breathing rate. Occasionally, chemotherapy can cause myelodysplastic syndromes or acute myeloid leukemia, which are both conditions where bone marrow cells are permanently damaged. It is best to prevent getting pregnant during a woman's chemotherapy treatments because it could cause birth abnormalities. Chemotherapy side effects that may last a lifetime include early menopause, pregnancy complications, and osteoarthritis (Choi et al., 2013).

2.5.3 Palliative care

The patient is given palliative or supportive care, which may include drug therapy, dietary adjustments, and stress reduction, to treat any adverse effects and to meet their physical, emotional, and social requirements (Phippen et al., 2013). In advanced stages, the goal of hospital care is to deliver the best quality of life likely for cancer patients with an effort to progress their value of life through palliative care and specialized equipment.

2.5.4 Clinical trials

Clinical trials are carried out to assess the safety and efficacy of innovative treatment modalities. It requires patient consent under fundamental ethical and legal rights. It might involve a novel pharmacological strategy; a mix of conventional therapies, new dosages of therapeutic intervention, and it might even be the only way to find more effective treatment options for cervical cancer. The National Cancer Institute (NCI) clinical trial web page provides details on ongoing clinical trials. The procedure's specifics, how it differs from usual care, its hazards, and any additional requirements for the patient should all be presented to the subjects before they are enrolled [**www.cancer.gov].**

2.5.6. Immunization

The development of the first-ever HPV vaccination, known as Merck's Gardasil and authorized by the FDA in 2006, was a significant step toward the prevention of CC. The main agents that cause CC are HPV 16 and 18, and this vaccine is effective against infections caused by HPV6, 11, 16, and 18 (Villa et al., 2005). The vaccine is made from recombinant HPV type-specific capsid proteins that can trigger type-specific neutralizing antibodies and must be intramuscularly administered at 0, 2, and 6 months in both females

and boys aged 9 to 15 years (Stanley, 2007). Cervarix, another vaccine, is also currently in use. These vaccines are created by creating immune responses to the viral E6 and E7 components that result in aberrant cell development [www.gco.iarc.fr/].

2.5.7 Follow-up care

The survivors of cervical cancer must improve their lifestyle choices and prohibit using nicotine, alcohol, cigarettes, and other recreational drugs. This entails leading a healthy lifestyle, maintaining a nutritious diet, developing good eating habits, and maintaining exercise regime. The ACS's tobacco cessation coaching program is very helpful in increasing the likelihood that a smoker will stop (Elit et al., 2010).

Clinical follow-up may be done each 3-6 for the initial 2 years, each 6-12 months for the rest of the 3 to 5 years, and then annually in patients treated for higher-risk cervical cancer. Further tests and examinations, including scans, will be conducted if there are any concerning clinical findings or indications. The American Society of Clinical Oncology (ASCO) advises that patients who have received treatment for precancerous lesions undergo human papillomavirus (HPV) testing a year later as a follow-up procedure.

2.6 Advancement in cancer treatment

Novel therapeutics are required to enhance clinical outcomes for patients with metastatic, or recurrent cervical cancer. Researchers from all the globe are working hard to find better cervical cancer treatments and cures, thus significant progress has been made in the treatment of cervical and other forms of cancer that are explicated beneath.

2.6.1 Immunotherapy

Immunotherapy can also be considered to be a form of targeted therapy, which uses monoclonal antibodies that primarily target specific protein in cancer cells (Morris et al., 1999).

First off, the monoclonal antibody identifies the malignant cells at specific regions that are not present in normal cells. It makes the immune system work more efficiently to identify and target towards eradication of cancer cells. For instance, immunotherapy medications

target the protein PD-1, which causes cancer cells to undergo cell death (Topalian et al., 2012). Although, cancerous cells recommence their uncontrolled proliferation by producing more growth factor receptors. The monoclonal antibody binds to these receptors and prevents the growth signal from being transmitted, which inhibits the cancer cells' abnormal growth. Radio-immunotherapy is a popular form of immunotherapy that employs radioactive materials linked to the monoclonal antibody (Harris et al., 2013). Using this therapy, radiation can be delivered selectively to malignant cells instead of affecting normal cells. These radio-immunotherapy medications include tositumomab (Bexxar), ibritumomab, and tiuxetan (Zevalin). Brentuximab vedotin (Adcetris) is an antibody–drug conjugate (ADC) medication used to treat specific forms of Hodgkin lymphomas [HL] (Meister et al., 2015).

Clinical investigations are being conducted on a drug called Bintrafusp Alfa (bifunctional fusion protein), which is composed of TGF-β and the monoclonal antibody avelumab. This drug can be used along with other anticancer medications to destroy cancer cells. Sentinel lymph node biopsy is a component of immunotherapy, helping to minimize the number of lymph nodes that must be stripped away to monitor the spread of malignancy.

The method involves a radioactive tracer-containing probe to locate the cancer cells. Moreover, the dye-radioactive probe complex containing lymph nodes are located and eliminated. Robotic aided laparoscopy is one of the most recent developments in Sentinel lymph node biopsy (Niikura et al., 2012).

2.6.2. Chimeric antigen receptor modified-T cells (CAR T) cell therapy

CAR T-cell therapy is a form of immunotherapy, in which T cells are extracted from the blood by leukapheresis from the particular patient and re-engineered by adding a gene for (CAR) in the laboratory, which aids the T cells in adhering to a particular cancer cell antigen (CCA). The patient is afterward re-injected with modified CAR T cells (**Figure 16**) (Miliotou & Papadopoulou, 2018). It makes CAR T-cell therapy work as "living drug" as it is customized for each distinctive patient. The U.S. FDA approved six CAR T-cell treatments for the management of specific hematologic malignancies since 2017.
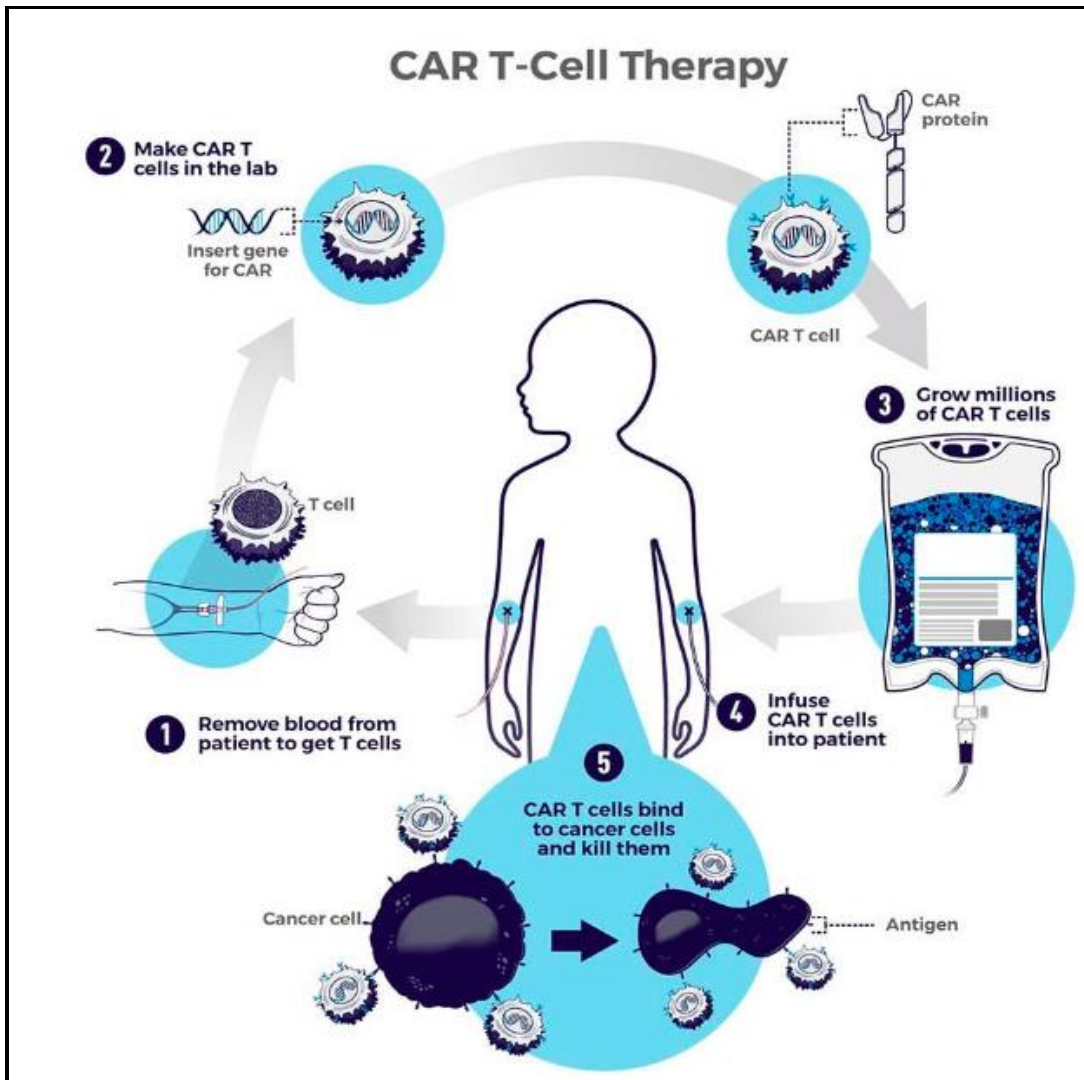
Figure 16: Schematic representation of CAR T-cell Therapy (https://www.cancer.gov/)

Human papillomavirus (HPV) is recognized by TCR engaged against an epitope of HPV16 E6 (E629-38) that is displayed by HLA-A*02:01 (E6 TCR) according to Doran et al., 2023. They studied the progression of HPV-related cervical cancer (E6 TCR T cells) using genetically modified T cells that expressed the E6 TCR. LN-145 is utilized to treat CC using an autologous adoptive cell transfer procedure to investigate the capacity of tumor-infiltrating lymphocytes (TILs) against cancer. This treatment is now being developed by Iovance Biotherapeutics (Sarnaik, 2019). Similar to this, 78% of patients with chronic,

recurrent, or metastatic CC who received cryopreserved LN-145 TIL treatment coupled with chemotherapy and the injection of (IL-2) experienced a safer decrease in cancer progression (*NCT03108495*, n.d.).

A recent clinical trial by Shenzhen Geno, Immune Medical Institute evaluated the effectiveness and reliability of CAR T cells in patients with CC, who have identified with the markers mesothelin, Mucin 1, markers disialoganglioside (GD2), prostate-specific membrane antigen (PSMA), cell surface-associated (Muc1), and few other markers (Poondla N. et al., 2021).

2.6.3 Non-specific immunotherapy

Non-specific immunotherapies (NSI) improve immune function more widely, instead of targeting specific sites in cancer cells. This can result in an overall improved immunity against cancer cells. Non-specific immunotherapies may be administered in addition to other cancer therapies like chemotherapy or radiation therapy or even as the primary cancer therapy. Certain non-specific immunotherapies are administered alone to treat cancer. Some are used as adjuvants (in addition to a primary treatment) to strengthen the immune system.

NSI treatment includes interleukins, interferons, Colony-stimulating factors (CSFs), extract from *Agaricus blazei Murill Kyowa, Mycobacterium tuberculosis* (Z-100), and Streptococcal preparations (OK-432), which have been evaluated for CC (Schmidt, Mona W., et al.2022). Among this group, Interleukins (IL-12, or aldesleukin (Proleukin)) and interferons (interferon α, Roferon-A, Intron A, and Alferon) are the key players. Furthermore, moderate flu viruses, allergic reaction (rashes), low count WBCs, hair thinning, loss of appetite are side effects of interferon therapy (Lifson et al., 2006).

2.6.3.1 Brachytherapy: (Patient experiences)

Patient experiences with brachytherapy caused to pain, anxiety, and distress as their no clear evidence of their experiences. In a study two cohorts of patients with their experiences of recently had brachytherapy and post brachytherapy and their views were prolong use of

applicators leads to discomfort in them and use of interstitial needles cause pain in them (Humphrey et al 2024). Psychological behavior of women also matter and after abnormal screening tests many women tend to show anxiety, distress and mentally upset and after surgery fear of recurrence, concern (Lloyd et al 2014).

2.6.4 Hyperthermia

Hyperthermia is a targeted, controlled elevation of body temperature through the average limits of 39^0 to $43\ ^0$C, using different electromagnetic waves, such as microwaves and radio waves, or acoustic waves. The type and stage of cancer, and its location, however, will determine the main source of energy **(Figure 17).** The treatment of CC involves local hyperthermia and regional hyperthermia. Hyperthermic intraperitoneal chemotherapy (HIPEC) is another kind of regional hyperthermia. HIPEC is a combined therapy that involves surgical removal of malignant tumors, followed by the administration of heated chemotherapeutic drugs to subsequently destroy the remaining cancer cells. The electro-hyperthermia is intended to improve the treatment efficacy by administration of 13.56 MHz short radiofrequency waves on targeted cancer cells. It prevents the recurrence of cancer and increases the longevity of the patient (Behrouzkia et al., 2016). Recent advancements in this field is the use of nanoparticles, which, when uniformly distributed inside the tumor, may transfer heat uniformly. The oncologist has to employ hyperthermia more frequently as it forms a component of comprehensive management in cervical cancer together with other therapeutic interventions.
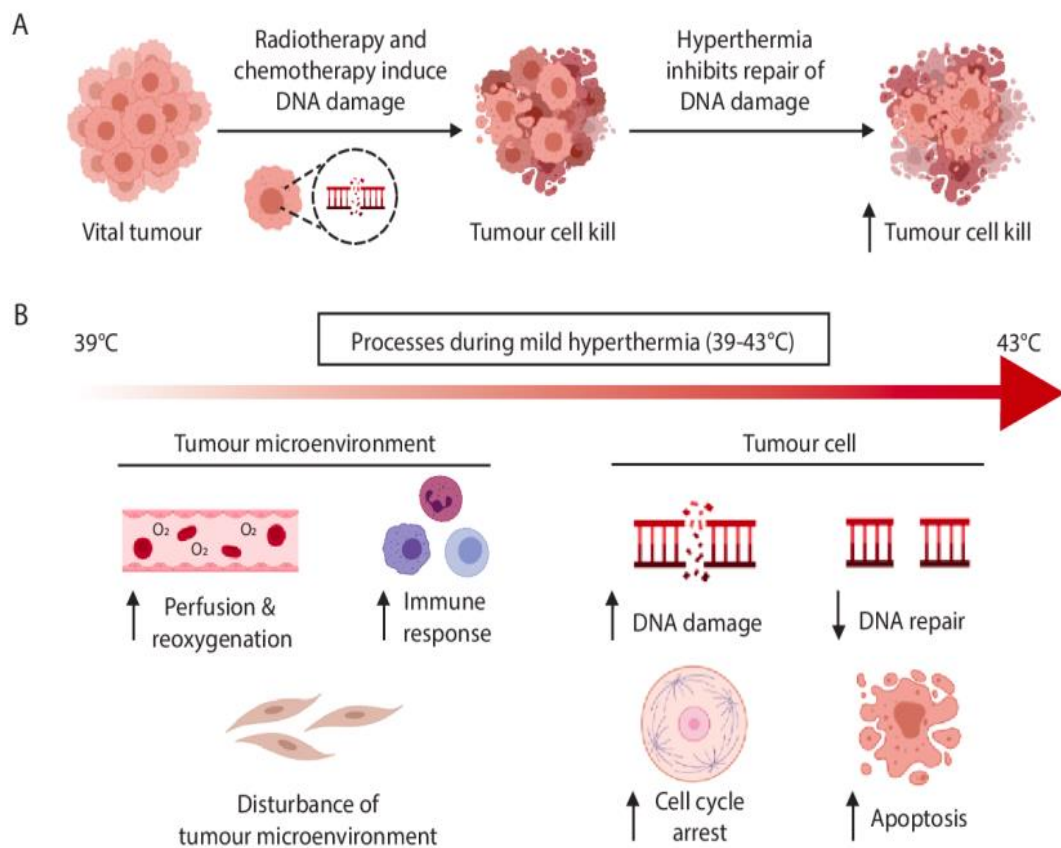
Figure 17: (A) Hyperthermia can substantially inhibit repair of radiotherapy or chemotherapy- induced DNA damage, results in increased death of tumor cells. (B) Hyperthermia affects on both the tumor cell and its surrounding environment. Through increasing perfusion and reoxygenation, hyperthermia causes discomfort at lower temperatures, starting around 39°C. Beginning around 41 °C, hyperthermia can impair DNA repair mechanisms for a brief period, leading to a buildup of DNA breaks and leads to apoptosis (IJff et al., 2022).

2.7 Epigenetic alterations in cervical cancer

Many epigenetic changes in the HPV and host genomes have been found to occur in cervical cancer at various stages throughout the disease's development. It is crucial to take into account both the epigenetic changes happening in the HPV viral genome that can affect carcinogenesis as well as epigenetic alterations in the host cellular genome since cervical

cancer development requires infection with hrHPV strains. It mainly includes histone modification, hypermethylation and hypomethylation of DNA, and RNA interference. Because epigenetic alterations manifest relatively early in the development of a cancer, they may be cast-off as molecular biomarkers for preliminary diagnosis from a clinical perspective.

Epigenetic alterations in the HPV genome, notably methylation, have been demonstrated in studies to play a significant role in the evolution of low- and high-grade CIN in HPV-associated invasive carcinoma (Baylin, 2004; Bird, 2002). It has been discovered that methylation CpG islands of the E2 protein binding site (E2BS), which are found in the (LCR) of HPV cellular genome and late promoter genes, may disable the E2 regulatory function or the expression of the L1 and L2 late genes, thereby promoting HPV-induced invasive cervical cancer (Baylin, 2004).

Several studies have found that aberrant promoter hypermethylation in cervical carcinoma has silenced or reduced the expression of p53 and pRb for many functional pathway (**Table 6**)

Gene	Rate	Function	Reference
O-6-methylguanine-DNA methyltransferase (MGMT)	5–81%	DNA repair	(Kang et al., 2005; Virmani et al., 2001; Zambrano et al., 2005)
E-cadherin (CDH1)	28–80.5%	Wnt/β-catenin signaling pathway	(C. L. Chen et al., 2003; Kang et al., 2005; Widschwendter et al., 2004)
Phosphatase and TENsin homolog (PTEN)	58%	Wnt/β-catenin signaling pathway	(Cheung et al., 2004)

antigen-presenting cell (APC)	11–94%	Wnt/β-catenin	(Dong et al., 2001; Zambrano et al., 2005)
Fanconi anemia group F (FANCF)	30%	FA-BRAC pathway	(Shivapurkar et al., 2004)
Tumor suppressor in lung cancer 1 (TSLC1)	58–65%	Tumor suppressor	(J. Li et al., 2005; Steenbergen et al., 2004)
Death-associated protein kinase 1 (DAPK)	45–100%	Metastasis/cell death	(Dong et al., 2001; Kang et al., 2005; Zambrano et al., 2005)
MutL homolog 1(hMLH1)	5%	Mismatch repair	(Virmani et al., 2001)
BReast CAncer gene 1(BRAC1)	6.1%	FA-BRAC pathway	Neveling, K. (2007)
RAS association domain family protein 1a (RASSF1A)	0–45%	Negative ras-effector	(Cohen et al., 2003; Kang et al., 2005; Kuzmin et al., 2003; Yu et al., 2003)
Estrogen receptor alpha (ER α)	25%	Steroid hormone receptor	(Zambrano et al., 2005)
Retinoic acid receptor beta (RARβ)	33–66%	Cell differentiation	(Feng et al., 2005; Ivanova et al., 2002; Zambrano et al., 2005)
hypermethylated in cancer 1 (HIC1)	18–45%	Transcription factor	(Dong et al., 2001)
Tissue Inhibitor of Metalloproteinase (TIMP2/TIMP3)	47%/1–10%	Tissue inhibitor MTs	(Ivanova et al., 2004; Kang et al., 2005; Widschwendter, Müller, et al., 2004)

Caveolin-1 (CAV1)	6%	Caveolae membrane	(Chan et al., 2003)
Fragile Histidine Triad Diadenosine Triphosphatase (FHIT)	11–88%	DNA repair/cell death?	(Dong et al., 2001; Kang et al., 2005; Virmani et al., 2001; Zambrano et al., 2005)

Table 6: Hypermethylated Tumor suppressor genes in Cervical cancer

A study of 50 CC patients identified methylation of either DcR1 or DcR2, the apoptosis related genes, indicating that cervical carcinoma is promoted by downregulating decoy receptor expression (Shivapurkar et al., 2004).

In most malignancies, the Wnt/β-catenin pathway is the one most often aberrantly affected and thus, facilitates tumorigenesis. The metabolic regulator PTEN tumor suppressor gene affects the WNT pathway by preventing the stimulation of ILK, which in turn inhibits GSK-3 and leads to the buildup of β-catenin (Martelli 2016). A research study reported that phosphatase and tensin homolog (PTEN) gene mutations were not present in 62 patient samples of squamous cell carcinoma (SCC), 58% of cases showed methylation of promoter gene. It's interesting to note that individuals with chronic disease who passed away from the disease had much greater incidence of PTEN methylation levels than patients with no signs of recurrence. Similarly, Dueñas-González et al (2005) reported that the tumor suppressor gene E-cadherin (CDH1) was methylated with a frequency that varies between 28 to 80.5% in cervical carcinoma (C. L. Chen et al., 2003; Kang et al., 2005; Widschwendter, et al., 2004). Hypermethylation occurs more frequently in endometrial cancer than in cervical cancer, with rates ranging from 10 to 95% (Dong et al., 2001; Zambrano et al., 2005), especially in MSI+ endometrial tumors.

It has been reported that the dysfunction of mismatched repair genes such as hMLH1, hMSH2, hMSH3, hMSH6, and hPMS2, exhibit higher mutation rates compared to normal cell (Jass 2004). However, there is limited data available on CC in terms of hMHL1 expression and methylation.

2.8 Histone alterations in cervical cancer

The remodeling and transcriptional capability of the chromatin are affected by post-translational changes of histone tails, such as histone acetylation, histone deacetylation, methylation, histone phosphorylation, binding of small ubiquitin-related modifier (SUMOylation), and ubiquitination. These chromatin modifications are important for the management of cell cycle, epigenetic inheritance, cellular proliferation and differentiation, and stem cell regulation (Bornelöv et al., 2018; Soto et al., 2017).

The molecular basis of the chromatin modifications was yet unknown. According to a recent study, cancerous cells lost their monoacetylated and trimethylated histone H4 forms during the tumorigenesis, primarily at the acetylated Lys16 and trimethylated Lys20 residues of histone H4, which are linked to hypomethylation of DNA repetitive sequences, a characteristic of cancer cells (Fraga et al., 2005).

Histone H3 alterations have been strongly linked to the development of CC from early stage CIN I to advanced stage CIN II and CIN III (Anton et al., 2004). High-risk HPV-E7 proteins interact to HDACs, and this interaction takes place through Mi2 protein (SWI2/SNF2 type helicase/ATPase domain-containing protein) that is a component of the nucleosome remodeling and histone deacetylation (NRHD) complex, which has the capacity of chromatin remodeling through both histone deacetylation and reorientation of ATP-dependent nucleosome (Longworth & Laimins, 2004).

2.9 Methylated genes as biomarkers for CC

Many methylated genes serve as indicators in cervical cancer that may be used for HPV screening and early diagnosis. The methylation of genes found in the serum or plasma of individuals with CC has been investigated in relation to its prognostic factors.

A recent study reported that different methylation rate of five genes from the serum sample of cervical cancer patient which includes calcitonin related polypeptide alpha (CALCA (62%), Telomerase reverse transcriptase (hTERT) (0%), Myogenic Differentiation 1 (MYOD1) (25%), progesterone receptor (PR) (79%), and tissue inhibitor of matrix

metalloproteinases (TIMP3) (4%). MYOD1 methylation was highly linked to short term median survival being 1.9 for methylated versus 6.1 years for unmethylated patients of cervical cancer, respectively (Widschwendter, Müller, et al., 2004). In a similar study, a cohort of 93 patients were evaluated for diagnosis and treatment importance with respect to methylation of CDH1 (E-cadherin) and CDH13 (H-cadherin) genes. The results revealed that 43% of the patients of cervical cancer showed methylation either in E-cadherin (CDH1) or H-cadherin (CDH13). It showed a median survival of 1.2 years for patients with methylated genes compared to 4.3 years for unmethylated CDH1/CDH13 in CC patients (Widschwendter, Ivarsson, et al., 2004). Yang et al. (2001) studied (DAPK 1), O-6-methylguanine-DNA methyltransferase (MGMT), and p16 gene to observe methylation of promoter genes in cervical cancer. The methylation was observed as normal occurrence in case of CC with methylation frequencies of 40% for DAPK, 10% for p16 and 7.5% for MGMT gene. This also established a significant correlation between primary tumor progression and methylation biomarkers in serum (H. J. Yang et al., 2004). Therefore, these findings facilitate additional research to identify a group of methylation genes with prognostic importance as well as potential use as proxy indicators of the success of epigenetic therapy.

2.10 Whole Exome Sequencing in CC

Whole exome sequencing (WES) is an advanced sequencing technology which analyze all the exome i.e. protein coding sequences in a particular genome. This technology provides a clinically comprehensive framework for patient diagnostics since the majority of disease related genetic aberrations fall into this group. Effective treatment strategies continue to be a medical necessity for patients with advanced stages of cervical cancer. WES plays a significant role for improved understanding of carcinoma genomics. We can compare the WES data of cancer cells with the normal healthy cells from the same patients. It is possible to discover elements that trigger tumorigenesis that is induced by genetic mutations found in exomes of cancer cells. By examining numerous mutations found in cancer cell exomes, it is feasible to discover mechanisms responsible for undesirable tumor growth. WES enables the detection of copy number variations (CNVs) and (SNVs). The data set of SNV

variant calling can be used together with machine learning-based pipelines for better understanding. An essential step in the detection of such brief variants is accurate variant calling. The most widely used variant calling applications are MuTect (Cibulskis et al., 2013), VarScan2 (Koboldt et al., 2012), Somatic Sniper (Larson et al., 2012), Strelka (Saunders et al., 2012), and FreeBayes (Garrison & Marth, 2012). Furthermore, these tools can be used in various clinical investigations for variant calling, either in the integrated way or alone. Also, a combination of these technologies was employed for variant calling in a number of clinical studies (Carbone et al., 2017; Cristescu et al., 2018; Hellmann et al., 2018; Hugo et al., 2016; Le et al., 2015; Riaz et al., 2017; H. Rizvi et al., 2018; N. A. Rizvi et al., 2015; Snyder et al., 2014, 2017; Van Allen et al., 2015). Figure represents the basic procedure of WES and Table represents a complete list of all commonly used available tools.

A recent study examined 65 CC cell lines, the majority of which had the HPV type-16/18, by using whole-exome sequencing (WES). Researchers used exome analysis to find recurrent somatic missense mutations in 22 genes and to count 12 copy number gains and 40 copy number losses. The chromatin remodeling, apoptosis pathways and signaling through ERBB2/PI3K/AKT/mTOR genes were linked to copy number alterations and single nucleotide variations. The study suggests that the considerable percentage of cervical cancers (>70%) may benefit from ERBB2/PIK3CA/AKT/mTOR-targeted medications, according to research using completely sequenced cell lines, xenografts, and current inhibitors and medicines (Zammataro et al., 2019).

Bao et al (2021) performed whole-exome sequencing (WES) on 25 cases of CIN and 10 cases of CC with surrounding neighboring tissues (Bao et al., 2021). The C>T transitions were detected in both CINs and CCs. Contrarily, a t-test with a P value of 6.60E-04 revealed that CCs had a larger proportion of somatic mutations than CINs. In addition, PIK3CA, Rho GTPase Activating Protein 5 (ARHGAP5), and Adhesion G protein-coupled receptor B1 (ADGRB1) were discovered as putative driver genes in this study, with ADGRB1 being the first to be described in CC. PIK3CA was one of the most frequently

mutant genes linked to CC. Previous studies of Ojesina et al and Huang et al. reported that PIK3CA was exclusively found in CCs and was significantly greater than that reported in earlier genome sequencing studies. Variant calling can offer the most fundamental and thorough data foundation. It can also precisely, and effectively assess the differences between genomes by analyzing the entire genome sequence, and obtain the most comprehensive information for the development of effective molecular biomarkers in cervical cancer.

Some of the tools used for the analysis in whole exome sequencing steps (Table 7)

Procedures	Tool	Website Link
Genome browser	Ensembl	www.ensembl.org
	UCSC	http://genome.ucsc.edu
Quality control	FastQC	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Short read mapping	Bowtie	http://bowtie-bio.sourceforge.net/index.shtml
	Bfast	http://bfast.sourceforge.net
	Mosaik	https://github.com/wanpinglee/MOSAIK
	BWA	http://bio-bwa.sourceforge.net/
Manipulate NGS data	Picard tools	https://broadinstitute.github.io/picard/index.html
	SAMTools	http://www.htslib.org/doc/samtools.html
Variant calling	GATK	https://software.broadinstitute.org/gatk/
	SAMTools	http://www.htslib.org/doc/samtools.html
Variant annotation: (1) Coding	SnpEff	http://snpeff.sourceforge.net/
	VEP	http://ensembl.org/info/docs/tools/vep/index.html
	SIFT	http://sift.jcvi.org/
	PolyPhen2	P http://genetics.bwh.harvard.edu/pph2/

effect predictions		
Variant annotation: (2) Conservation	PhyloP	http://compgen.bscb.cornell.edu/phast
	GERPflfl	http://gvs.gs.washington.edu/GVS147/
	CADD	http://cadd.gs.washington.edu/
Variant annotation: (3) Gene-level	MSC	http://lab.rockefeller.edu/casanova/MSC
	GAVIN	https://molgenis20.gcc.rug.nl/
Variant annotation: (4) integrative	ANNOVAR	http://annovar.openbioinformatics.org/en/latest/user-guide/download/
Knowledge-based annotation	HGPS	http://hgc.rockefeller.edu/
	KEGG	https://www.genome.jp/kegg/
	REACTOME	www.reactome.org/
	MPO	www.informatics.jax.org/humanDisease.shtml
	GEO	www.ncbi.nlm.nih. gov/geoprofiles
	GXA	www.ebi.ac.uk/gxa
	BioGPS	http://biogps.org
	STRING	http://string-db.org
	ToppGene.	https://toppgene.cchmc.org
	GeneMania	http://genemania.org

Table 7: Common tools and website links for WES data analysis pipeline
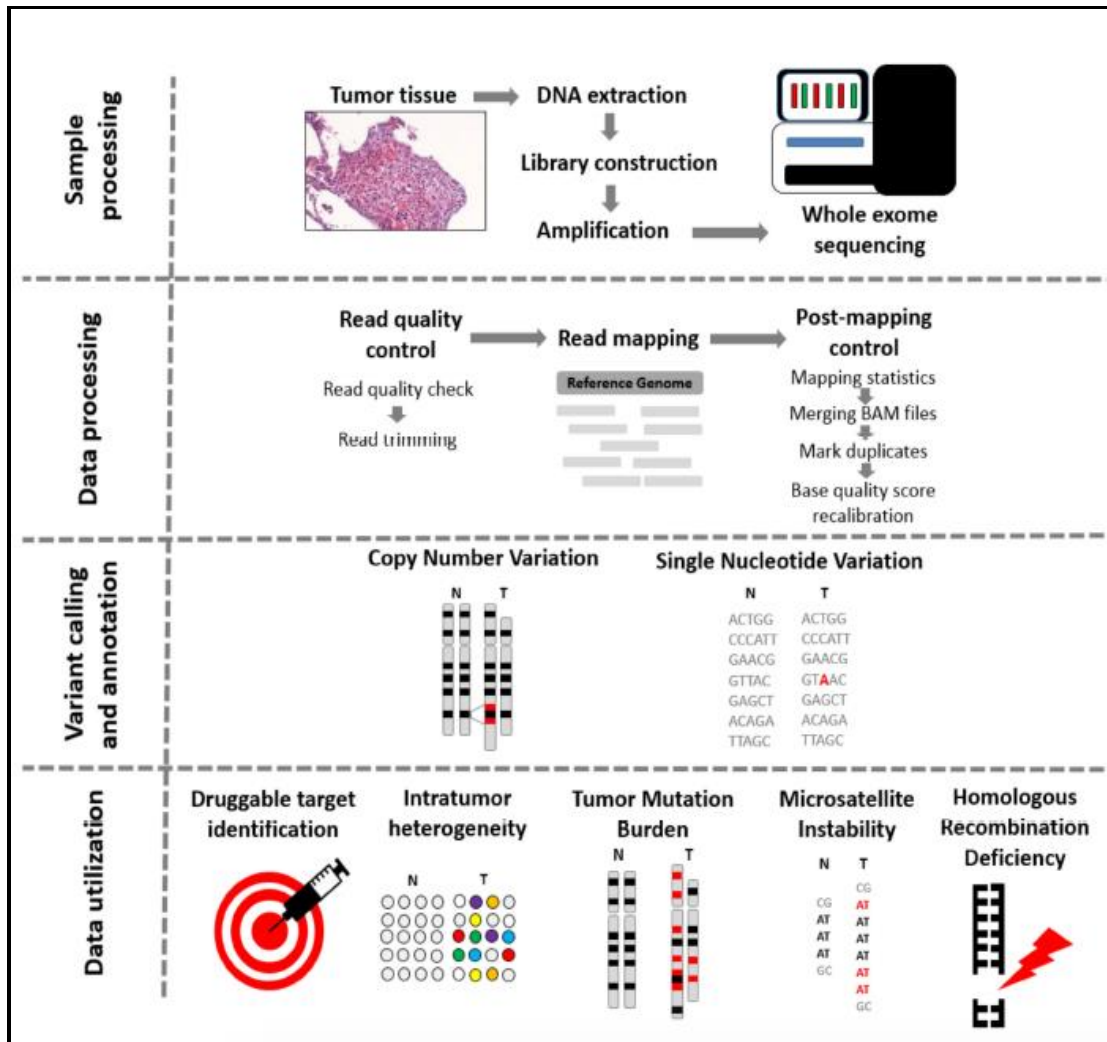
Figure 18: Steps of Whole exome sequencing, sample processing, library preparation and variant calling and data utilization (Courtesy to Bartha & Gyorffy 2019)

The sample preparation starts with the DNA extraction from tumor tissue, followed by library preparation, amplification and sequencing. An accurate quality check is essential for a valid data analysis. Following raw data quality check (QC) and preprocessing, the reads are mapped to the reference genome. QC is followed by variant analysis, which is comprised of three distinct steps: variant calling, annotation, and prioritizing. is Binary Alignment Map (BAM) files are examined to analyze single nucleotide variants calling, copy number variation, insertion –deletion data. The dataset can also be used to various other parameters (Bartha & Győrffy. 2019) **(Figure 18).**

Transcriptomics

One unregulated gene that has been identified is bone gamma-carboxy glutamic acid-containing protein (BGLAP). Osteocalcin, also known as, is an 11-kDa protein hormone secreted by osteoblasts. BGLAP primarily regulates bone matrix mineralization and blood coagulation. The transcriptome research conducted by Bettadapura et al. in 2018 revealed elevated BGLAP expression in cervical carcinomas. Their research found that BGLAP overexpression is present exclusively in cancerous tissues, not in normal tissues, suggesting it could be a valuable target for therapy in this type of cancer. Five separate transcriptome datasets of cervical cancer were analyzed, revealing 113 down-regulated differentially expressed genes (DEGs) and 199 upregulated DEGs. The dysregulated core genes encode proteins related to hormones, signaling molecules, enzymes, receptors, transporters, and structural proteins. Down-regulated proteins consist of KAT2B, ESR1, WNK1, and FGFR2. On the other hand, the proteins that are increased in expression are GSK3β, PARP1, PCNA, and CDK1. Their research identified disrupted genes related to metabolism pathways in colon cancer by integrating transcriptome data with the genome-scale metabolic network. Their study found that the arachidonic acid metabolism pathway is a potential therapeutic target, and the mentioned indicators could be used for CC screening or treatment (Kori and Arga 2018).

Another factor contributing to cancer is the variability in gene expression that leads to its development. MicroRNAs (miRNAs) have a beneficial impact on gene expression and regulation. MicroRNAs are short single-stranded non-coding RNAs that primarily interact with the 3' untranslated regions (3' UTRs) of mRNA to regulate gene expression by either inhibiting or fully suppressing the mRNA. Cancer tissues or cells typically exhibit upregulation of oncogenic miRNAs or downregulation of tumor suppressor miRNAs, leading to the promotion of cancer expression. Under these circumstances, analyzing the transcriptome using RNA-sequencing (RNA-seq) has become a potent technique for examining gene expression patterns. Around 250 miRNAs exhibit differential expression in CC cells. MiRNA-214, also known as miR-214, is decreased in all three stages of

cervical intraepithelial neoplasia (CIN) (Yeung et al. 2017). (Sen et al., 2020) conducted transcriptome analysis to evaluate gene expression in C33A and CasKi CC cell lines following miR-214 knock-down and knock-in. Research has found that higher amounts of cell death (apoptosis) occur when miR-214 is overexpressed, and this effect is reversed when miR-214 is knocked out. Therefore, these genes could also serve as potential biomarkers for characterizing tumor prognosis in different cancer stages, including CC.

Through the utilization of an Illumina microRNA microarray platform, J. H. Li and colleagues (2011) detected several miRNAs with varying expression levels in cervical squamous cell carcinoma and nearby non-cancerous tissues. 1145 miRNAs were discovered, and 7 of them, such as miR-886-5p, showed significant variations between tumor and non-tumor tissues. miRNA microarray data and bioinformatics research showed that HPGD (15-hydroxyprostaglandin dehydrogenase) can restrict cell proliferation and movement, since it is a specific target of mi146b-3p in CC (Yao et al. 2018).

Long non-coding RNAs (lncRNAs) are another form of RNA that actively regulates transcription factors. These are long functional RNA molecules that do not code for proteins and consist of over 200 nucleotides. In 2018, Liu et al. conducted microarray tests that revealed 1621 lncRNAs and 1345 mRNAs expressed differently in high and low-risk squamous cell carcinoma. In 2018, a different team of researchers (J. Huang et al.2018) utilized an HT microarray to analyze and compare the expression profiles of lncRNAs and mRNAs in malignant and normal tissue samples. 5844 lncRNAs and 4436 mRNAs were identified as having altered levels of expression in CC compared to normal tissues. LncRNA microarray tests have demonstrated that lncRNAs are crucial in regulating cancer genes through either increasing or decreasing their activity. Jing et al. (2015) conducted research on the lncRNA HOTAIR because of its significant impact on controlling the expression of many carcinogenic genes. The researchers detected HOTAIR expression in four CC cell lines (CaSki, C33A, HeLa, and SiHa cells) and observed significantly greater

levels of HOTAIR in HeLa and SiHa cells compared to CaSki and C33A cells. Additionally, it was found that the lncRNA HOTAIR suppresses p21 in CC, facilitating cell cycle advancement, growth, movement, and infiltration, thus acting as an oncogene that causes CC when overexpressed. LncRNA-LET functions as a tumor suppressor, and decreased levels of it result in CC. A study conducted by Jiang, Wang, and Yang in 2015 found that patients with colon cancer who had decreased levels of lncRNA-LET experience a notably reduced overall survival rate compared to those with increased levels of lncRNA-LET (Duppala et al.2022)

Combining WES, Metabolomics and transcriptomic analysis has recently shown a significant progression in CC research. Metabolomics is widely used to understand cancer metabolism and identify biomarkers to infer the onset and progression of cancer. (Yang et al., 2017) used metabolomic profiling with transcriptomics data to validate the potential diagnostic biomarkers of CC. They selected five metabolites as CC biomarker candidates, which had a brilliant performance in distinguishing between CC and normal. It has been proven to be a viable approach for CC screening and diagnosis. There were several confirmed combinational analysis of metabolomics and transcriptomics as promising methods to examine the mechanism of carcinogenesis and discover more consistent biomarkers.

2.11 Databases, Artificial Intelligence and System biology approaches

Machine learning (ML) employs iterative algorithms and statistical approaches and generate patterns from huge amounts of complex data set (Beniczky et al., 2021). These virtual models can be used for disease prognosis, drug interactions, drug sensitivity, and use of chemotherapeutics (Chu et al., 2020; Huang et al., 2022; Oyaga-Iriarte et al., 2019; Xia et al., 2021; Y. Yang et al., 2022). Among the multiple supervised ML algorithms for medical data, a random forest (RF) approach demonstrated greater accuracy (Uddin et al., 2019). There are various databases which can be used for ML algorithms such as The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), Kyoto Encyclopedia of Genes and Genomes (KEGG) etc for better understanding of biological systems. The

Cancer Genome Atlas (TCGA) is a repository of the most extensive sequencing data, offering researchers complete cancer genomic datasets on different stages of tumor, malignant transformation, prognosis, treatment history, gender, and associated clinical data.

Li et al, (2017) analyzed the altered mRNA gene expression in from the 252 samples of cervix squamous cell carcinoma and 3 samples of normal tissue using RNA-Seq data from The Cancer Genome Atlas (TCGA) (X. Li et al., 2017). The gene subnetworks were determined for gene profiling, molecular and functional characterization of a biological pathway, protein –protein interaction (PPI) network analysis, using different dataset such as GSEA and KEGG pathway analysis. (Figure) The combined strategy demonstrated a substantial correlation between the rate of survival and DNA repair, which included 18 histone genes (Histone cluster 1 H2A, B, and H4). Additionally, five of these histone genes were shown to be highly expressed in three groups of CC, determined by the Oncomine database. It showed that two systemic lupus erythematosus SLE-associated gene sets (*HIST1H2BD, HIST1H2BJ; HIST1H2BD, HIST1H2BJ, HIST1H2BH, HIST1H2AM, and HIST1H4K*) can be employed as predictive markers for cervical cancer survival prediction.
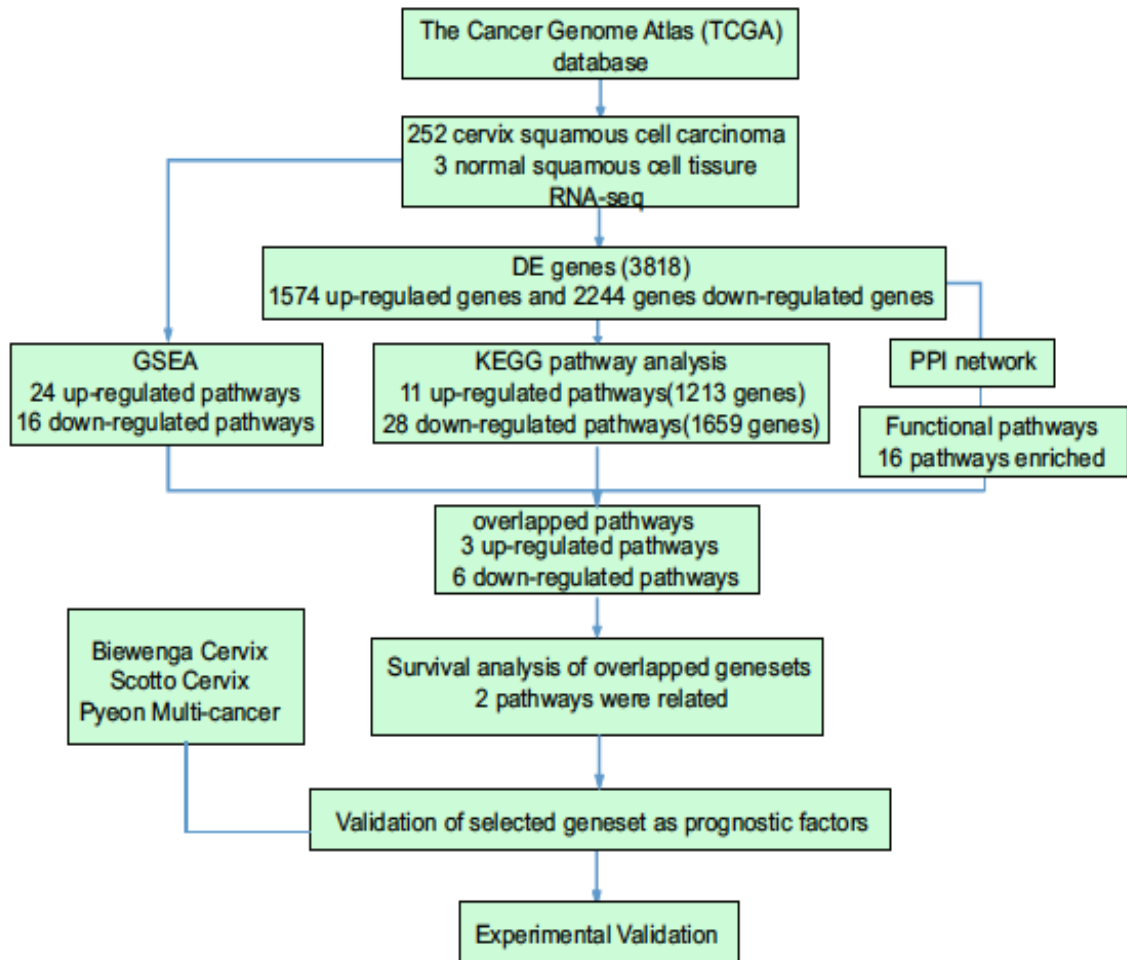
Figure 19: Basic workflow of the study (X. Li et al., 2017)

Lin et al. (2021) worked to identify viable antiviral treatments for HPV infection by comparing a number of ML models (Lin et al., 2021). For the purpose of creating prediction models, 6 different types of machine learning classifiers suitable for binary classification were chosen. Support Vector Machine, Random Forest, AdaBoost32, Logistic Regression, Naive Bayes33, and K-Nearest Neighbor classifier were the predictors that were selected. K-Nearest Neighbor classifier and Adaboost, among others,

reported with strong predictive abilities for identifying miRNA-disease associations (X. Chen et al., 2017; Y. Zhao et al., 2019).

However, they identified the optimized Support Vector Machine and K-Nearest Neighbor classifier with high precision score as the best two predictors (0.80 and 0.85 respectively) using 182 pairs of antiviral-target interaction dataset. Among 864 pairs of antiviral-HPV protein connections, we successfully predicted 57 pairs of antiviral-HPV protein interactions.

ML can analyze large amounts of data and build models from the dataset. ML entails the molecular make-up and genetic composition of cancer tumors in specific patients. The virtual models can predict cellular alterations that might be promoting malignancy and provide insight to create customized therapies.

Based on the above review of literature this study is designed for further hypothesis and workflow on whole exome sequencing, transcriptomics and system-based approaches the work is established.

Chapter 3

Hypothesis

The link between HPV infection and CC is well established, but there are gaps in understanding the specific mechanisms by which certain HPV types lead to the development of CC. Further, research into molecular and genetic factors could improve preventive strategies and therapeutic interventions. There are enormous technologies that enhance the prognosis, diagnosis, and treatment methods in the current generation. Several scientists and researchers are working on cervical cancer treatment and therapeutic approaches, but still, there is a lack of chemoradiotherapy for advanced metastasis and recurrent cervical cancer patients. Identifying reliable and specific biomarkers for early detection of CC and recurrent and metastasis stages, there is a need to find new biomarkers that can enhance the sensitivity and specificity of screening methods, improving the chances of early diagnosis and successful treatment. There is a need for patient-personalized treatment, and this helps clinicians diagnose the patient based on the data generated, which will enhance the proper treatment. Developing the biomarkers to predict individual responses to specific therapies requires more exploration to predict which patients are likely to respond best to particular therapies.

HPV genotyping enables precise **risk stratification**, identifying women who are most at risk of progressing to cervical cancer. This helps focus prevention strategies on individuals with persistent high-risk infections, improving the efficiency of screening programs and vaccine deployment. SSCP helps identify **early genetic mutations** that may precede visible cellular changes. This can lead to **earlier detection** of precancerous lesions or cancerous changes, allowing for timely intervention before the cancer progresses to an invasive stage. WES provides critical insights into **individual genetic variations** that influence cancer development, progression, and response to treatment. This knowledge enables the development of **personalized treatment plans**, identifying patients who may benefit from specific targeted therapies or immunotherapies. It also aids in discovering new therapeutic targets for drug development and bioinformatics bridges the gap between complex genomic data and **clinical application**. It enables the identification of

biomarkers for early detection, prognosis, and response to therapy. These technologies like HPV genotyping, SSCP, Whole Exome Sequencing and bioinformatics data analysis are considered to specifically identify the gaps in the HPV and CC.

The present study has been planned to evaluate the insights into integrating whole exome sequencing data in cervical cancer that will elucidate novel somatic mutations and novel mutational signatures that identify potential therapeutic targets, thus enhancing our understanding of the genomic landscapes and molecular mechanism underlying cervical carcinogenesis and also the integration and its impact will be assessed **(Figure 20).**
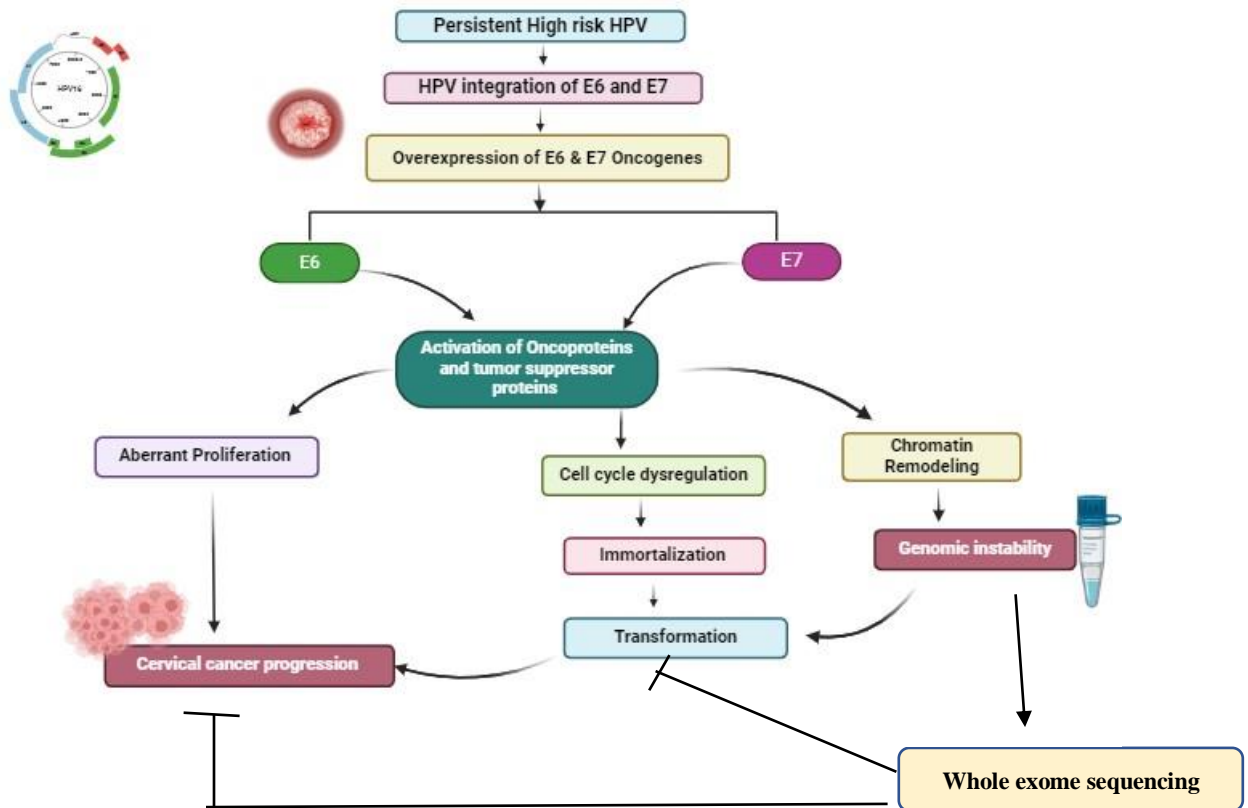
Figure 20: Schematic View of the Representation of the Hypothesis

Chapter 4
Aims and Objectives

Aim

To integrate system biology approaches through Whole Exome Sequencing in Cervical cancer samples.

Objectives:

1. To catalogue the Human papillomavirus (HPV) through genotyping and phylogenetic analysis.

2. To evaluate the genetic variants through whole exome sequencing.

3. Prioritize variants, validate specific markers (SSCP, Sanger sequencing, Methylation and RNA-Seq analysis), and establish their associations with disease causation per se.

4. To perform integrative analysis using system biology-based approaches.

Chapter 5
Materials and Methods

5.0 To catalogue the Human papillomavirus (HPV) through genotyping and phylogenetic analysis

5.1 Epidemiological Population study

Controls and malignant cervical lesions were obtained from untreated patients aged 24-70 years. The tissue samples of cervical cancer were gathered from Mehdi Nawaz Jung (MNJ) government cancer hospital in Hyderabad, Telangana, and Homi Bhabha Hospital, Vishakhapatnam from the southern part of India, were confirmed through histological examination. Control samples of cervical tissues were obtained from women who underwent hysterectomy at Shroff Memorial hospitals for varied gynecological conditions. The MNJ institutional and departmental ethics committee board of Osmania University, Hyderabad, accepted the comprehensive protocols for collecting samples, preservation, and procedure. Written consent was obtained through signature or thumbprint, and epidemiological data such as family history, smoking status, hormonal and menopausal state, rate of parity, dietary habits, age around the time of marriage, and financial situation were collected from all the patients who participated in this study.

5.2 Selection criteria

5.2.1 Inclusion criteria

The participants were chosen to be females who were at least 21 years old, had a cervix that was not damaged, and had no previous history of cervical cancer.

5.2.2 Exclusion Criteria

The study excluded pregnant women, individuals with diabetes, individuals with hypertension, and anyone diagnosed with any other form of cancer*

5.3 Sample collection

5.3.1 Collection of control samples

Biopsies were attained from women who underwent for gynecological issues, specifically from normal cervical tissue or surrounding tissue. The obtained tissues were quickly placed in RNA later solution (100 to 80ml) and subsequently preserved in a -80 ^{0}C and -20^{0}C freezer until needed. At the same time, a portion of the biopsy sample was placed into a vial comprising a solution of formalin and dispatched to the pathology department. This study only included histopathologically confirmed normal tissue samples.

5.3.2 Collection of Patient Samples

The study included patients who exhibited histopathologically verified untreated squamous cervical malignancy. A biopsy sample was preserved in RNA later solution to facilitate DNA and RNA analysis. The study exclusively encompasses samples that were verified using histopathology. The female patients who visited the hospital with abnormal discharge, heavy bleeding, itching, lower abdominal pain, painful intercourse, back pain, foul smell discharge, colour change, and infertility. They all underwent liquid-based cytology tests and pap smear tests for examination. The cervical scrape was collected using cytopathic cytobrush and transferred to phosphate-buffered saline containing 450 ml, immediately kept at -20 degrees, and a smear was prepared for cytological examination.

5.4 Demographic Study Method

5.4.1 Age

The study included healthy women aged between 24 and 70, while the group of women with diseases also ranged from 24 to 70 years old. The mean age and parity exhibit a statistically significant difference ($p<0.05$) between the patients and controls. There wasn't any notable disparity in the average age of marriage among sick and healthy women (Deepti et al 2023).

5.4.2 Smoking

Women who engaged in the consumption of smoking whether through cigarettes or chewing tobacco products such as pan masala or gutka, were classified as smoking/tobacco users. On the contrary, women who never used tobacco were categorized as non-smoking/tobacco users.

5.4.3 HPV status

If woman is detected with HPV types of low-risk or high-risk they are categorized under HPV positive and those who do not show any genes or HPV types they are classified as HPV negative.

5.4.4   Financial Status and Education

Most of the members are from very low class and middle class and few are illiterate. A total sample of seventy were included in the study control samples were 5 and tumor tissue samples of 20. Most of them are illiterate and financially weak.

5.5 DNA extraction or genomic DNA isolation

5.5.1 Principle

Through the use of sodium dodecyl sulfate (SDS), the cell wall was broken down, and DNA was shielded from the action of endogenous nucleases through the utilization of EDTA to chelate the essential co-factor Mg2++ ions. Proteinase K was utilized to break down the proteinase enzyme. Chloroform and phenol were utilized to denature the proteins, which resulted in the separation of the denatured protein layer through centrifugation. Lastly, DNA was precipitated through the use of chilled 100% ethanol or isopropanol.

5.5.2 Materials and Reagents Used

The QIAamp DNA Mini Kit, with catalog number 51304 is used which have following contents:

- Collection Tubes (2 ml)

- Buffer AL

- Buffer ATL

- Spin column

- Spinner

- Proteinase K solution Buffer AW1

- Buffer AW2

- Buffer AE

- Water bath (80^0 heated)

- Microcentrifuge

- Tissue homogenizer

5.5.3 Procedure (based on manufacturer's Qiagen protocol)

- Using a sterile scalpel, finely cut 25 milligrams of tissue as an example.

- Place it in a centrifuge tube with a capacity of 1.5 milliliters, and then add 20 microliters of proteinase K and 180 microliters of buffer ATL. Incubate the mixture at a temperature of 56 degrees Celsius for a period of one to three hours.

- Incubate for ten minutes at a temperature of seventy degrees Celsius and vortex for fifteen seconds.

- After adding 200 µl of 100% ethanol, it was vortexed for 15 seconds. After that, it was very carefully collected into the spin column, and then it was centrifuged for one minute at 8000 revolutions per minute.

- Following the addition of 500 microliters of AW1 and AW2 buffer into the column, the column was centrifuged for one minute at 8000 rpm and 140000 rpm. The supernatant was discarded, and the DNA was eluted in 200 microliters of elution buffer. The DNA was then stored at a temperature of -20 degrees Celsius, after being checked in 0.8% agarose gel. (HB-0329-005-HB-QIAamp-DNA-Mini-Blood-Mini-0623.pdf)

5.5.4 DNA PCR Check

To ensure the integrity of DNA quality, a bio spectrophotometer (Eppendorf) was utilized to determine the concentration of DNA. Additionally, β-globin polymerase chain reaction (PCR) was carried out with specific primers. A list of primer sequences for β-globin is provided below.

GH20(F):5'-GAAGAGCCAAGGACAGG TAC-3 PC04(R): 5'-CAACTTCATCCACGTT CACC-3 (Silva et al., 2005) has been described as the following:

5.5.5 Cycling conditions

S.NO	Steps	Temperature	Duration	No. of cycles
1	Initial Denaturation	95°C	5min	1
2	Denaturation	95°C	30sec	
3	Annealing	58°C	30sec	35
4	Extension	72°C	30sec	
5	Final extension	72°C	5min	1
6	Hold	4°C		

Table 8: PCR conditions for DNA extraction

5.6 Agarose gel Preparation

A gel composed of 2% Agarose Preparation:

1. Two grams of agarose, labeled with the SeaKem® LE Agarose Catalogue number #50004, was put to a conical flask that contained one hundred milliliters of 1X TBE buffer.

2. The flask was heated in a microwave oven until the agarose was entirely dissolved as the temperature increased.

3. A volume of 5 microliters of ethidium bromide with a concentration of 10 milligrams per milliliter is introduced into the solution.

4. After pouring the solution into the agarose gel-casting tray, the solution is allowed to solidify at room temperature for some time.

After running the amplified PCR products on a gel composed of 2% agarose with a 100 bp ladder (HI-media) at 80 volts for thirty minutes, the bands were observed and recorded with the use of a gel documentation system. The Bio-Rad. Samples that were found to be positive for β-globin PCR were chosen for HPV PCR analysis.

5.7 HPV testing by PCR method

HPV screening was conducted on DNA samples using GP5+/GP6+ primers. The PCR product was subjected to agarose gel electrophoresis after amplification. The reference is from Maria and Husman's work published in 1995.

Primer sequences for GP5+/GP6+:

The forward primer, GP5+, has the sequence TTTGTTACTGTGGTAGATACTAC, while the reverse primer, GP6+, has the sequence GAAAAATAAACTGTAAATCATATTC.

5.7.1 Cyclic conditions

S. No.	Steps	Temperature	Duration	No. of cycles
1	Initial Denaturation	95-96°C	5min	1
2	Denaturation	95°C	30sec	
3	Annealing	52°C	30sec	40
4	Extension	72°C	1min	
5	Final extension	72°C	5min	1
6	Hold	4°C		

Table 9: PCR cycling conditions for HPV testing

The PCR reaction consists of 20 microliters containing the following components: 50 millimolar potassium chloride (KCl), 10 millimolar Tris-hydrochloride (Tris-HCl) with a pH of 8.3, 200 micromolar deoxyribonucleotide triphosphates (dNTPs), 1.5 millimolar magnesium chloride (MgCl2), and 20 picomoles of forward (F.P) and reverse primers (R.P) (GP5+/GP6+). Use 1 unit of Taq DNA polymerase and 1x Taq buffer according to the settings mentioned above. The band observed at 150 base pairs in agarose gel

electrophoresis was verified to be a positive sample of HPV DNA template. This sample was used as a reference ((Badillo Almaraz 2014).

5.8 HPV Genotyping

5.8.1 Reagents

- Nuclease-free water
- Buffer
- Effendoff Tubes
- PCR
- Primers
- 10% Polyacrylamide
- *RsaI* enzyme
- Gelstar or gel documentation (Bio-rad)

5.8.2 Procedure

After Gel checks, the positive samples are subjected to RSA1 restriction enzyme, *RsaI* enzyme dilution is performed for the stock solution of 20 units, and 40 micro-litres for 20 samples are prepared accordingly. For stock enzyme 2.0 µl, Nuclease free H20 – 20 µl and buffer 18 µl was added to make the working solution and distributed in effendorff tubes. 8.0µl PCR Product, 4 µl *RsaI,* are incubated at 37^0 C overnight or for 8 to 9 hrs in a water bath, and after that, performed PAGE gel check is at 100V.

HPV genotyping was performed in all the cytology samples with specific primers of L1 gene MY9 and MY11 and following the cost-effective method of Restriction fragment length polymorphism (RFLP) used for amplification. The consensus DNA fragments of approximately 450bp and 70 bp (Molano, M., et al. 2002), HPV positive ladder is also used to compare the DNA strands or base pairs. The experiment was conducted twice to remove any inappropriate results. Each primer was specific and checked for no false positive effect. PCR–RFLP was performed as it was a monetarily advantageous method to perceive and characterize the HPV-DNA in clinical cases when compared. The HPV-positive samples were genotyped with the RFLP method. The restriction enzyme reaction was performed

discreetly by final volume 20 microliter, using 1g of MY9/11 PCR product 2l of 10x restriction buffer was added to 10 units of *RsaI* according to instructions given by the manufacturer. The digestion was performed in 10% polyacrylamide gels electrophoretically in the occurrence of 50bp DNA; gels were stained with Gelstar for visualization (Nobre et al.2008)

5.9 Collection of data from databases

The data relating to low-risk and high-risk types of HPV were collected from previous literature. The types differed from several types of cancers like Oral cancer, Vulvar and Vaginal cancer, anal cancer, cervical cancer, and Head and neck cancer. The types HPV-6, 11, 16, 18, 26, 31, 33,35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 61, 66, 68, 70, 72, 73, 81, 82, CP6108 are identified to be related to cervical cancer **(Table 10)**.

5.9.1 Low and HR HPV types

The table consists of HPV low and high-risk types and their presence in oral, vulvar, vaginal, anal and cervical cancers.

Cancer Types	Low Risk -HPV	High - Risk -HPV	References
Oral	HPV-6, 11, 42, 43, 44	HPV-16, 18, 31, 33, 35, 45, 51, 52, 56, 58 and 59	Feller et al.2009
Vulvar and Vaginal	HPV 6, 11, 40, 42, 43, 44, 82, 83, 84, Iso39, 71, CP6108, 81, 26, 34, 53, 54, 55, 57, 61, 70, 72, 73.	HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68	Molano, M., et al.2002
Anal	6, 11, 32, 34, 40, 42, 43, 44, 53, 54, 55, 61,	HPV 16, 18, 31, 33, 35, 39, 45, 51, 52,	Cornall et al.2013 & Hoots, et al.2003

	70, 72, 73, 81, 83, 84, 89 and Pap155	56, 58, 59, 66 and 68.	
Cervical	6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81, and CP6108	16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73 and 82.	Muñoz et al.2003 & Tampa et al.2013

Table 10: Low and High-Risk Types of HPV

5.9.2 Retrieval of HPV sequences related to cervical cancer of E6 and E7

The sequences for HPV linked to cervical cancer of E6 and E7 were retrieved from PAVE database (https://pave.niaid.nih.gov/). (attached as Supplementary Data)

5.9.3 PAVE Database

The PapillomaVirus Episteme (PaVE) offers meticulously organized and regulated data and resources on papillomavirus genetics, specifically focusing on the Papillomaviridae viruses (Van Doorslaer et al.2012). The PaVE consists of a set of databases and web-based apps to store, assess, and transfer information. The PaVE utilizes an open source software methodology and prioritizes the incorporation and repurpose of preexisting tools **(Figure 21).**
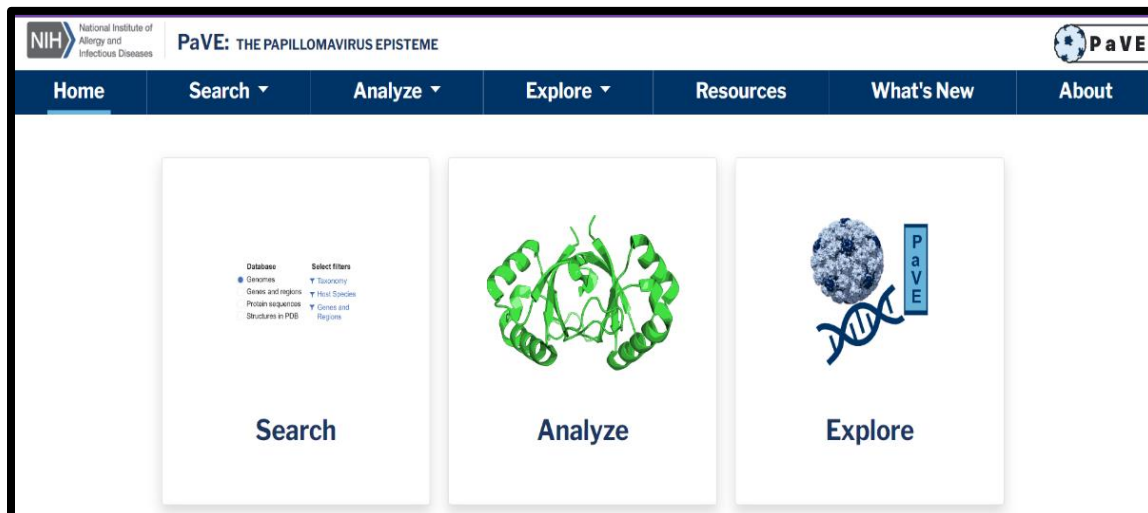
Figure 21: High and low risk HPV types from PAVE database

Figure 22: Retrieval of complete genome of HPV LR and HR types

5.10 System Properties

The system properties for the present in silico protein interaction work have been as follows.

Intel® Core™ i5CPU 650@3.20GHz with 8GB RAM and 64-bit Operating system.

5.10.1 Phylogenetic analysis
Principle

Within a phylogenetic tree, the relatedness between two species carries a highly precise significance. The degree of relatedness of two species can be assessed by the recurrence of their common ancestor (Knowles 2009). Species with a more topical common ancestor are considered to be more thoroughly related, while those with a less recent common ancestor are considered to be less closely related.

5.10.2 MEGA X software

MEGA X 64 bit installed and downloaded in the computer to further process the multiple sequence alignment and phylogenetic analysis (https://www.megasoftware.net/downloads/dload_win_gui).
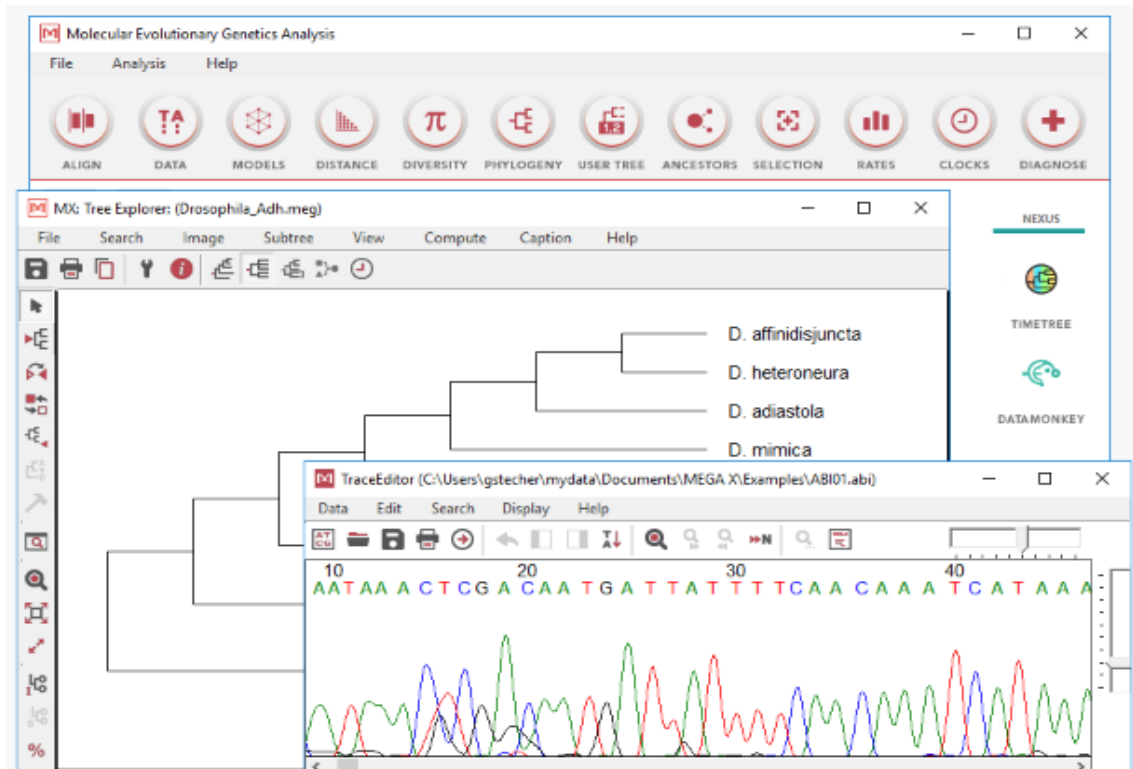
Figure 23: MEGA X software download and the analysis can be performed by this platform

5.10.3 Alignment of the sequences using ClustalSW or inbuilt Mega X software

Steps to be followed:

1. In the provided dialog box, input your set of sequences. Each sequence should be organized with the '>' sign followed by the sequence name (FASTA style), followed by enter, and then the sequence itself (Golding et al 2000).

2. The DNA/protein sequences are also able to be submitted by file upload by selecting the "choose file" option. It is important to ensure that all the sequences are in the same format.

3.The user has the flexibility to independently choose the parameters for pair-wise alignment alternatives and multiple sequence alignment choices, including selecting the scoring grids and scoring levels. Typically, the settings are set to their default values.

4. Users can receive email notifications by using the email notification option

Figure 24: Pairwise Alignment /multiple sequence Alignment complete steps

5.10.4 Construction of Phylogenetic tree

Construction of a phylogenetic tree based on E6 & E7 Proteins of High Risk and Low-Risk HPV using MEGA X software. Pairwise alignment of sequences used for the construction of a phylogenetic tree, it is used to see the similarity between the structural, evolutionary, and functional relationship among the sequences.

5.11 Gene/Protein-Protein Interaction studies (PPI)
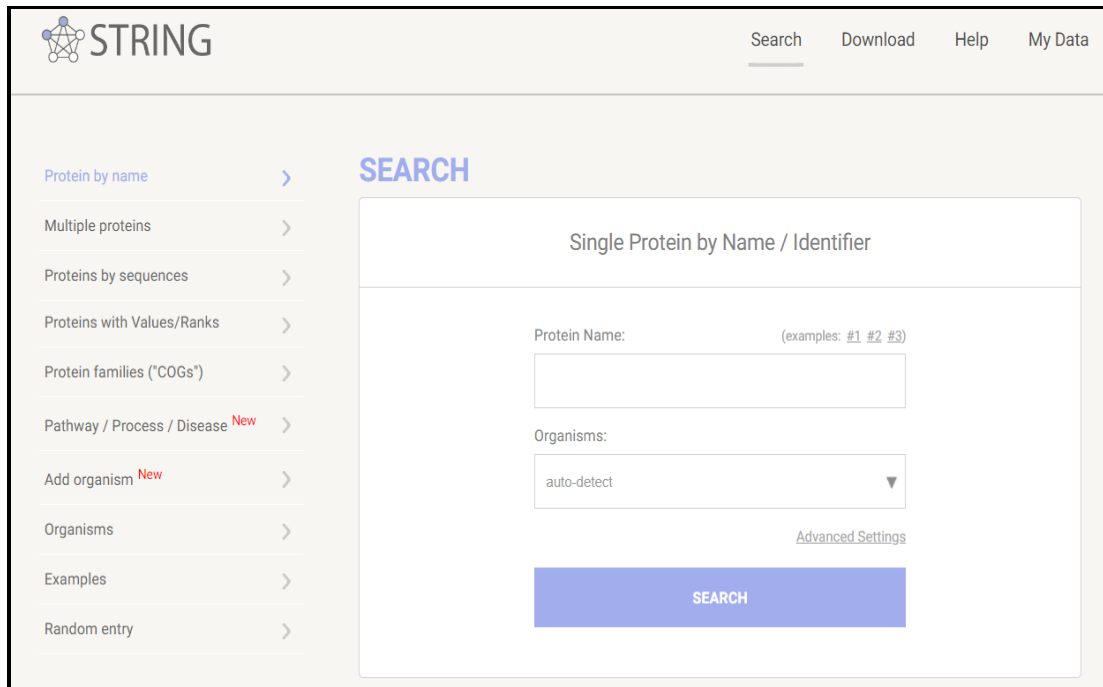
To anticipate molecular and cellular activity, gene and/or protein networks are employed as a cardinal representation of various biological processes (Raman, Karthik 2010).

5.11.1 String – functional interaction network of proteins

Structural and functional protein interactions are prudently collected and integrated into the String DB and the website is (https://string-db.org/). The information comes from various places including databases of interaction studies, conserved genomic context databases, computational interaction models from co-expression, and computerized text mining of the scientific literature.

Procedure

Input Single gene or protein/ multiple genes or protein vs Homologous organism and find the search results and the interaction pathways.

Figure 25 (A): String Software to provide the genes input

5.11.2 Cytoscape

Molecular interaction networks may be visualized with Cytoscape, an open-source bioinformatics application that works with gene expression profiles and other types of state information. Plugins allow for the addition of new capabilities. Network and molecular profiling analyses, fresh layouts, expanded file format support, database integration, and wide-ranging network searches are all possible via accessible plugins (Szklarczyk et al., 2023).
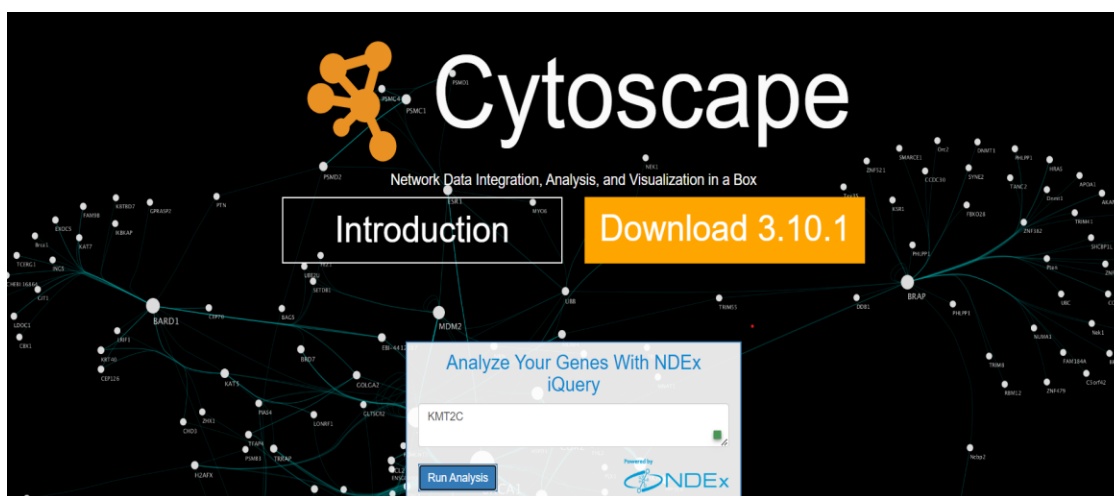
Figure 25 (B): Cytoscape to download and can directly use by NDEx and Visualize in cytoscape

5.11.3 Procedure

The Cytoscape software is downloaded to the computer using (https://cytoscape.org/download.html) and online for and genes/proteins TSV files of String data are uploaded into software and it gives results of interactions. Through Cytoscape visualization of genes and interactome is very clear.

5.12 Docking Studies of HPV Molecules (Gardasil 9)

5.12.1 Collection of data

PDB files of all HPV Molecules are retrieved from protein Data Bank (PDB) (https://www.rcsb.org/).
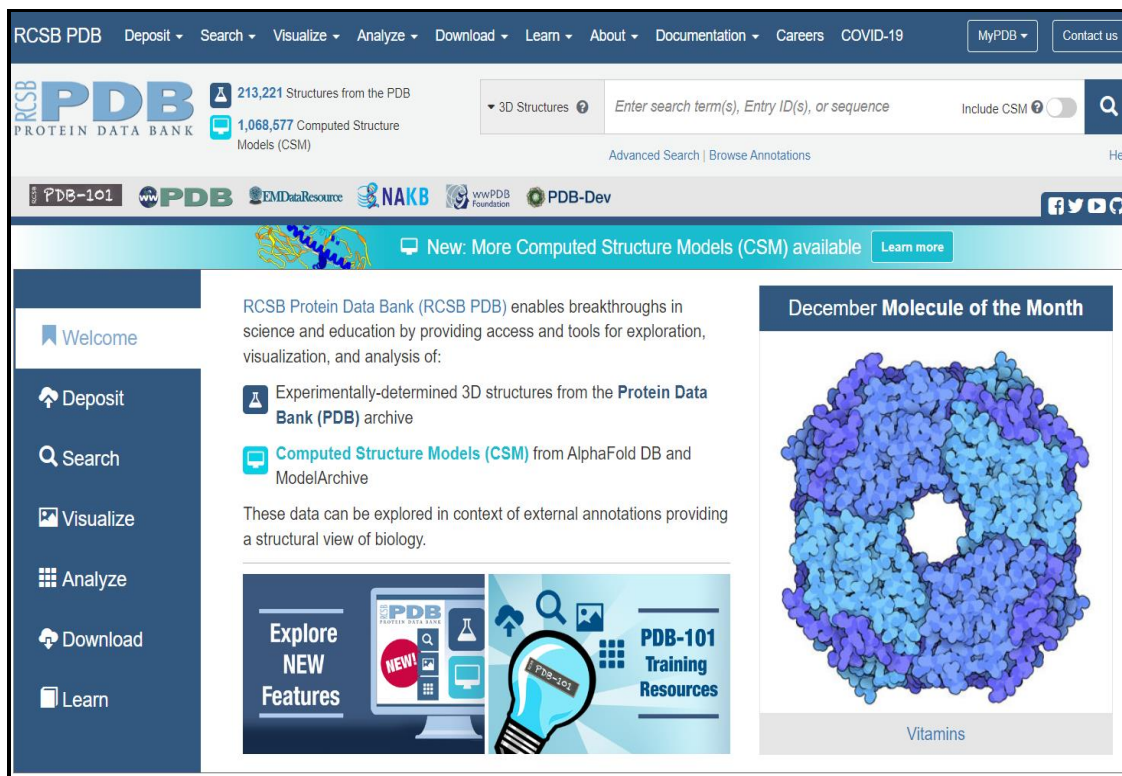
Figure 26: PDB files of ligands and receptors are retrieved from the protein data bank

5.12.2 Ligands

The protein sequences of HPV 18 L1 capsid protein, HPV 31 L1 capsid protein, and HPV 45 L1 capsid protein are retrieved from the PAVE database. The sequences are submitted to the Swiss model, and the best structures are selected as ligands. The Ligands of HPV6 (6L31), HPV16 (1DZL), HPV 11 (2R5K), HPV 18 (2R5I), HPV 33 (6IGE), HPV52 (6IGF) and HPV 58 (5Y9E) are retrieved from Protein Data Bank (PDB). In addition, the docked structure for HPV6+HPV11+HPV16+HPV18 using Hex and ClusPro software is also selected as Ligand.
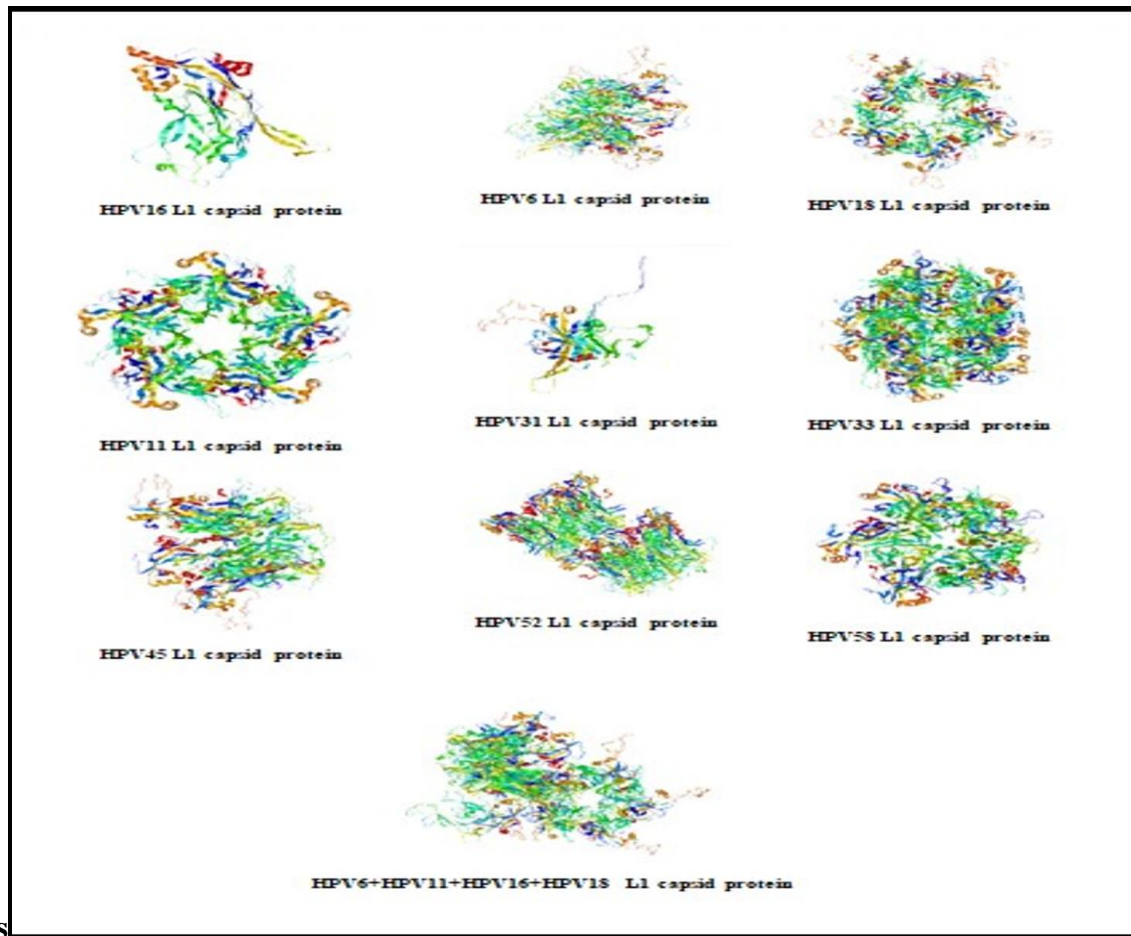
Figure 27: The PDB ligands from Protein Data Bank for docking studies

5.12.3 Receptors

The receptors Interferon 1, Laminin 5, Estrogen, Immunoglobulins, Endothelin A, FcγRIII, ATM, Heparin Sulfate, VEGF, EGFR, GRP78, and Integrin α6β1 are retrieved from the protein database
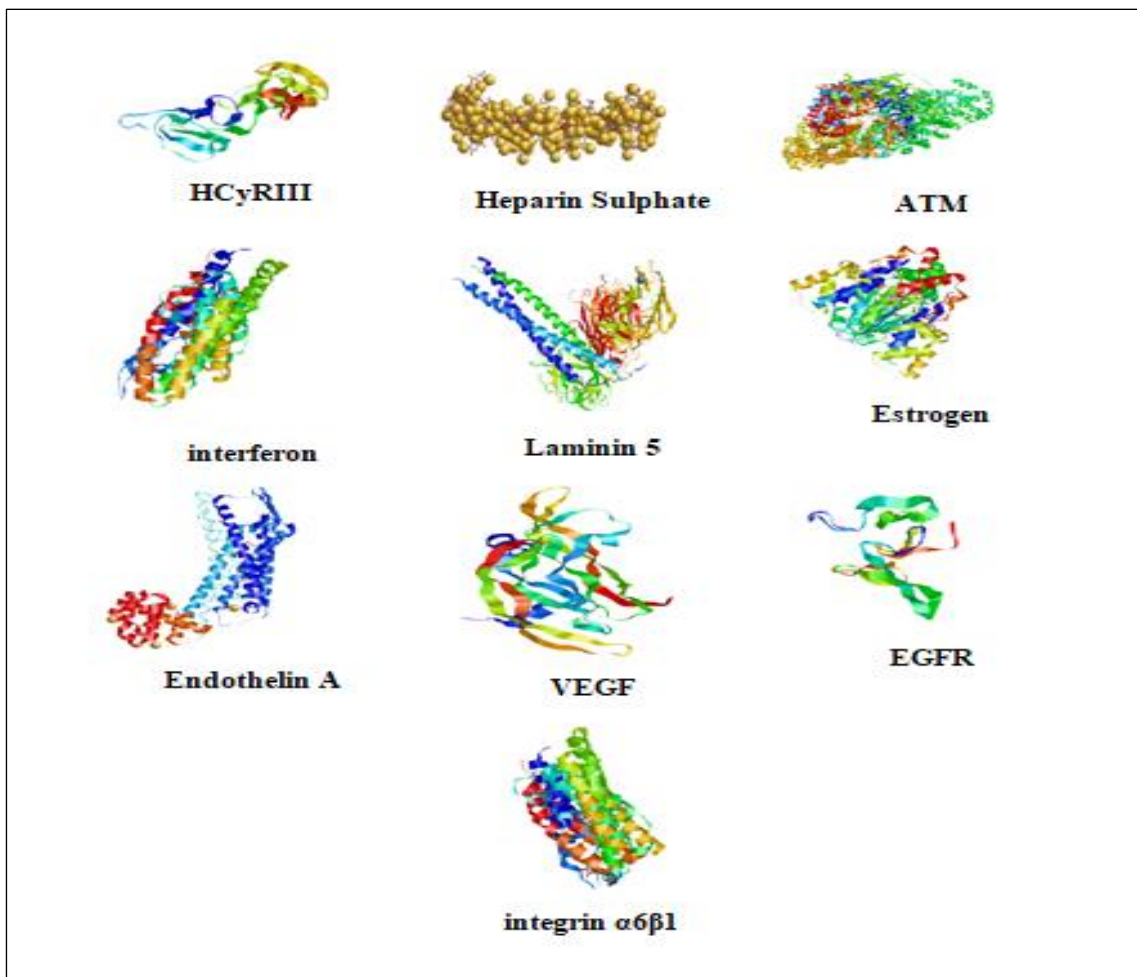
Figure 28: Receptors retrieved for docking studies using Hex 8.0 and ClusPro 2.0

5.12.4 Protein-Protein Docking using ClusPro 2.0 and Hex 8.0:

In computational and experimental biology, there is a consistent need for computational procedures for checking modeling and interactions between proteins. PPI is important for understanding cellular functions and organizations (Hu, Lun, et al.2021). The ClusPro server is centered on Algorithm PIPER, which accomplishes the sampling. It represents the contact energy between two proteins using an expression of the given form.

5.13 Protein–Protein Interaction Using String, Genemania, and Cytoscape

Flowchart of the work
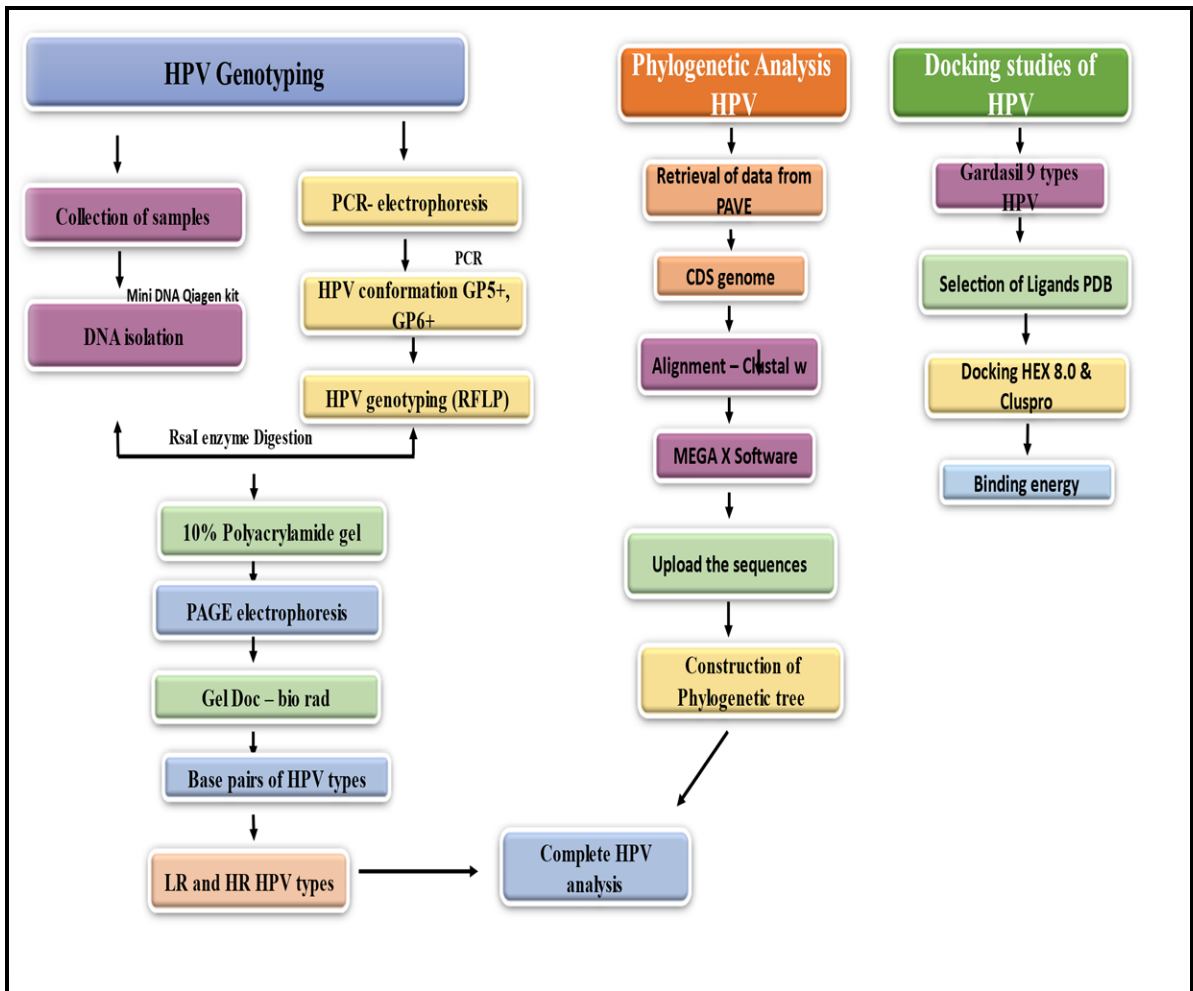
Figure 29: Overall representation of HPV Genotyping, Phylogenetic analysis, Docking studies.

5.14 To evaluate the genetic variants through whole exome sequencing (WES)

5.14.1 Whole Exome Sequencing (WES) samples

The samples were collected from MNJ Cancer Hospital, Hyderabad, India. All ethics clearances were accorded as per the ethics committee of MNJ Cancer Hospital Hyderabad, and informed permission was judiciously taken. Five normal and five malignant CC specimens with formalin-fixed samples with stages TNM II and III were

considered for the study. For DNA WES, formalin-fixed fresh cervical tumor samples were taken and sent to DNA Xperts, Ghaziabad, New Delhi. The CC samples are labelled for identification, as shown in Table 11. briefly, the management of five CC samples and five surrounding tissue samples of moderately differentiated squamous cell carcinoma and adenocarcinoma having TNM stages I, II, and III were included after surgical hysterectomy procedures. All the samples were subjected to DNA WES on the Illumina platform.

Sample ID	Specimen	Diagnosis/ Condition
T1- 2151/2020 S1- 2151/2020	Uterus-Cervix, Bilateral adnexa	Squamous Cell carcinoma TNM stage IIB
T2-2094/2020 S2-2094/2020	Uterus-Cervix with bilateral ovary	Adenocarcinoma, TNM stage IIIC
T3-2126/2020 S3-2196/2020	Uterus -Cervix	Squamous Cell carcinoma, TNM stage I
T4-2130/2020 T4-2130/2020	Uterus-Cervix with bilateral adnexa	Squamous cell carcinoma, TNM stage-II
T5- 2125/2020 T5-2125/2020	Uterus-Cervix	Adenocarcinoma, TNM stage – II

Table 11: Whole exome sequencing of tumor and control samples and the stage of the patient samples

5.14.2 Sample Preparation and DNA Isolation

Cervical tumor formalin-dipped DNA samples are washed and cleaned with PBS (Phosphate Buffer Saline), and later isolation is performed using a Qiagen FFPE DNA extraction kit. Then (0.8%) agarose gel is prepared to run the genomic DNA at 110 V for 40 minutes **(Figure 30)**. To determine concentration and quantification, each sample is estimated using Qubit Fluorometer. Further, library preparation was performed using the SureselectXT target enrichment system (TES). Moreover, 200ng of each sample was used for fragmentation. To analyze FFPE-derived DNA samples with extensively degraded

DNA, utilize the amount of amplifiable DNA obtained by qPCR and employ the maximum quantity of DNA within the range of 100–200 ng. Following the fragmentation of DNA, it undergoes end repair.
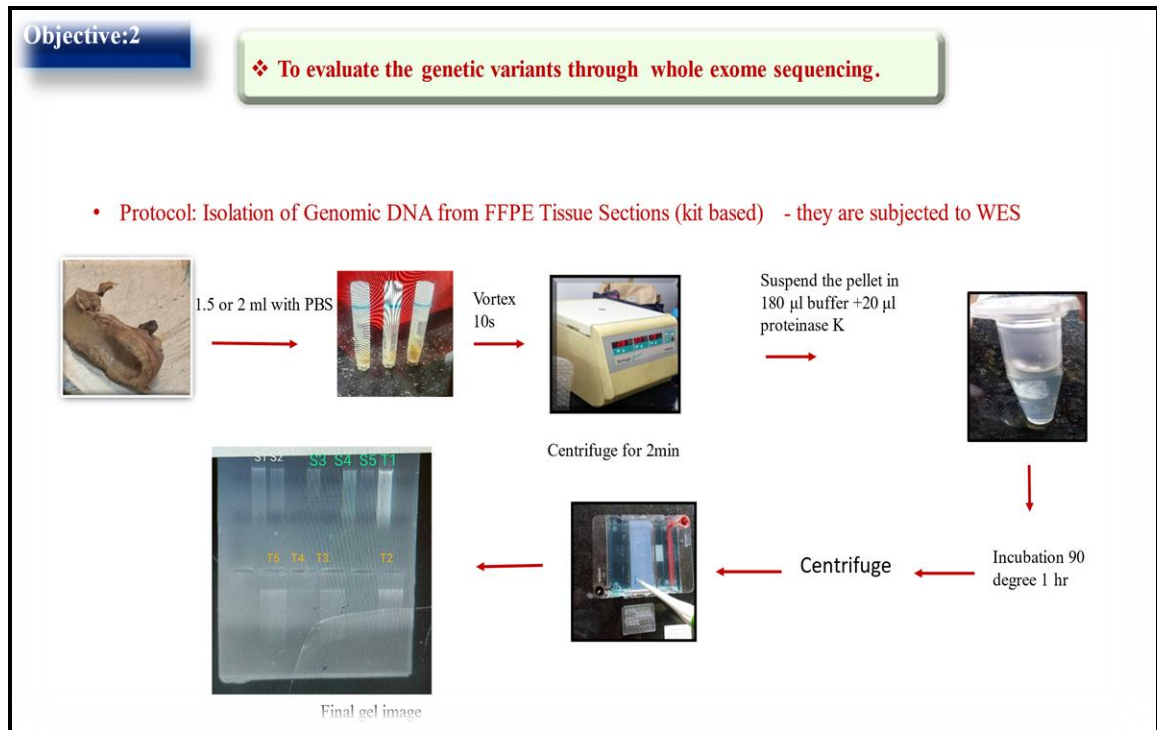
Figure 30: Sample Preparation and Preprocessing steps for whole exome sequencing

5.15 Library Preparation

5.15.1 Exon Capture

WES was performed on the ten samples using the Illumina platform. The reads were captured with an average target coverage depth of 100X per sample. This was immediately followed by in-solution extraction of genomic DNA using the Sureselect Human All exon V5 UTR kit, targeting 40Mb of exons. The DNA obtained from every sample is subjected to multiplexed paired-end sequencing, where two distinct sets of samples are used on the Illumina platform. (Pal and Aruna Pal. 2022).

5.16 Linux Terminal

After performing whole exome sequencing on Illumina platform, the raw reads are obtained and further we have analyzed the raw reads data on the Linux terminal. I have installed the LINUX and processed further by providing command lines. There are three steps in the processing and for further analysis:

1. NGS Pre-processing
2. Variant Discovery
3. Variant Prioritization
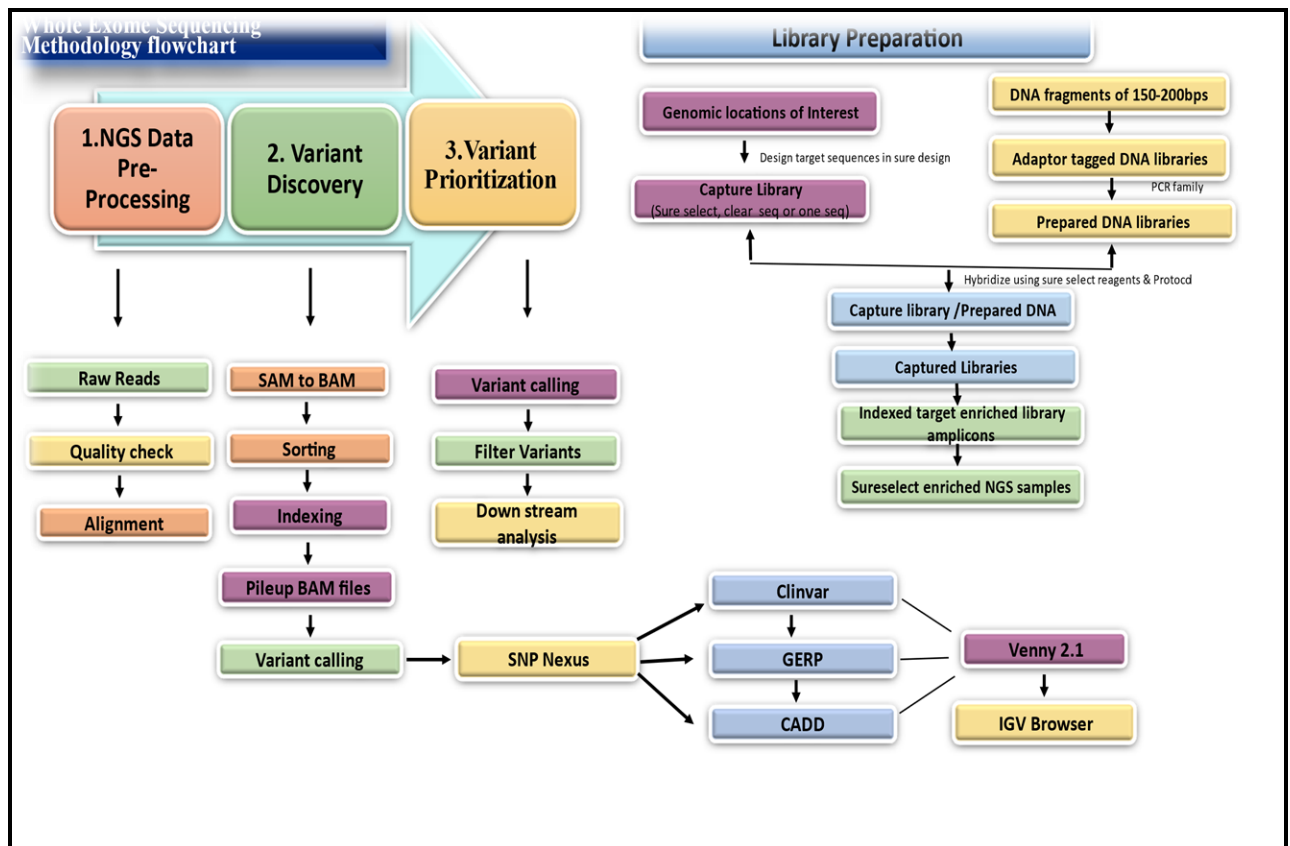
Figure 31: Flowchart of whole exome sequencing steps includes 1. NGS preprocessing 2. Variant Discovery 3. Variant Prioritizatio

5.17 NGS Pre-Processing

5.17.1 Quality check

To check the quality and GC content of all the samples, raw readings are run via our in-house benchmarked pipeline. FastQC is the command line with a sample name that should be entered. Example T1_1. fastq where T for tumor tissue samples and S1_1. fastq whereas S stands for surrounding tissue or control sample. All the files were downloaded in html format using proper command line (**Table 12**) (**Figures 32, 33**).

Sample ID	FastQC	GC content	Specimen	Diagnosis/ Condition
T1-2151/2020 S1-2151/2020	T1_1. fastq T1_2. fastq S1_1. fastq S1_2. fastq	47 47 47 47	Uterus-Cervix, Bilateral adnexa	Squamous Cell carcinoma TNM stage IIB
T2-2094/2020 S2-2094/2020	T2_1. fastq T2_2.fastq S2_1.fastq S2_2.fastq	48 47 47 46	Uterus-Cervix with bilateral ovary	Adenocarcinoma, TNM stage IIIC
T3-2126/2020 S3-2196/2020	T3_1. fastq T3_2. fastq S3_1. fastq S3_2. fastq	47 47 47 47	Uterus -Cervix	Squamous Cell carcinoma, TNM stage I
T4-2130/2020 T4-2130/2020	T4_1. fastq T4_2. fastq S4_1. fastq S4_2. fastq	47 47 47 47	Uterus-Cervix with bilateral adnexa	Squamous cell carcinoma, TNM stage-II
T5-2125/2020 T5-2125/2020	T5_1. fastq T5_2. fastq S5_1. fastq S5_2. fastq	48 48 44 44	Uterus-Cervix	Adenocarcinoma, TNM stage – II

Table 12: FastQ of the samples and their GC content and grade of the cancer

Figures 32 and 33: Quality check of all the raw reads files run by command line Fastq and the files are downloaded in html format.

5.17.2 Alignment

The alignment procedure is done by the bowtie2 command line, the sequences are aligned to further process for next steps. The command line example bowtie2

5.17.3 Variant Discovery

In Variant discovery the reads or files are converted first from secondary alignment memory (SAM) to BAM files i.e. Binary alignment memory so that the computer can read and encode the command line. Then sorting is performed by giving the command line of the sorted bam file, the conversion of sorted bam file is then indexing, pile up of files to mpileup is performed.

5.17.4 Variant calling

The samples were processed at three levels: pre-processing, the discovery of variants, and

90

prioritization of variants. A crosscheck for contamination was undertaken to see the heterogeneity and filter the variants. To detect heterozygous variations, all sequential processes such as sorting indels are conducted using the variant calling tool Varscan. The heterozygous variations are enumerated with awk/bash command one-liners, and the average depth of >5 is taken into account (**Figure 34**).

Figure 34: Variant discovery steps including sorted bam files and also mpileup and indels

5.17.5. Variant Prioritization

Variant Prioritization includes variant calling by varscan command and filtering of variants was performed, the filtered variants are converted to VCF files. The VCF files of CC tissue samples and surrounding or control tissue are downloaded. While running the pipeline for variant identification, we have taken the depth of the variants that mainly fall between ≥DP 5 and ≤20 (**Figure 35).**
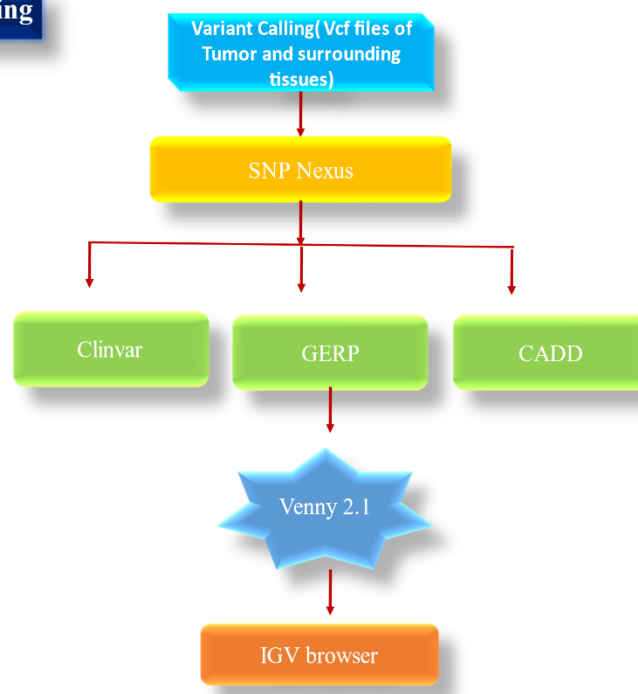
Figure 35: The flowchart depicts that after variant calling and Prioritization the following steps are included in the process.

5.18 Downstream analysis
5.18.1 SNP Nexus

The filtered variants and samples' VCF files go through downstream analysis using NCBI and various bioinformatics tools. The VCF files are uploaded in SNP nexus first and downloaded all the results of samples. https://www.snp-nexus.org/. we get an Excel sheet with several variants and these variants with CADD scores, GERP scores, 1000 genomes, pathway analysis, reactome etc.

5.18.2 Clinvar – NCBI

Clinvar is used to check the pathogenicity and non-pathogenicity of the variant, or is it a Benign

5.18.3 CADD and GERP

The cut-off value for CADD is >10 and the GERP value is >2, the filtered variants after these cut values are considered for further analysis using Venny 2.1

5.18.4 Venny 2.1

Venny 2.1 is used to draw venn diagrams for multiple samples according to their need. The data is uploaded for tumor samples and surrounding tissue samples final common variants are filtered out and considered for our study.

5.18.5 dbSNP -NCBI

After the variants are filtered the rs IDS of the variants, allele variation, and where the variant is missense, coding or non-coding, synonymous or nonsynonymous annotation done through dnSNP. The Minor allele frequency of the variant should be <0.05 value and GnomAD value considered for this study (**Figure 36).**
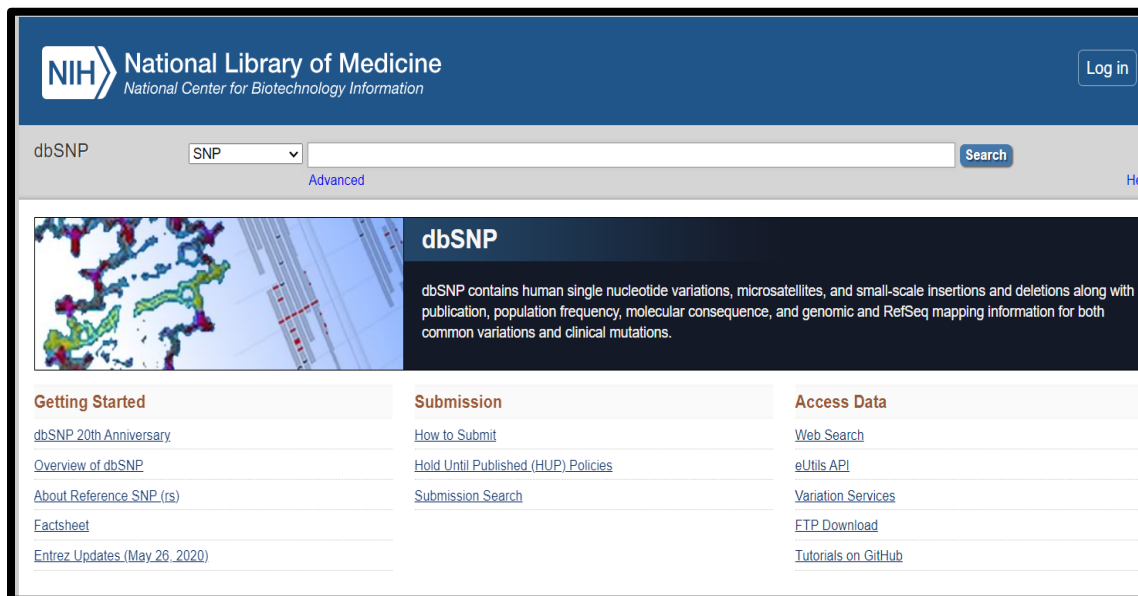
Figure 36: NCBI dbSNP used to view the SNP chromosome number and rs IDs

5.19 Prioritize variants, validate specific markers (SSCP, Sanger sequencing, Methylation and RNA-Seq analysis), and establish their associations with disease causation per se.

5.19.1 Integrative Genomic Viewer (IGV)

IGV tool representing Sashimi plots were used for quantitative alignment of RNA-seq reads for visualization and also to compare the exon coverage across the sample (Garrido, et al.2018 & El Husseini et al. 2019). First, install and download the IGV in the computer and then upload the human genome hg 38 file and also upload the sorted BAM file to view the result. It is used for visualization and validation of variants (**Figure 37).**

Figure 37: IGV installation to upload the file hg38 and tissue samples of sorted bam file

5.20 Polymerase Chain Reaction Single Nucleotide Conformation Polymorphism (PCR SSCP)

Through SSCP mutations are visualized in the form of bands, with polyacrylamide gel. Further here for PCR SSCP other than WES samples of the study, laboratory samples of 64 in which 15 control and remaining tumor tissue samples of grade 2,3, and 4 were taken for the study to correlate the WES samples data of variants with another cohort of CC. DNA isolated and further according to the protocol the steps were followed

5.20.1 Materials and Reagents
DNA template, PCR primers specific to the target region and PCR master mix <u>Denaturing solution</u> (formamide-containing buffer)

- Formamide: 95%
- Bromophenol blue: 10 mM EDTA

<u>14% SSCP gel</u> (non-Denaturing Polyacrylamide Gel)

- 49:1 Acrylamide: 28%
- Glycerol: 0.25%
- 10xTBE: 10%
- dH2O: 61.25%
- APS: 0.4%
- TEMED: 0.09%

<u>Gel running buffer</u> (1X TBE)

1X TBE buffer:

- Tris base: 89 mM
- Boric acid: 89 mM
- EDTA: 2 mM

<u>Silver stain solution</u>

- Ethanol: 10%

- Acetic acid: 0.5%

- AgNO$_3$: 0.2%

Developer solution

- NaOH: 3%

- Formaldehyde: 0.5%

Fixative solution

- Ethanol: 10%

- Acetic acid: 0.5%

5.20.2 Procedure

1. PCR Amplification

a. Design and order specific primers for the target DNA region.

b. Set up a PCR reaction as follows:

- 2x PCR Master mix: 1x volume

- Forward primer: 20pM volume

- Reverse primer: 20pM volume

- DNA template: 100ng

- dH2O: make up to desired volume

c. Perform PCR with the following cycling conditions:

- Initial Denaturation: 95°C for 5 min

- 35 Cycles of: 95°C for 30 sec, 60°C for 0.15 sec, 72°C for 30 sec

- Final Extension: 72°C for 5 min

- Hold at 4°C

d. Verify the success of the PCR reaction by running a small aliquot of the reaction on an agarose gel to check for the presence of the target amplicon.

2. <u>Preparing SSCP Gel</u>

a. Prepare a non-denaturing polyacrylamide gel appropriate for the target fragment size.

b. Pre-run the gel in 1x TBE buffer to equilibrate the temperature across the gel.

c. Make sure the gel tank is maintained at a constant temperature, typically 4-20°C.

3. <u>Sample Preparation</u>

a. Mix the PCR products with a denaturing solution with the ratio of 1:10 respectively.

b. Heat the samples at 95°C for 5 minutes to denature the DNA.

c. Immediately chill the samples on ice for 5-10 minutes.

4. <u>Loading and Running the SSCP Gel</u>

a. Load the denatured DNA samples onto the pre-run polyacrylamide gel along with the DNA ladder.

b. Run the gel O/N with 60-100 volts of electricity at an appropriate temperature (usually 4–20 °C) until the samples have properly separated.

5. <u>Visualization</u>

a. Stain the gel with a silver stain for 10 min and rinse with dH_2O once.
b. Add a developer solution to the gel and shake it back and forth until the single-stranded DNA bands are clearly visible.
c. Then discard the developer and pour fixative solution

6. <u>Analysis</u>

a. Compare the SSCP patterns of the samples to identify variations or mutations in the target DNA region.

b. Variations will result in shifts or changes in the band patterns.

7. <u>Confirmation</u>

If any intriguing band patterns are found, sequence the DNA from those samples to verify the alterations.

5.21 Sanger Sequencing of *KMT2C and C1QTNF 9*

The DNA was extracted from (FFPE) tissue using a Qaigen kit. Genotyping tests were carried out utilising DNA samples with high molecular weight on agarose gel and the nanodrop A260/280 ratios ranging from 1.8 to 2.0. The flanking regions of the variants KMT2C and C1QTNF 9 were amplified. (rs62478356 and rs138908625) followed by direct sequencing. A PCR reaction was carried out with 100 ng of DNA, using 10 pmol/µl (Forward primer- 5′ACACAAACCTTTAGCACCTAGT3′, Reverse primer 5′GGTTCATGATGCAAATCCTTTCA3′ (rs62478356, rs138908625) of KMT2C and Forward primer- 5′ TGCACACCAAAGATGCTTACATGA 3′, and reverse primer 5′ CCACAAGAAACGCACAGACGAG 3') of C1QTNF 9 and 1× EmeraldAmp® GT and PCR-Master Mix in a vol. of 25 µl. The samples underwent incubation in a thermal cycler, the following steps and starting with an:

Initial Denaturation: 94^0C for 5 min

No. of Cycles: 35

Temperature: 98^0C for 30 Sec

Temperature: 60^0C for 30sec

Temperature: 72^0C for 60 sec

Final extension step: 72^0C for 5 min

Holding the temperature: 4^0C

Prior to purification and sequencing using an ABI sequencer (Life Technologies, USA), the product size of all amplicons was verified on an agarose gel, with sizes of 345bp and 349bp. The genotypes were analyzed using FinchTV software.

5.22 Transcriptome Samples for Data Analysis

10 women with CC and healthy controls participated in the study. Cervical punch biopsy tissue samples were obtained from patients in the MNJ Cancer Hospital's oncology unit (Hyderabad, India) and Shroff hospital, collected 10 normal and cervical tissue biopsies from women extending in age from 28 to 60 years (median age) who had undergone hysterectomy for non-malignant gynecological problems. For histopathological analysis, a small portion of the biopsy specimens were removed, formalin-fixed, and embedded in paraffin.

5.23 Transcriptome/RNA sequencing gene expression profiling

RNASeq is a novel transcriptome profiling method that makes use of deep sequencing technologies. Furthermore, compared to other methods, RNASeq provides far more accurate transcript levels and isoform measurements. As previously stated, the Clariom-D microarray was performed on 10 tissue samples of control and malignant samples. RNA was isolated, tagged, and subjected to hybridization using the (Affymetrix) following the instructions provided by the manufacturer (Yu, Min, et al.2012). Staining with phycoerthyrin fluorochrome is done, and the transcriptome array produced fluorescent signals that were saved as DAT files. The (AGCC) software was employed to transform scanned pictures of DAT files into insensitive CEL files. The unprocessed CEL and CHIP files were loaded into the (EC) and (TAC) software to produce a catalog of genes that exhibited differential expression.

5.24 To perform integrative analysis using system biology-based approaches

5.24.1 COSMIC and cBioportal Database

Cancer for somatic mutations to visualize the rs IDs of cancer samples and the link to the DB https://cancer.sanger.ac.uk/cosmic Through the Genome Browser COSMIC database the SNVs were annotated for 138908625, 62478356 respectively
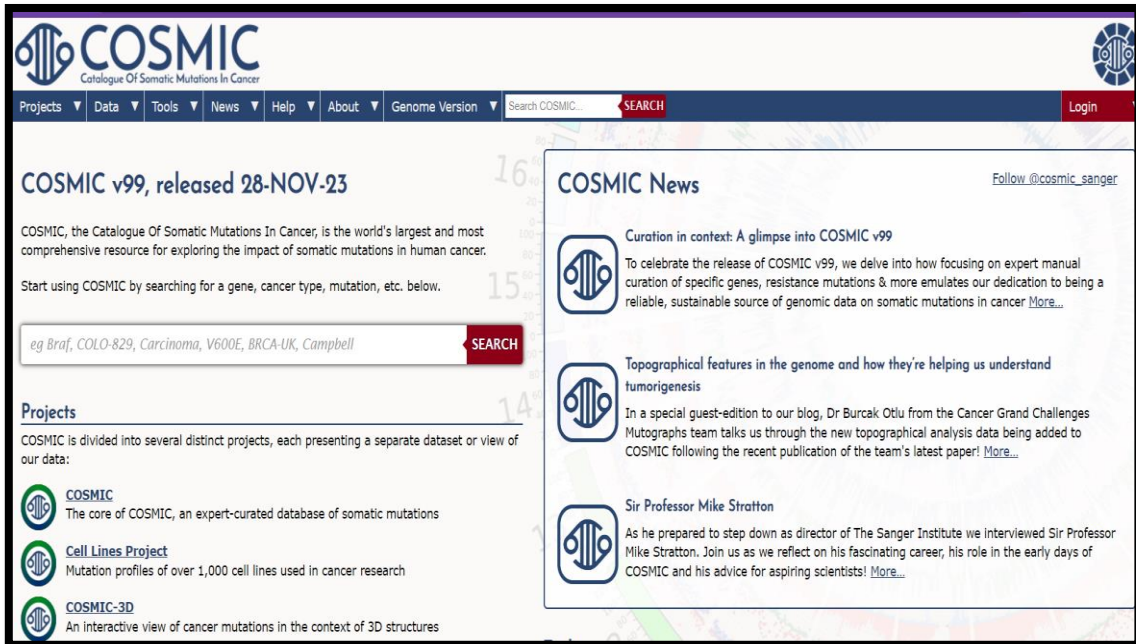
Figure 38: Cosmic Database to input rs IDs for annotation

5.24.2 cBioportal Database

The WES SNVs of CC were compared with data of PAN cancer studies of 278 samples of Cervical squamous cell carcinoma (CSCC) and normal cervix from the cBioportal database, KMT2C plot of overall survival against the probability of overall survival that denotes the altered and unaltered groups (https://www.cbioportal.org/).

5.25 Data mining of KMT2C gene

The bioinformatic analysis utilized six datasets sourced from (GEO) and the (TCGA) dataset (He, Simin, et al.2019 & Gao, Li, et al.2018). The CC microarray gene expression (GE) datasets were acquired from the GEO (https://www.ncbi.nlm.nih.gov/geo/) using the accession numbers GSE6791, GSE29570, GSE63514, GSE67522, and GSE138080. The

dataset comprises many samples, from which we specifically selected just the ones that tested positive for cervical cancer, in comparison to the control samples. The platforms used for GSE6791 were derived from GPL570 [HG-U133_Plus_2]. The dataset GSE29570 utilized the AHG U133 Plus 2.0 Array, which was designed based on the GPL6244 platform (HuGene-1_0-st). The GSE63514 dataset was generated using the AHG 1.0 ST Array, which is the transcript (gene) version. This array is based on the GPL570 platform, also known as HG-U133_Plus_2. The (AHG) U133 Plus 2.0 Array, GSE67522, utilized the GPL10558 Illumina HumanHT-12 V4.0 expression beadchip platform. On the other hand, GSE138080 was based on the GPL4133 Agilent-014850 Whole Human Genome Microarray 4x44K G4112F platform.

5.26 GEPIA Dataset

GEPIA is a cooperating tool for analysing TCGA and GTEx RNA sequencing expression data. The gene expression of KMT2C mRNA in cervical tumour and normal tissues was compared using GEPIA.

5.27 TIMER 2.0

TIMER2.0 is a robust program used for investigating immunological infiltrates (Li, Taiwen, et al.2017). The configuration and prevalence of immune cells in the cancer microenvironment exert a significant impact on cancer advancement and the effectiveness of immunotherapy. Due to the limitations of direct measuring methods, computer algorithms are commonly employed to deduce the makeup of immune cells from bulk tumor transcriptome profiles (Li, Taiwen, et al.2020). The study employed TIMER2.0 to examine the association between KMT2C expression and the levels of 6 immune infiltrates in cervical cancer: CD4+ T cells, CD8+ T cells, B cells, neutrophils, macrophages, and dendritic cells. The scatterplots display the partial Spearman's rho values that have been adjusted for purity.

5.28 Kaplan-Meier Plotter (KM)

The KM survival curve represents the likelihood of living during specific time intervals when time is divided into numerous short intervals. The KM estimate is a highly effective

method for determining the proportion of individuals who survive for a specific duration following treatment (Goel, Manish Kumar et al.2010). The KM Plotter, available at https://kmplot.com/analysis/, is a comprehensive dataset that utilizes a large sample size to forecast the survival of cancer patients. The predictive status of KMT2C expression in cervical cancer was assessed using the KM. $P<0.05$ was used to determine statistical significance. The hazard ratios (HRs) and their complementary 95% confidence intervals (CIs) were provided (Nagy et al.2018).

5.29 KEGG and STRING database analysis

KEGG (https://www.genome.jp/kegg/). The KEGG DB main aim is to give genes and genomes responsive connotations at both the molecular and greater levels (Kanehisa et al.2017). Pathway motfits, orthologous groups are most commonly encoded through this pathway that contains cellular connections (Kanehisa et al 2000). All known and anticipated protein interactions are included in the STRING database. The STRING database is online resources devoted to organism-wide protein interaction networks (Szklarczyk et al.2021& Mering, et al.2011). The associated pathways and network of protein-protein interactions for the KMT2C gene were discovered using these two databases. Furthermore, transcription factor binding site analysis was carried out by using Alibaba TRANSFAC software to identify the associated transcription factor for KMT2C gene exon variant C>T (SNP ID: rs 62478357).

5.30 Ualcan Database

Ualcan is a methylation-based database that can show both hypomethylated genes and hyper methylated genes. DNA methylation includes beta values for unmethylated and methylated genes. Across all the cancer was also checked.

5.31 Docking studies of KMT2C

The KMT2C gene encodes a histone methyltransferase that regulates gene transcription by modifying chromatin structure. KMT2C ligands and receptors Artemisinin from PDB bank. Docking was performed using Clus pro 2.0 with Artemisinin, Bucidarasin A, Betulin, Shikonin etc.

5.31.1 KMT2C PDB structure:

The structure of KMT2C/ MLL3 from the protein data bank, retrieved the image for docking studies KMT2C gene contains 60 exons and spans more than 216 kb. It is preceded in the 5-prime region by a 1.8-kb CpG island.

Figure 39: The crystal structure of MLL3 SET domain (5F59)

5.32 Interaction networks by GeneMania, String, and Cytoscape

The protein-protein interaction networks are constructed using a string database for single or multiple genes and if we download the TSV file of string and upload to the Cytoscape, the visualization of the network and nodes and internodes are visible.

5.33 I -TASSER

I-TASSER is a method for predicting protein structure and annotating protein function based on a hierarchical technique called Iterative Threading ASSEmbly Refinement. The initial step involves the identification of structural templates from the Protein Data Bank (PDB) using the LOMETS. Subsequently, full-length atomic models are generated by iterative template-based fragment assembly simulations. The function insights of the target are obtained by re-analyzing the 3D models using the protein function database BioLiP. I-TASSER achieved the top ranking among servers for predicting protein structure in a recent community-wide evaluation (Zhang. 2008).

We have first submitted the wild type (original sequence) and mutation sequences (two sequences with variations) of exon 8 region rs ID 13908625 protein sequences in I Tasser. Wild type sequence Arginine is mutated to Histidine in query 1 sequence and Arginine to Leucine in other mutated sequence (Query 2).

5.33.1 Original sequence of KMT2C exon 8 region (wild type)

The below sequence is been submitted to I Tasser

>protein

KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKRAGWQCPECKVC
QNC

5.33.2 Ligands and receptors

KMT2C acts as a ligand and receptors from the literature of cervical cancer chosen compounds such as Artemisinim, Shikonin, Bucidarasin A, Betulin.

5.34 Protein-Protein Docking

Cluspro 2.0 is used for docking as a protein-protein study. The compounds used for docking were Artrmisinin, Shikonin, Sitoindoside IX, Bucidarasin A, and Betulin taken from the literature. These compounds are used in the treatment of CC (Figure 40).
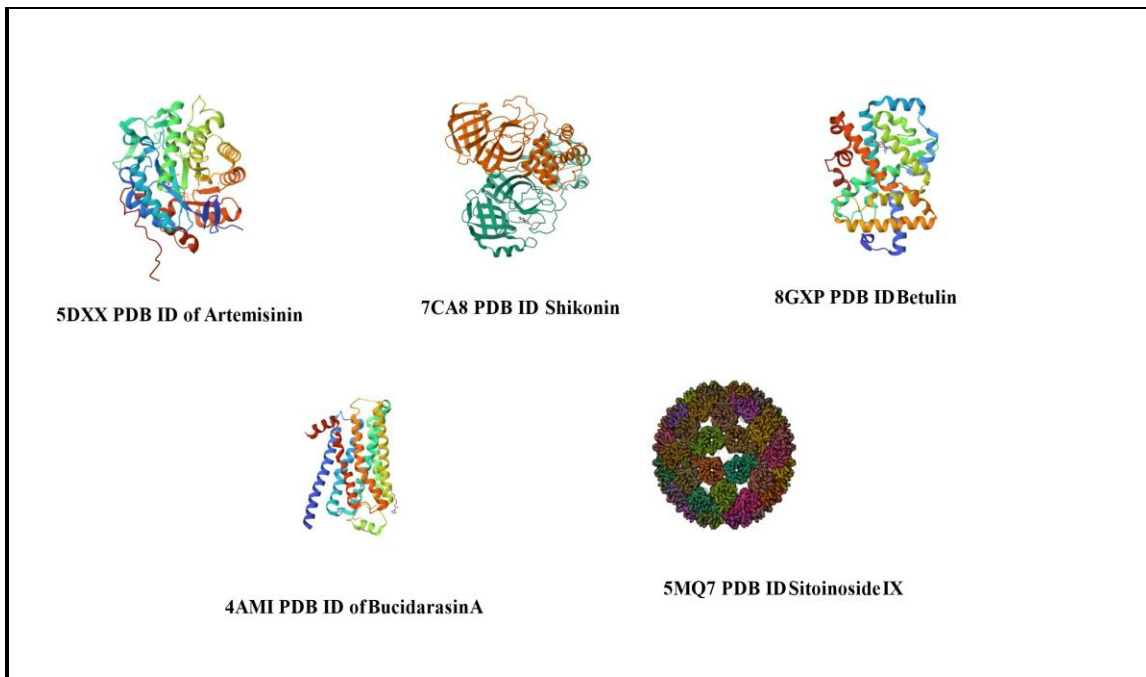
Figure 40: The PDB structure of compounds that to bind with KMT2C

Chapter 6

Results and Discussion

To catalogue the Human papillomavirus (HPV) through genotyping and phylogenetic analysis.

6.1 DNA Extracted from biopsy and tumor tissue samples

DNA isolated with a mini Qiagen kit and where 50bp ladder and HeLa as control samples were taken and these are the images of DNA bands on PCR gel image (**Figure 41**).
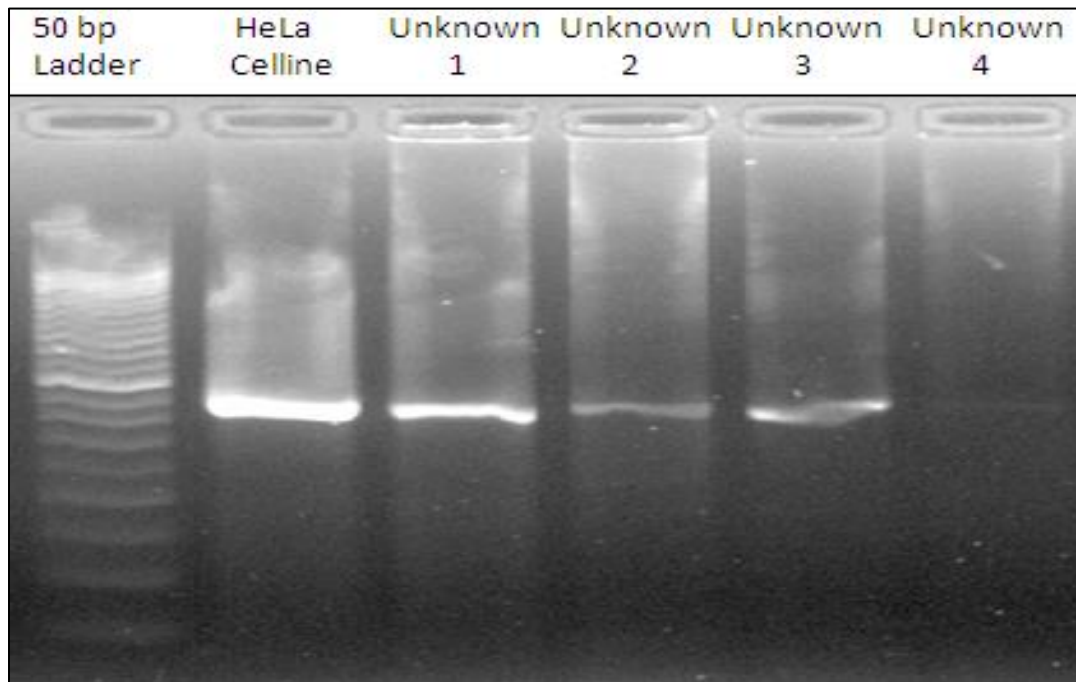
Figure 41: DNA PCR Gel image

6.2 HPV Genotyping from CC samples

HPV screening was performed through pap tests, liquid-based cytology, or detection to detect precancerous lesions, inflammation, or cancer. The detection can be more effective than cytology in knowing the type of HPV responsible for cancer. HPV Genotyping using the RFLP method reveals few HR & LR HPV Bands. Instead, they have shown a particular base pair (bp) for a specific variant, and these were categorized based on the length of the bps, Cervical cancer patients of these samples however, **(Figures 42,43)** showed HPV 6

having base pairs of 67bps, 72bps, 149bps, 159bps, HPV 16 having 70bps, 72bps and 310bps, HPV18 showing 38bps, 72bps, 85bps, 125bps, and 135bps and HPV 33 with 39bps, 72bps, 102bps, and 236bps. LR HPV was indicated as the reason for genital warts; however, even HR HPV was shown to cause both cancers and genital warts.
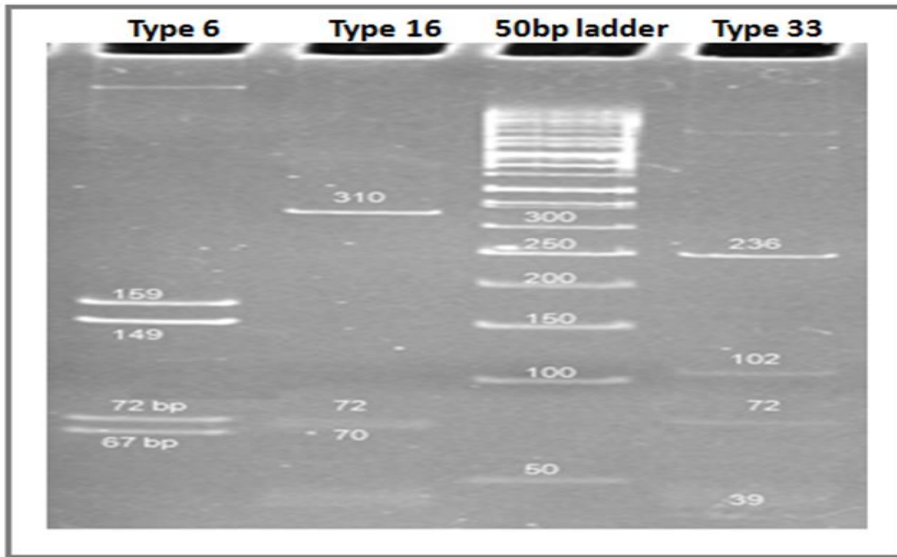
Figure 42: Gel image of HPV Genotyping of CC patients with base pairs as follows - HPV 6: shown 67bps, 72bps, 149bps, 159bps, HPV 16: shown 70bps, 72bps and 310bps, HPV 33: shown 39bps, 72bps, 102bps and 236bps
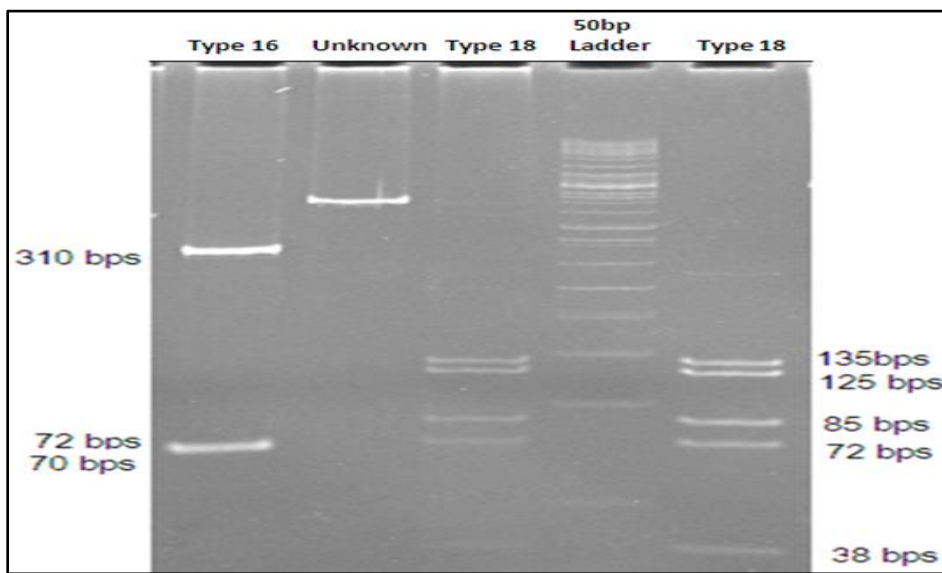
Figure 43: HPV Genotyping and detection as HPV 16 showing 70bps, 72bps, 310bps & HPV 18 showing 38bps, 72bps, 85bps, 125bps and 135bps.HPV genotyping of cervical cancer patients was predicted with MY09/MY11 primers, and HPV types 6, 16,18 and 33 were observed.

Anal cancer (in over ninety percent of cases), vaginal cancer (in between sixty-five percent of cases), penile cancer (in between forty percent of cases), vulvar cancer (in twenty percent to fifty percent of cases), oropharyngeal cancer (in forty-five percent of cases), oral cancer (in twenty-four percent of cases), and laryngeal cancer (in twenty-two percent of cases) have all been linked to HR-HPV infections (Sitarz et al.2022). It was proposed that the most prevalent screening approach for cervical cancer would be the HR-HPV genotyping test, with a 5-year screening interval. Additionally, complete HR-HPV genotyping has increased in popularity due to the useful epidemiological data it provides (Rosário et al.2023).

As CC is caused due to HPV, a Pap test for the cytological examination, HPV DNA screening is used to determine the type of HPV, i.e., LR or HR HPV, associated with an individual that causes cervical squamous cell carcinoma, LSIL, Cervical intraepithelial neoplasia, Adenocarcinoma, HPV genotyping with PCR- RFLP in house method results in **(Figure 42 and 43)** depicts HPV 6, 16, 18, 33, whereas HPV 16, 18 and 33 causes, Low grade squamous intraepithelial lesions, HSIL, and CC (Rosário et al.2023). Therefore, people with normal cytology and HPV genotypes 16 and/or 18 should have rapid colposcopy rather than delaying a year, and people with HSIL cytology and HPV type 16 should have accelerated therapy (rather than colposcopy). However, the consequences of incorporating HPV testing and genotyping into a program for detection & screening remain uncertain.

Multiple worldwide and national research has consistently shown that HPV 16 has the highest occurrence rate in women, both with normal and atypical cytology (Golfetto et al. 2018). The findings suggest that our internally conducted HPV genotyping testing, namely

using PCR-RFLP analysis, has the potential to be utilized as the main method for screening HPV, particularly in economically disadvantaged nations. Presumably, the restricted number of cases constrained our findings. However, in the case where numerous forms of HPV are detected in the main test, a more detailed test, such as the PapilloCheck®, hybrid capture, may be conducted. At present, there is no universally accepted approach for determining the type of HPV. The selection of a method for therapeutic purposes should be based on a careful evaluation of its strengths and weaknesses.

6.3 Venn diagram data

Figures 44 A, B, and C show that HPV-26, 53, 73, and 82 are at LR for Anal, Vulvar, and Vaginal cancer but will transform into high risk for cervical cancer and commonly HPV 6,11 for LR and 16, 18, 33 from our study results show HR HPV.
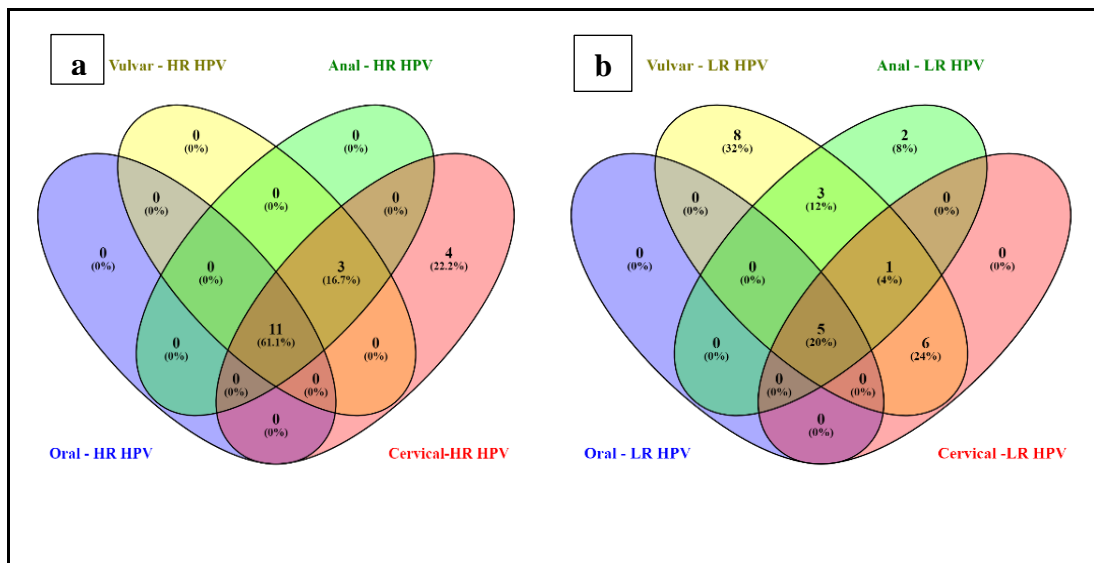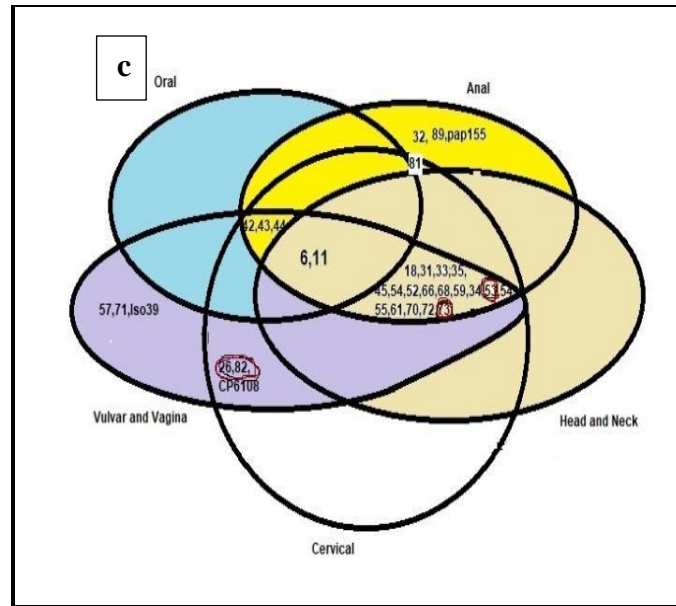
Figure 44a, 44b, 44c: Venn Diagram of High-risk HPV types and Low-risk HPV types causing Oral, Vaginal & Vulvar, Anal, Head & Neck and Cervical cancer

6.3.1 Phylogenetic Analysis

Phylogenetic analysis for the HPV types HPV-6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 61, 66, 68, 70, 72, 73, 81, 82 are shown in Figure 5D. The HPV types HPV-82, HPV-51, and HPV-26 are predicted in a single branch and are closely related based on phylogenetic studies. E6 and E7 oncogenic proteins have shown divergence with the relatedness of HR and LR-HPV convergences in the phylogenetic tree (**Figure 45**). The prediction shows that all these three types offer a high risk of HPV infection in the cause of cancer.
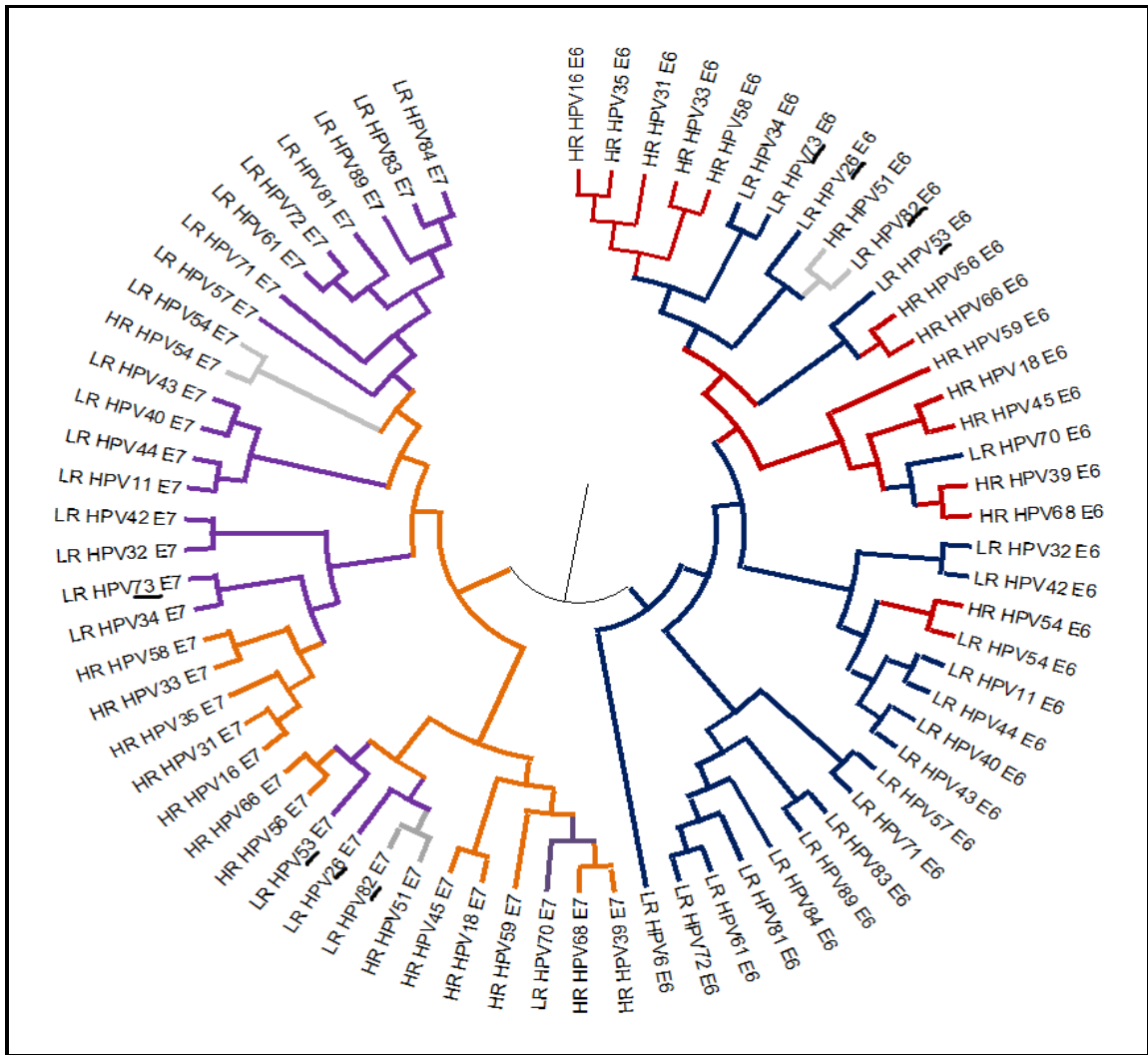
Figure 45: Construction of Phylogenetic Tree based on E6 & E7 Proteins of High Risk and Low-Risk HPV based on Maximum likelihood using MEGA X software

In our study, *Phylogenetic analysis* **Figure. 45** determines the evolutionary relationship of genes, genomes, or proteins of whole or partial biological communities developed from a selective habitat. Investigating the gene composition of HPV types from cervical cancer is one of the better methods for predicting the genes. Phylogenetic analysis was performed in HPV-6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 61, 66, 68, 70, 72, 73, 81 and 82 oncogenic protein sequences using MEGA 7.0 software. The PAVE database has fully sequenced about 150 genomes of HPVs, and more unknown HPV

types are to be further discovered that may be present between animals, humans, and their respective environments. The HPV types 16, 18, and 33 are predicted to be closely related based on phylogenetic studies.

6.4 Protein-protein Interactions (PPI)

6.4.1 PPI using String database

Protein-protein interaction studies aim to determine the nature of genes involved in HPV proliferation, Replication, and progression to cause cancer. These genes act as receptors and bind to the ligand. The input submitted for the string database is Interferon 1 (IFNAR1), Laminin 5 (TMEM5), Estrogen (ESR2), Endothelin A (ECE1), FcγRIII (no proteins by this name in Homo sapiens), ATM (ATM), Heparin Sulphate, VEGF (VEGFA), EGFR (EGFR), GRP78 (HSPA5) and Integrin α6β1 (no proteins by this name). The EGFR signalling pathway for HPV shows interactions with Interferon 1 (IFNAR1), Laminin 5 (TMEM5), Estrogen (ESR2), Endothelin A (ECE1), ATM (ATM), VEGF (VEGFA), EGFR (EGFR), and GRP78 (HSPA5) proteins (**Figure 4**). All these proteins of ligands and receptors are involved in the HPV signaling pathway that directs to proliferation, chronic inflammation, and survival. The number of nodes of PP interaction is 19, the edges are 63, the average node degree is 6.63, the average local clustering coefficient is 0.689, the overall expected number of edges is 39, and the PPI enrichment value is 0.000266.
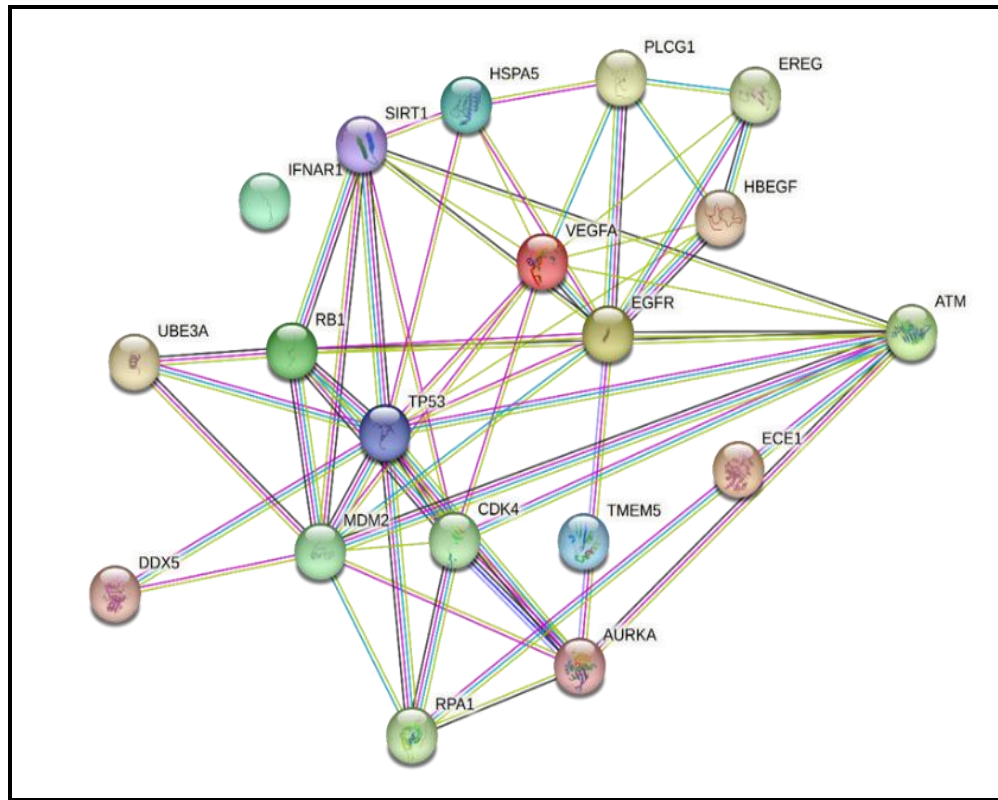
Figure 46: Protein-protein interaction between Genes Involved in the HPV pathway using String database.

6.4.1 PPI applying Cytoscape Interaction

Interactome formed by Cytoscape Genes retrieved from the STRING database, including EGFR and VEGFA, serve as membrane receptors of HPV to trigger multiple signaling cascades- ATM, which respond against DNA damage and response pathway. On the other hand, ESR2 acts as the direct target from VEGFA, which modulates EGFR, KDR, ATF6 & FLTe. In addition, ATF6 & FLT1 modulate angiogenic response. VEGF and EGFR further stimulate HSP90, which sensitizes in response to misfolded proteins (**Figure 47**). As a result, targeting E5 may be a sensible line of attack for cervical cancer chemoprevention and therapy, and knowing its oncogenic mechanisms may help us build innovative therapeutic techniques of HPV 16 E5 on cervical cancer development, specifically how it alters critical cellular signaling pathways associated with cell proliferation, angiogenesis, and apoptosis (Bhattacharjee et al.2022).
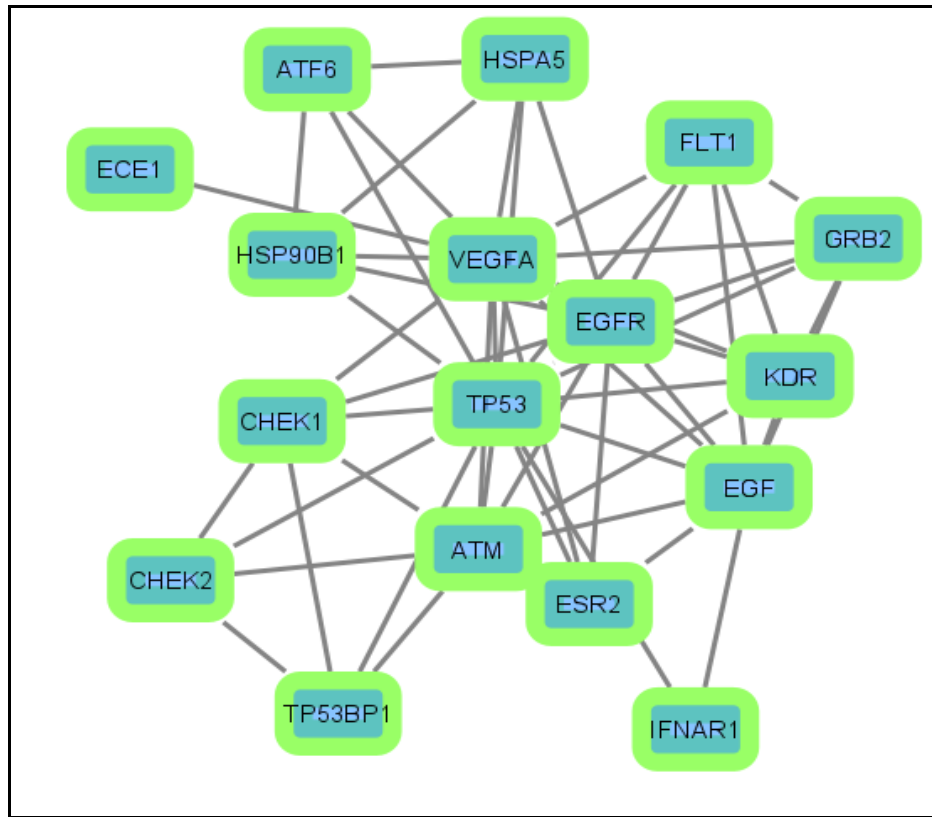
Figure 47: Cytoscape interactions from the Tsv file downloaded from string interactions for visualization

The PPI network of the above cervical cancer proteins such as P53, PRb, EGFR, ATM, VEGF, and ECE1, evaluates the proteins responsible for progression, and proliferation that cause carcinoma of different stages of grade I, II, III and metastasis through various pathways. The proteins or genes network shows biological molecular pathways that are responsible for different cancers that also cause cervical carcinogenesis.

6.4.2 GeneMania

The interactions of GeneMania show the given proteins interact, having Co-localisation and genetic interactions between the genes ESR1, ECE1, EGFR, IFNARI, ATM, PRKCA, VEGFA, and HSPA5. The Genetic interactions of all the genes show 37.35%, Physical interactions at 70.23%, Co-expression at 14.35% and Co-localization at 11.19%. ECE1 is involved in biologically active peptides, and the mutations cause cardiac defects and

autoimmune disorders. Estrogen receptor protein is associated with breast cancer and Estrogen resistance and EGFR, in which the receptor has seven other ligands. EGFR helps in cell growth and survival. The proteins like Laminin 5, Interferon, and Integrin are involved in the overexpression of cervical carcinoma and CIN cases **(Figure 48).**

The interactions from the String database, GeneMania, differ in their physical interactions with each protein or gene; however, each gene is responsible for inflammation, survival, tumor development stages, and angiogenesis.
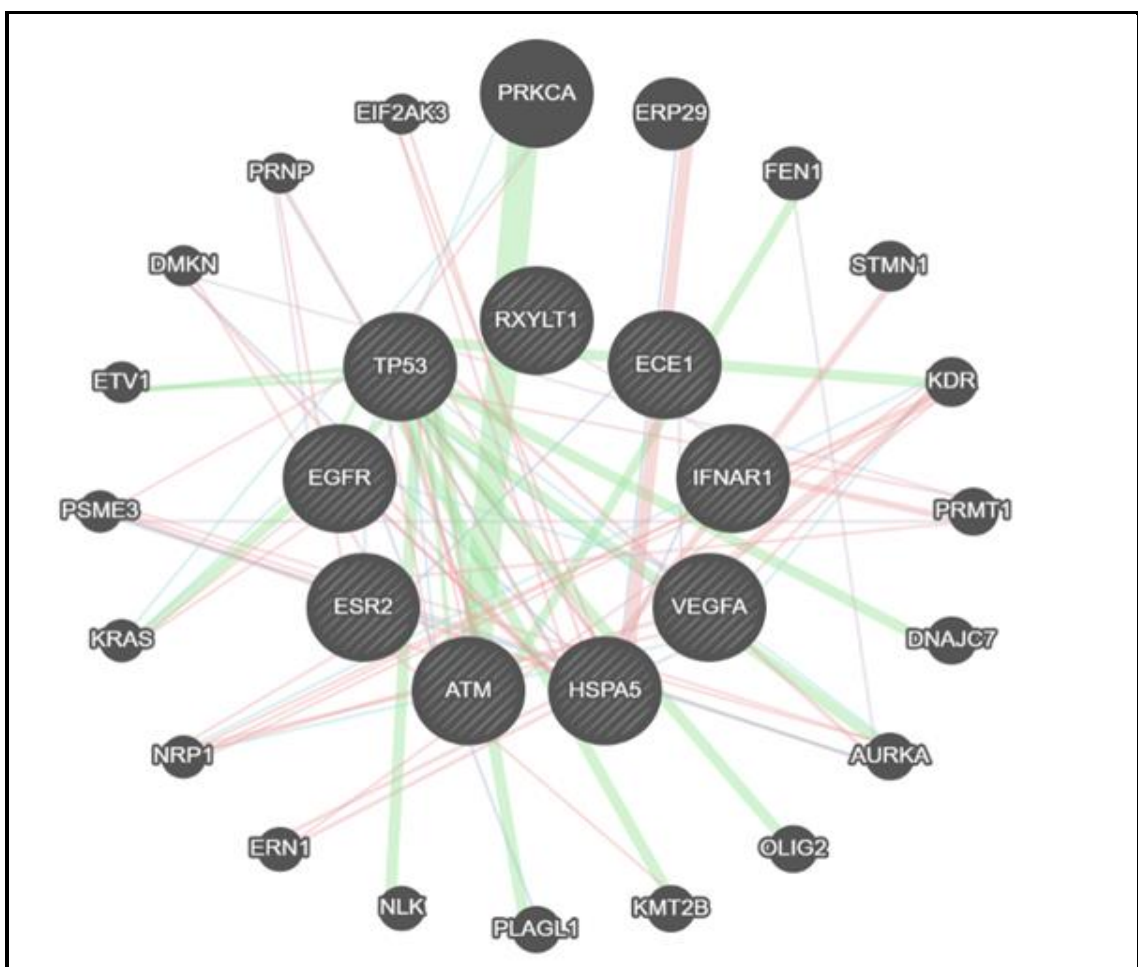
Figure 48: A concentric layout of multiple genes using GeneMania is displayed

Further, (**Figures. 46, 47 & 48)** the ligands and receptors collected for docking studies, we have compared the protein-protein interactions (PPI) studies using String software, GeneMania, and the network interactions connected to various biological pathways,

molecular and cellular pathways. Interferons (IFNs) are cytokines that generally exhibit potent anti-viral and immunomodulatory properties and have anti-proliferative properties due to the induction of interferon-stimulated gene (ISG) transcription factors. IFNs show mixed results in cancers induced by HR-HPVs or in treating low-grade lesions. For example, **(Table 13)** Interferon 1 showed good activity with HPV 16 (-576.8 Kcal/mol) and 31 types (-486.94 Kcal/mol); other types did not.

6.5 Protein- Protein Docking

The main aim of this in silico studies is to perform PPI interactions and Protein - Protein Docking studies to interpret the interactions among the molecules of E6, E7 and L1 capsid protein (Gardasil 9) as ligands of HPV types 6, 11, 16, 18, 31, 33,45, 52 and 58 with the receptors of several molecules to know the binding affinity, energy between the HPV type and the receptors. Gardasil is quadravalent and available all over the world and Gardasil 9 is nanovalent and available in only a few countries like the USA, Europe etc. In India Gardasil is available as a vaccine for girls 9-26 years of age. By using in silico Computational tools docking was performed to see the relatedness and differences between the HPV types and several receptors.

The PP docking was carried out using Hex software and ClusPro was used to validate the docking studies. EGFR has found the best interaction results with all inactive L1 capsid proteins using Gardasil 9 followed by Heparan Sulfate. L1 capsid proteins of HPV 16 and HPV 31 have shown activity with all the selected receptors; hence they may be used as better vaccine components for the control of HPV. The docking studies check the binding affinities of L1 capsid protein with the HPV types to elucidate the fundamental biochemical process. We completed the comparative docking studies using two software Hex 8.0 (**Table 13**) and ClusPro **Table 12:** Protein-Protein Interaction of Capsid proteins L1 from Gardasil with selected receptors using HEX 8.0 Software results.

Receptors	6L31 6	1DZL116	2R5K11	2R5I 18	Model 31	6IGE 33	Model 45	6IGF 52	5Y9E 58	6+11+16
Docking results (in Kcal/mol)										
Interferon 1	-0.23	-576.8	35.3	30.8	-486.	51.1	38.2	25.9	20.3	-26.9
Laminin 5	-68.73	-8.2	24.3	49980.8	-407.5	116.3	19955.2	81.8	54.5	-30.6
Estrogen	-106.3	-628.2	51064.3	-121.5	-492.6	-70.6	-61.3	-78.8	-72.9	-126.6
Endothelin A	-85.5	-0.13	-78.6	36814.7	-85.5	-106.1	36739.0	-103.1	-57.9	-112.1
FcγRIII Ig receptor,	-130.8	-728.8	-134.4	-311.21	-554.2	-119.0	-400.	-121.6	-159.2	-146.0
ATM	0.55	-197.7	197.2	0.37	-33.7	0	0.70	0	287.4	0.01
Heparin Sulphate	-257.1	-609.2	-471.3	-872.2	-471.8	-490.9	-856.3	-320.8	-67.1	-199.4
VEGF	-57.1	-673.3	34.1	-54.92	-649.0	-50.9	-32.7	-47.9	-39.4	-63.4

EGFR	GRP78	Integrin α6β1
-115.1	0.04	-67.5
-675.2	-213.4	-740.3
-537.9	30714.	-806.
-277.77	33793	-822.8
-523.8	-37.36	-575.2
-108.7	34027	-227.
-239.8	34657	-727.1
-298.1	17298	-51.9
-103.3	481.5	-49.3
-137	0	-63.4

Table 13: Docking studies of HPV Gardasil 9 types binding with receptors

The results were downloaded and analyzed using Pymol visualizer software. The lowest molecules and binding energies. Based on their binding or interactions, the two software predicted the highest cluster and lowest energy scores. The Hex 8.0 docking studies of HPV 9 types of Gardasil 9 with different receptors show the binding affinity of the molecules, and the binding energy results are highest in EGFR (-523.8 kcal/mol), immunoglobulins and then followed by Heparin Sulfate (-856.3 kcal/mol).
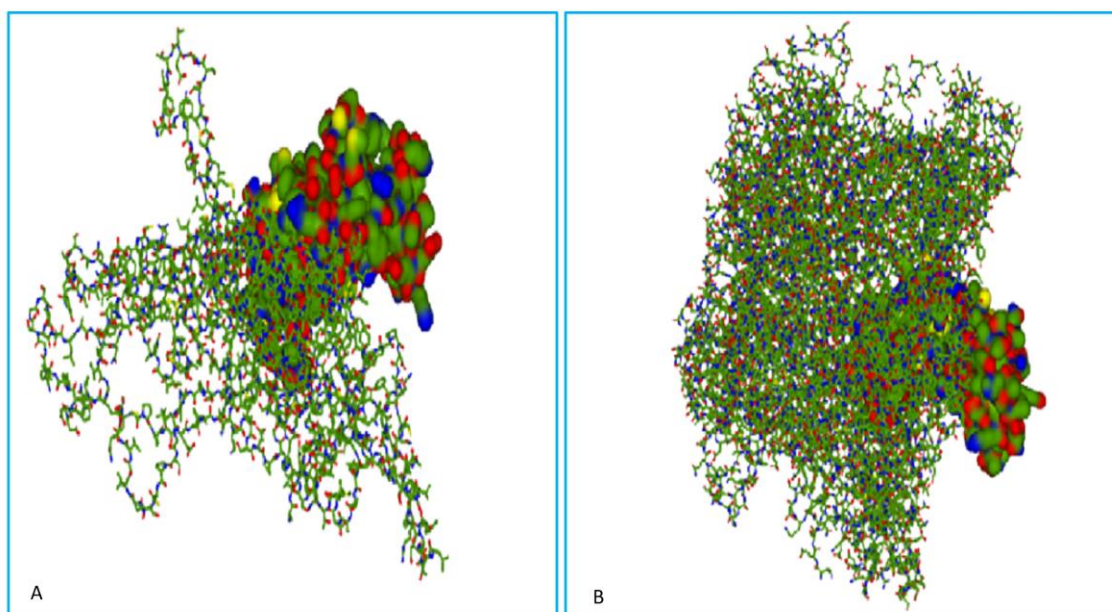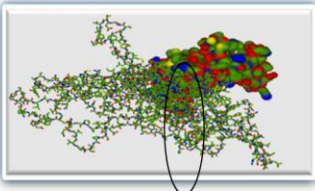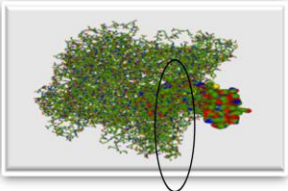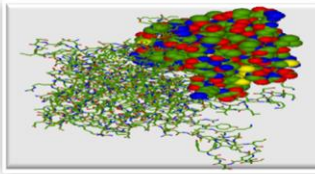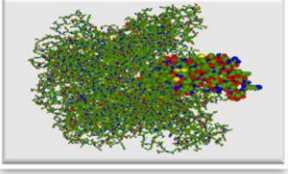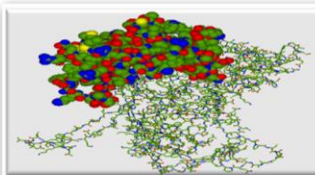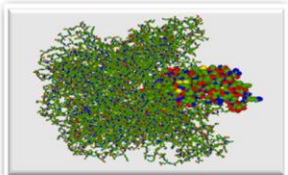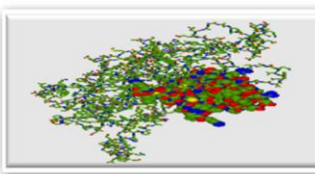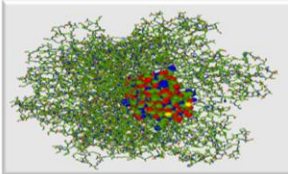
Figure 49: Protein-Protein Interferon between (A) HPV 16 EGFR with 1DZL (B) HPV1 16 Heparin Sulphate with 2R5K

Further, (**Figures 49**) the ligands and receptors collected for docking studies, we have compared the protein-protein interactions (PPI) studies using String software, GeneMania, and the network interactions connected to various biological pathways, molecular and cellular pathways. Interferons (IFNs) are cytokines that generally exhibit potent anti-viral and immunomodulatory properties and have anti-proliferative properties due to the induction of interferon-stimulated gene (ISG) transcription factors. IFNs show mixed results in cancers induced by HR-HPVs or in treating low-grade lesions. For example, **(Table 3, 4)** Interferon 1 showed good activity with HPV 16 (-576.8 Kcal/mol) and 31 types (-486.94 Kcal/mol); other types did not.

	1DZL	2R5K
H6E6		
H6E7		
H11E6		
H11E7		

Ig receptor, FcγRIII (CD16)		
ATM		
Heparin sulphate		
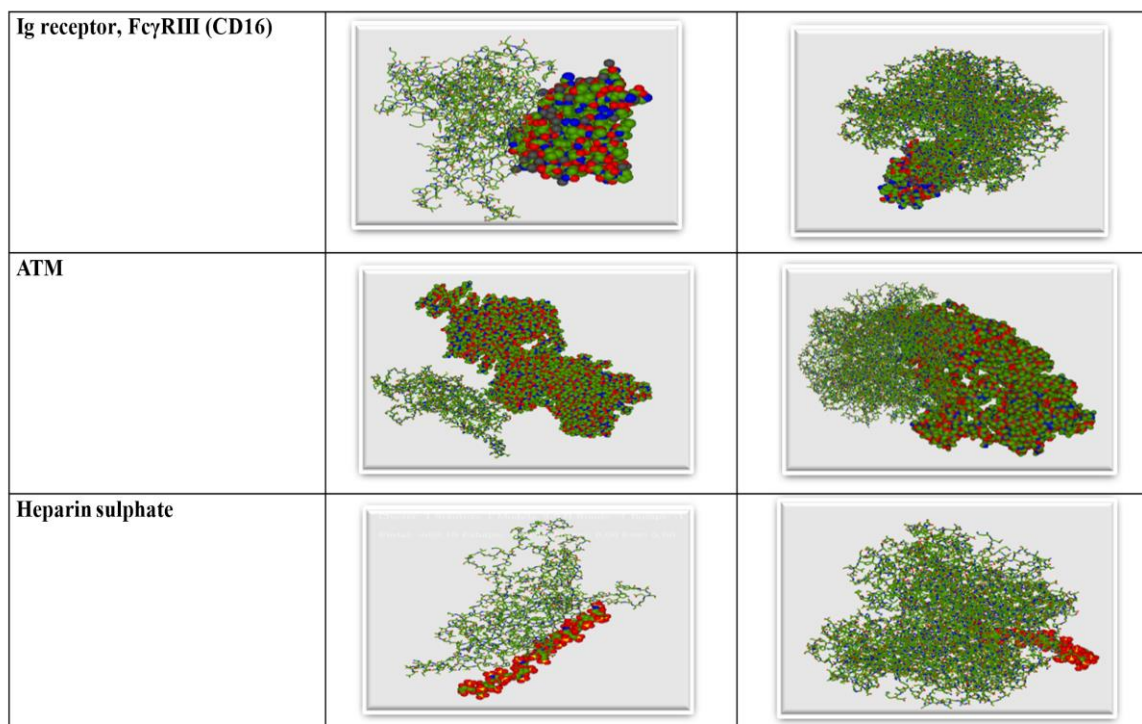
Figures 50: Docking results of HPV using HEX 8.0 with all the receptors

Receptor		Docking results (in Kcal/mol)									
		6L31 6	1DZL 16	2R5K 11	2R5I 18	Model 31	6IGE 33	Model 45	6IGF 52		5Y9E 58	6+11+16+18
Interferon 1	molecules	38	39	25	191	75	22	66	25	31	52
	Lowest energy	-1232.9	-1048.8	-924.4	-832	-925.0	-870.0	-995.5	-788.2	781.5	-851.2
Laminin 5	molecules	22	42	29	51	38	40	63	26	35	67
	Lowest energy	-1417.3	-1403.5	-1035.6	-1152.5	-1180.5	-1347.5	-1106.1	-1104.6	-1269.1	-1234

120

Estrogen	molecules	80	62	36	51	53	32	166	68	32	32
	Lowest energy	-948.8	-1358.9	-860.7	-980.2	-1168.9	-852.2	-1094.9	-978.4	-856.6	-1076
Endothelin A	molecules	45	45	19	119	37	50	69	41	29	43
	Lowest energy	-1018.7	-856	-766.7	-813.6	-1020.9	-852.8	-1209	-725.9	-765	-876
FcγRIII Ig receptor, (CD16)	molecules	85	64	63	147	117	56	105	38	30	40
	Lowest energy	-1118.3	-1006.7	-874.1	-863.8	-742.9	-742.4	-787.5	-769.6	-674.8	-927
ATM	molecules	28	45	12	29	35	62	69	68	14	23
	Lowest energy	-1279.8	-986	-834.9	-998.4	-1036	-954.3	-1209.6	-978	-844.8	-765
Heparin Sulphate	molecules	46	61	32	74	82	32	46	42	26	41
	Lowest energy	-1243.6	-1040.6	-1010.9	-934.3	-811.2	-1010.9	-943	-858.6	-927.3	-1129.4
VEGF	molecules	31	46	17	63	62	21	89	17	19	45
	Lowest energy	-1080	-1033	-789.7	-1000.5	-732.0	-806.5	-965.9	-713.0	-711.7	-879
EGFR	molecules	33	41	13	56	58	48	54	58	20	65

	Lowest energy	-1312.6	-1141.8	960.1	-1028.1	-865.9	-935.8	-1052.4	-1027.1	-876.2	-985
Integrin α6β1	molecules	38	39	25	191	75	22	66	25	31	-27
	Lowest energy	-1232.9	-1048.8	924.4	-832	-925.0	-870.0	-995.5	-788.2	-781.5	-934.8
GRP78	molecules	70	63	26	62	29	22	97	33	37	87
	Lowest energy	-1143.2	901.2	834.2	-737.9	-815.6	-750.8	-903.9	-828.4	-777.7	-879.8

Table 14: Protein -Protein docking results and their binding affinities using ClusPro 2.0 protein-protein docking software

The greater the negative value, the greater the binding affinity and E value. The molecules are docked, and figures are shown in the ball and stick model **(Figure 50).** The ClusPro results predict that the ligand-receptor docking studies of HPV Gardasil 9 show great results. The best docking cluster in our study represents lamnin5 (PDB ID: 5XAU), with all HPV types having 22 members showing -1417.2 as the lowest energy, and they are best docked binding affinity, next to it, EGFR -1312.6 Kcal mol, ATM -1279.8 Kcal/mol and Heparin Sulfate **(Figure 50 & 51)** **(Table 14).**
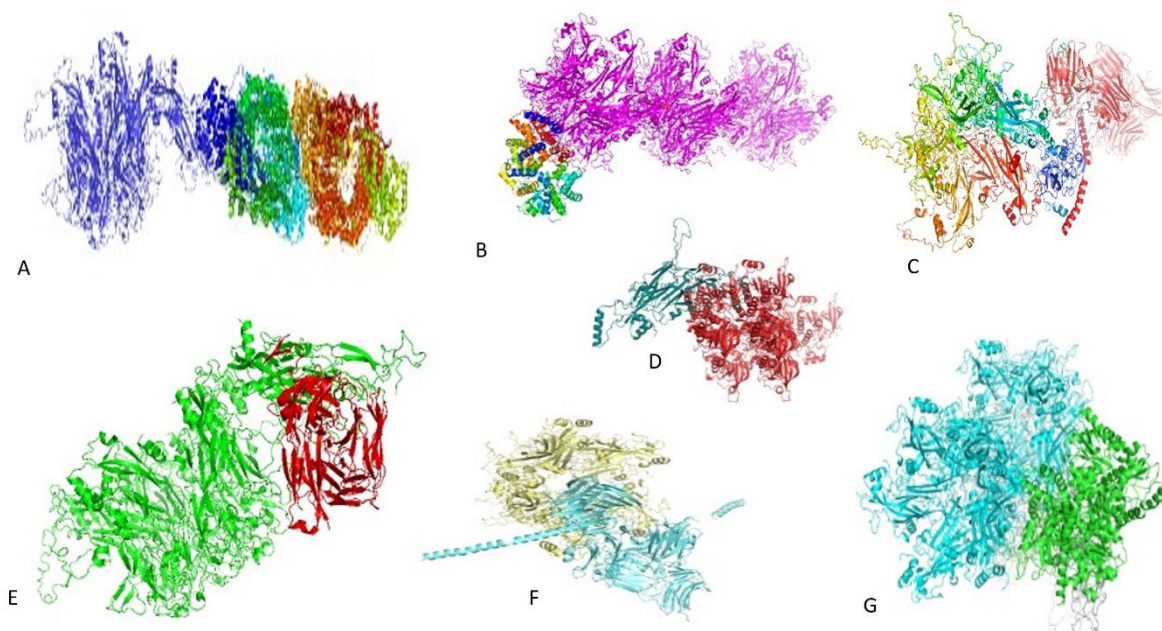
Figure 51: Protein -Protein Docking using Cluspro software (A) shows binding between receptor ATM (PDB ID: 7SIC) and ligand HPV6 (PDB ID: 6L3I) (B) Receptor Endothelin A (PDB ID: 3DWB) and ligand HPV31 (PDB ID: 2R5I. (C) Receptor Laminin 5 (PDB ID: 5XAU) with ligand HPV 6 (PDB ID: 6L3I). (D) Receptor EGFR (PDB ID: 5GTY) with ligand HPV 16 (PDB ID: IDZL) (E) Receptor EGFR (PDB ID: 5GTY) with ligand HPV 6 (PDB ID: 6L3I). (F) Receptor VEGF (PDB ID: LBJ1) with ligand HPV 6 (PDB ID: 6L3I). (G) Receptor Laminin (PDB ID: 5XAU) with ligand HPV 16(PDB ID: IDZL)
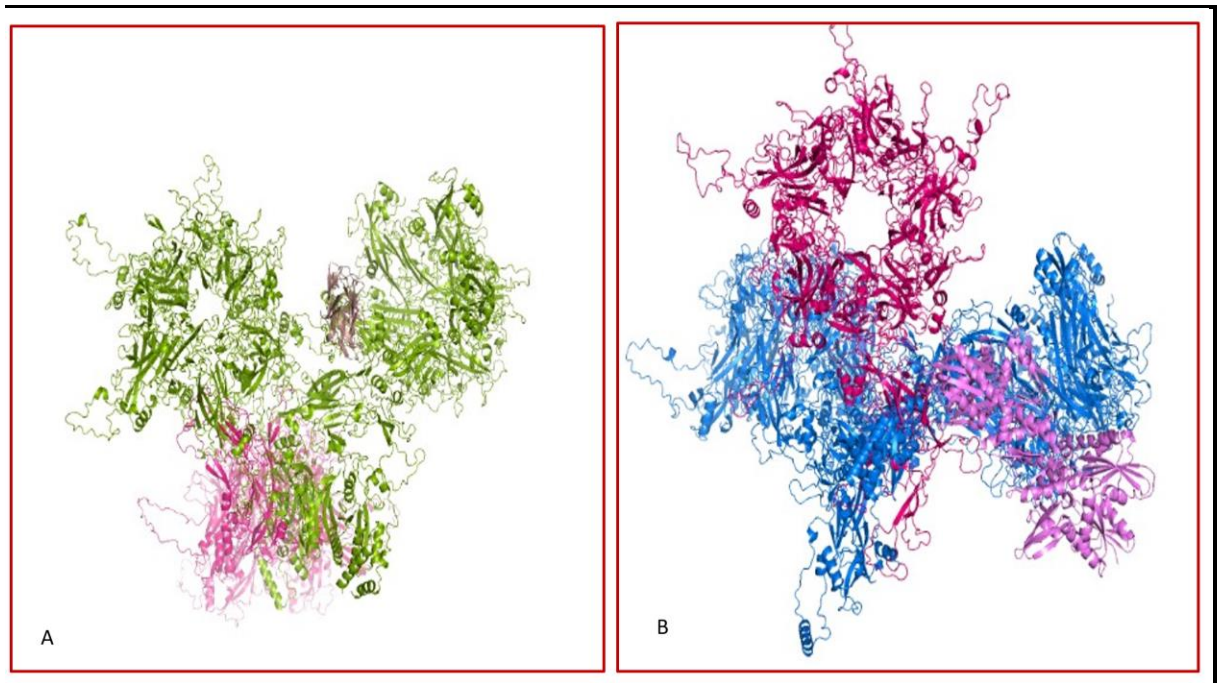
Figure 51: (A) Shows interaction between FcgammaIII (PDB ID: 6MJO) and HPV 6, 11, 16, 18 (6L3I, 2R5K, IDZL, 6IWD). (B) Shows interaction between GRPT8 (PDB ID: 3IAU) HPV 6,16,18,31 (6L3I,2R5K, IDZL, 6IWD)

The molecular docking studies show higher binding affinities and lowest energies through ClusPro software with Laminin 5, EGFR, and interferons, followed by ATM and Heparan Sulfate. HPV types with the receptors of these molecules show selective binding. IFNs have been magnificently treating patients with genital warts (GW), a sexually transmitted disease that was induced by LR HPV types. HPV 6 and 11 are significant in Genital warts (GW) associated with risk factors. A prophylactic quadrivalent HPV Gardasil vaccine (L1 capsid HPV types 6, 11, 16, and 18) is vastly effective in inhibiting disease-associated GW and HPV types 6, 11, 16, and 18. In addition, endothelin-1 (ET-1) and its receptors are released from HPV-transfected keratinocytes and can be directed for antitumor therapy. Laminin 5 is an extracellular component showing over-expression and increased localization of (CIN). Integrin is also associated with increased expression of cervical carcinoma (Malinowski D.2007). HPV virus-like particles (HPV VLP) and chimeric virus-

like particles are immunogens that act as potent anti-viral/tumor T and B cell responses. FcγRIII (CD16) is a molecule that recognizes VLP and can bind both on the foreskin epithelium and immune cells. In addition, L1 and L2 significant and minor capsid proteins of HPV-16 can self-assemble into virus-like particles (VLP). HPV VLP can be a lead candidate that acts as a prophylactic and therapeutic vaccine against several types of HPV infection (Da Silva et al.2001, Panatto et al.2015). The major and minor capsid proteins of HPV and their Interaction with Heparan Sulfate (HS) are essential for the triumphant entry of HPV16 (Richards et al. 2013) Cell-surface Heparan Sulfates are generally undeviating and highly negatively charged oligosaccharides that show antiviral effects against HPV (Drobni, and Näslund 2004).HR-HPVs type 16 is the main factor in cervical carcinogenesis. Estrus and estrogen treatment show the persistence of upregulation of transgene expression and controlling all stages of carcinogenesis. Estrogen and its receptors are concerned with preventing and promoting cervical cancer strongly associated with HPV infections (Chung & Sang-Hyuk,2010). HPV-16 oncoproteins promote the advance of non-small cell lung cancer (NSCLC), possibly by activating HIF-1α/VEGF-mediated tumor angiogenesis (Li, Gang, et al.2011). VEGF or telomerase reverse transcriptase (hTERT) is an inflammatory and angiogenic factor that plays a role in the regulation of HPV-induced cervical cancer (Li, Fang and Jinquan Cui. 2015).

Further, this infection leads to cervical cancer initiation and progression due to tobacco smoking. CaSki (HPV16) and SiHa (HPV16) are cervical cancer-derived cell lines. When exposed to cigarette smoke components, increased E6 and E7 levels are due to EGFR activation and c-Jun phosphorylation (Muñoz, Juan P., et al.2018). Activation of EGFR, cyclooxygenase (COX)-2, and ErbB-2 signaling develops cervical cancer via CIN lesions (Fukazawa et al.2014). Glucose-regulated protein 78 (GRP78) is mainly responsible for upregulating many cancers, including cervical cancer. Whenever GRP78 binding is inhibited, there will be a faster degradation rate of E6 by the host cell degradation machinery (Elfiky, Abdo A.2020). Sacchin αβ4- and α6β1-integrins are associated with ErbB-2 that can be used in the detection of CC (Du et al.2006). In addition, HPV16 infection efficiency has decreased expression due to the loss of β4 integrin (Aksoy et
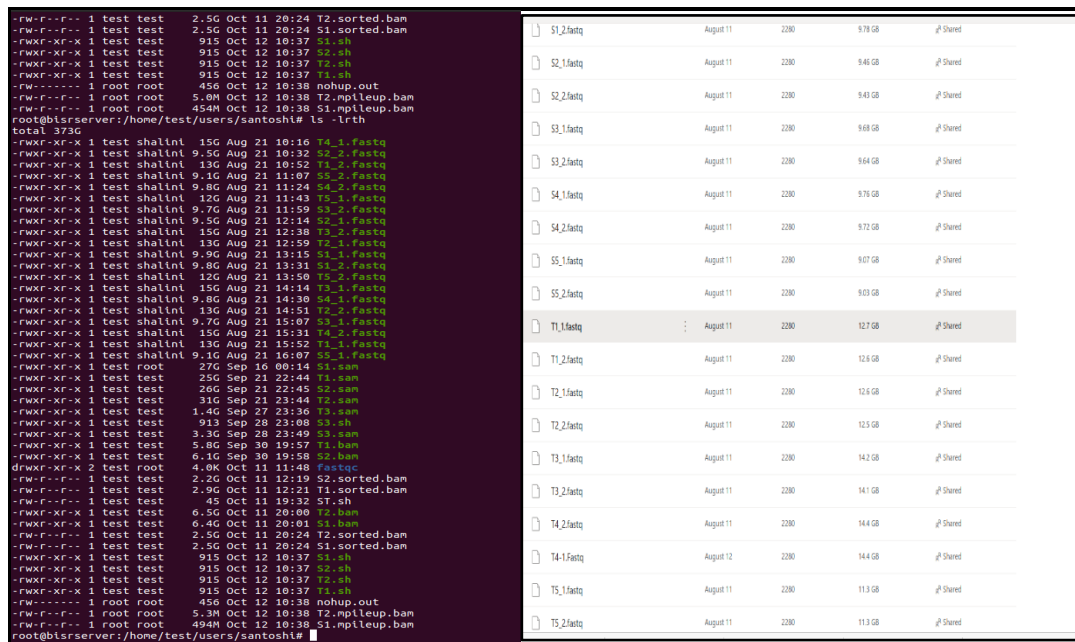
al.2014). The docking results showed that Gardasil 9 HPV vaccine was more effective with selected receptors like Interferon 1, Laminin 5, Estrogen, Endothelin A, FcγRIII Ig receptor, (CD16), Heparin Sulphate, VEGF, EGFR, and Integrin α6β1. Hence, earlier vaccine development programs commercialized numerous authorized prophylactic vaccines, all major capsid (L1) proteins. Because of the modest susceptibility supplied by the capsid L2 protein resulted in the development of second-generation clinical preventive HPV vaccines, none of which have yet been authorized (Bagheri et al.2021), the other side, post-translational proteins E 6 & 7 have been used to generate potential prophylactic vaccines. The *in-silico* design of therapeutic and preventive vaccination against HPV types is done here. It is our first pilot study combination of both invitro and in silico computational analysis performed, genotyping based on RFLP, in silico work on phylogenetic analysis, the combination of PP interaction and PP docking studies done, As E6 and E7 are early oncoproteins express an vital role in the Accumulation, proliferation, transformation etc., of cancer, In our study HPV Genotypes through HPV genotyping, was performed and found HPV 6, 16, 18, 33 from 20 samples, of which 6 samples were positive and remaining were negative for HPV. The HPV sequences were submitted through the PAVE database with multiple sequence alignment to see the evolutionary relationship of E6 and E7 oncoproteins So, we have performed a Phylogenetic analysis of E6 and E7 sequences of HPV constructed and predicted LR 6, 11, and HR HPV types 16, 18, 33. Further, added PP interactions and PP docking to know the binding affinity and energy of the E6, E7 and L1 capsid protein of HPV and the receptors like endothelin A, VEGF, Heparin Sulfate, Laminin, interferon, etc. to know the type of HPV and its binding energy, to predict the efficacy and therapeutic intervention through computational studies.

6.6 To evaluate the genetic variants through whole exome sequencing

1. NGS Preprocessing **2.** Variant Discovery **3.** Variant Prioritization

6.6.1 Quality Check

The raw reads after Illumina Sequencing are checked for contamination whether the base quality scores, per base sequence quality, per base GC content, Sequence duplication levels etc., show contamination is present or not by checking with per base sequence quality, if is >20 so the sequence of the samples is good and if it less than <20 there is a chance of contamination. The analysis was performed on the Linux terminal and the Quality check was performed by using the command line FastQ for forward strand and reverse stand. Below were the images of the Quality of tumor samples and surrounding tissue or control samples. All the samples were run and the files were downloaded in HTML files (**Figure 52, 53**).

Figure 52: Fastq files are downloaded and results are visible in HTML file

Quality check FASTQ files (**Figure 53** are mention below)

S1_1. FastQc
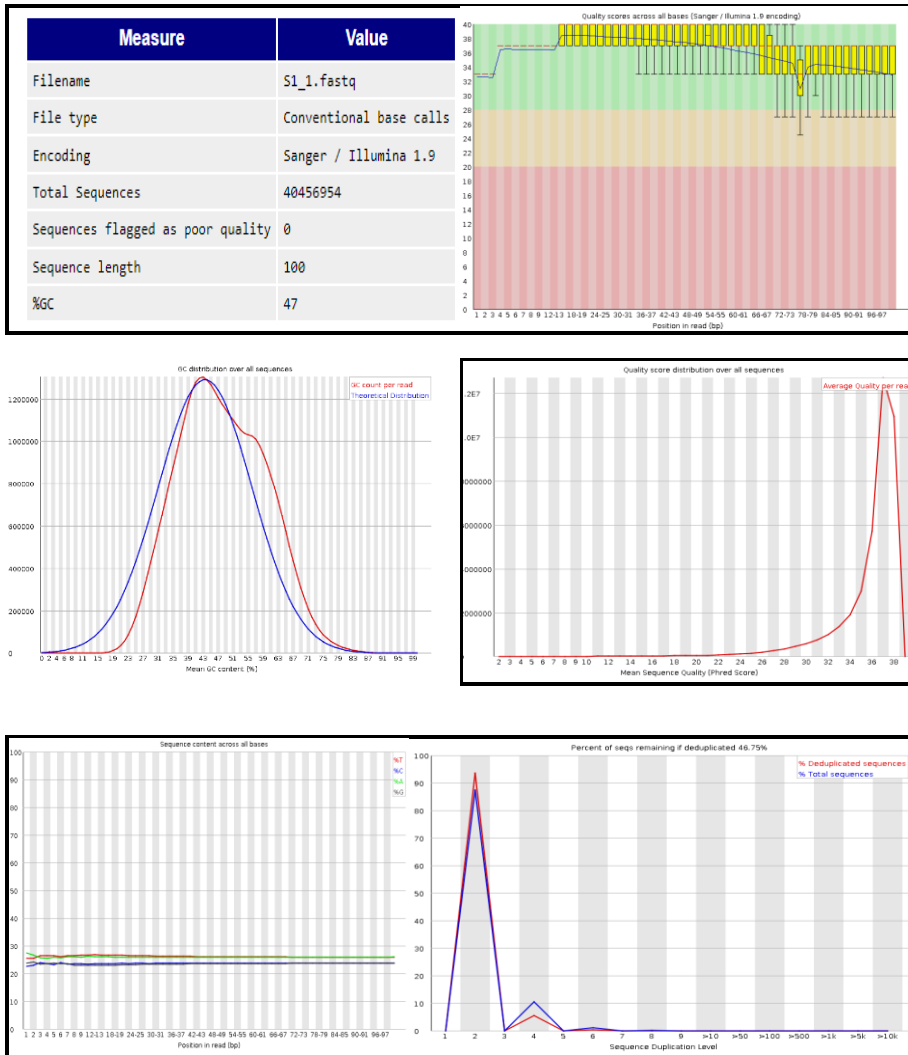
Figure 53 (a): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). S1 _ 1. FastQc

S1__2. FastQc
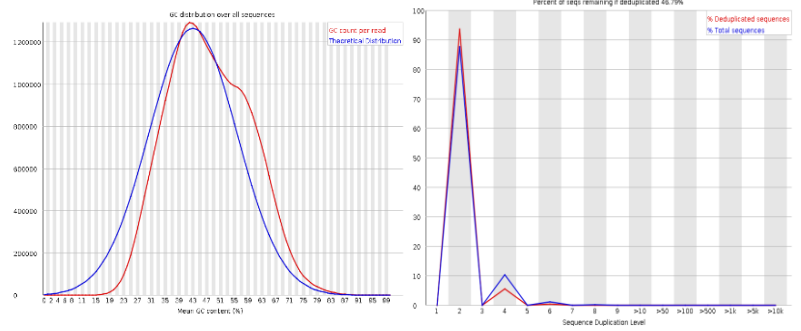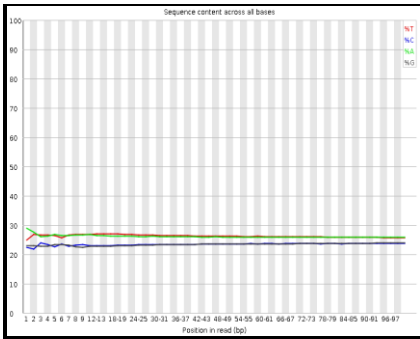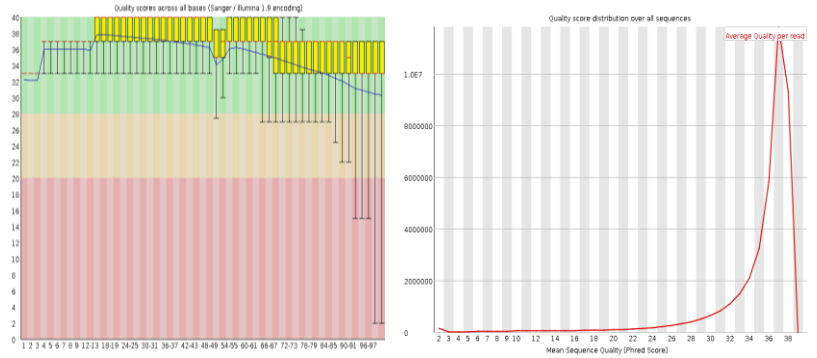
Figure 53 (b): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). S1 _ 2. Fast

S2_1. FastQc

Figure 53 (c): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). S2 _ 1. FastQc

S2_2. FastQc

Measure	Value
Filename	S2_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	39008904
Sequences flagged as poor quality	0
Sequence length	100
%GC	47

130
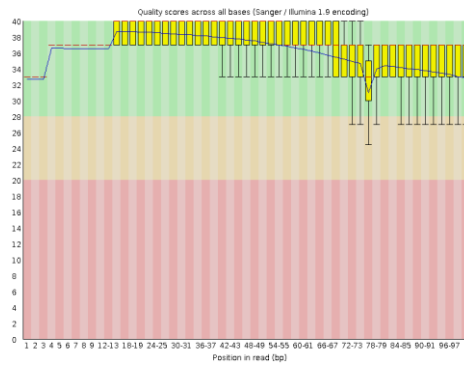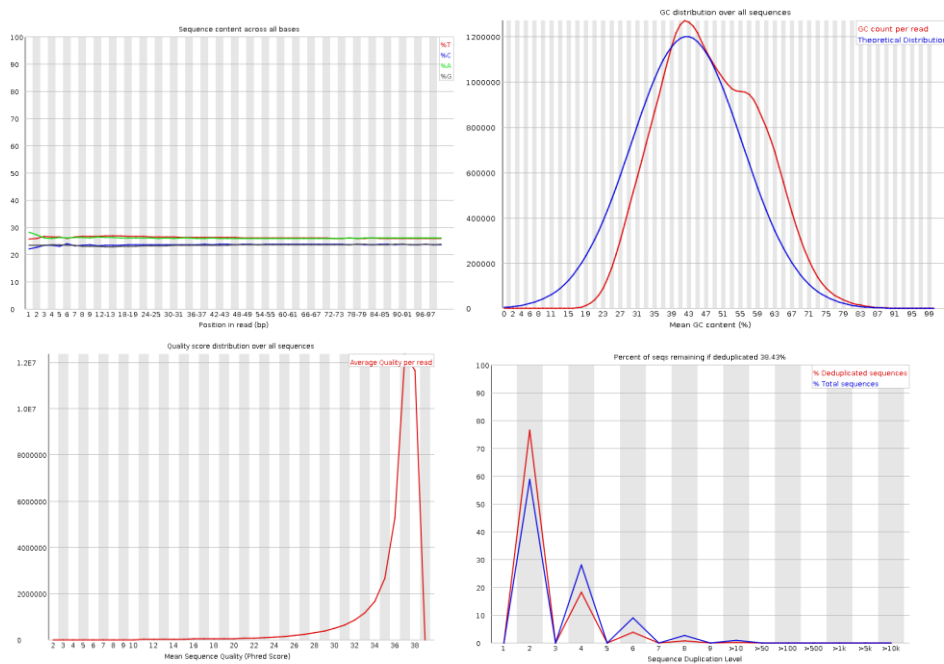
Figure 53 (d): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). S2 _ 2. FastQc

T1_1. FastQc

Measure	Value
Filename	T1_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	52667854
Sequences flagged as poor quality	0
Sequence length	100
%GC	47
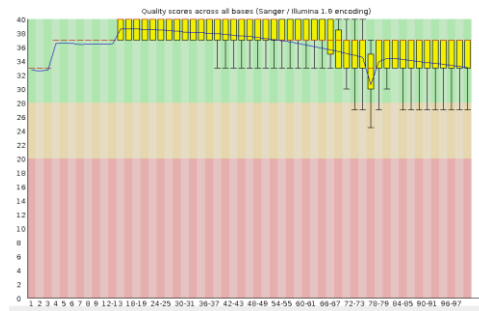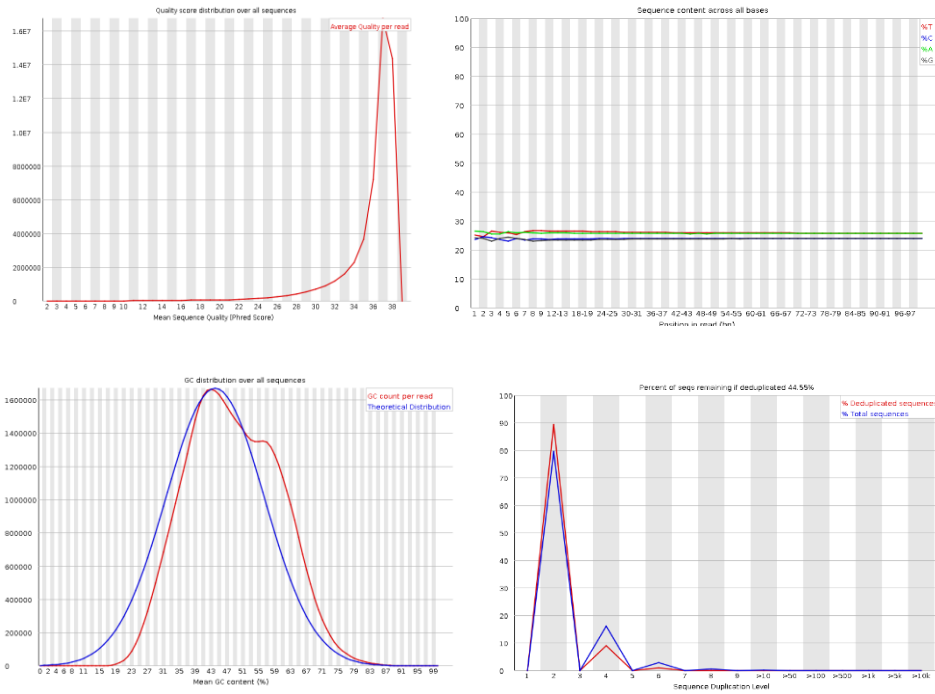
Figure 53 (e): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). T1 _ 1. FastQc

T2_1. FastQc

Measure	Value
Filename	T2_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	51782488
Sequences flagged as poor quality	0
Sequence length	100
%GC	48
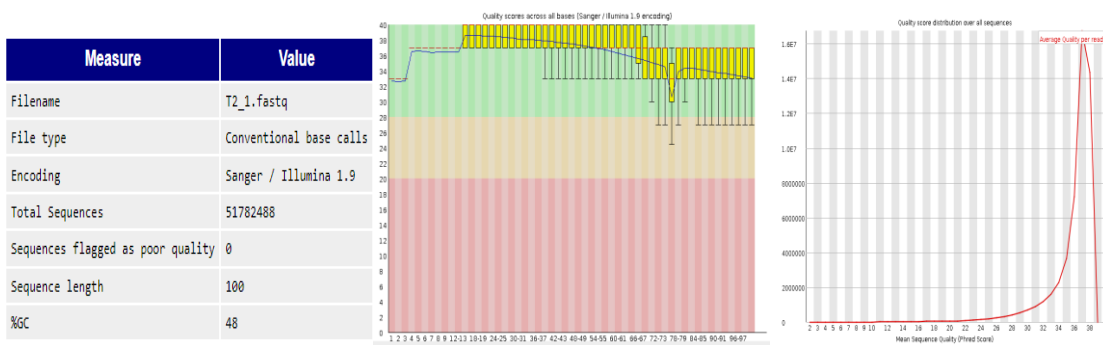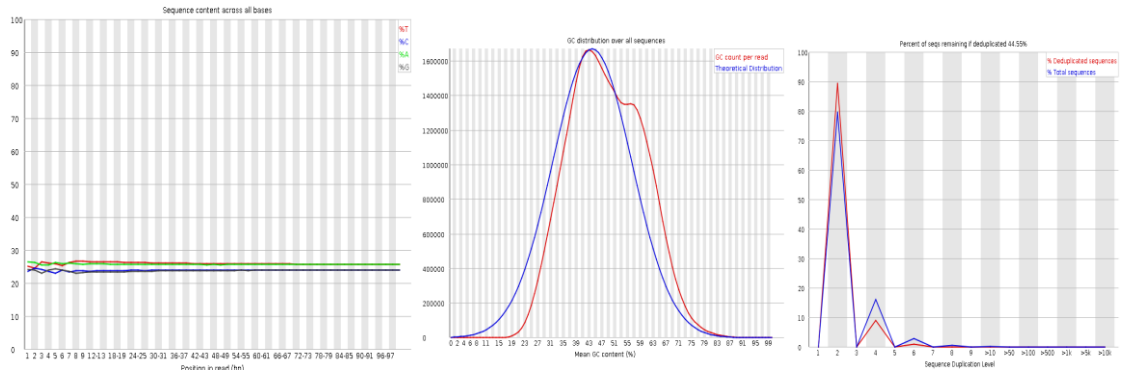
Figure 53 (f): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence levels duplication of all the tumor and control samples (T- tumor and S- control). T1 _ 2. FastQc

T2_2. FastQc

Measure	Value
Filename	T2_2.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	51782488
Sequences flagged as poor quality	0
Sequence length	100
%GC	48
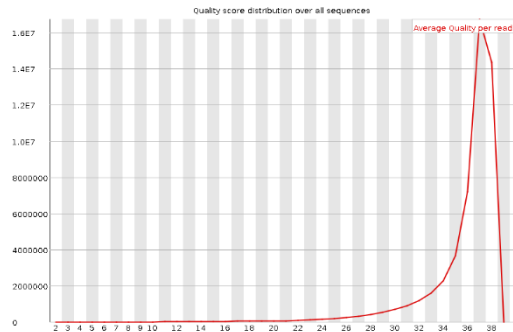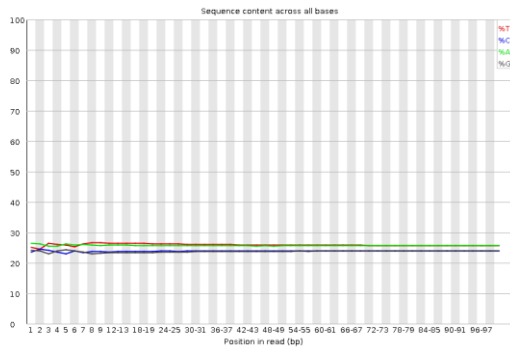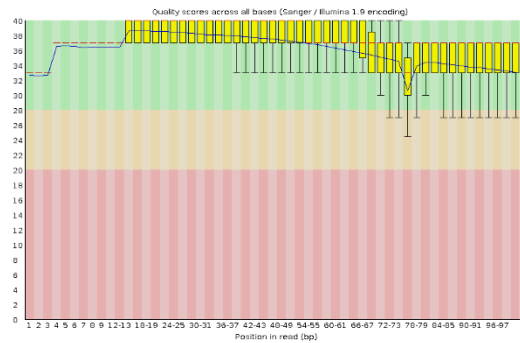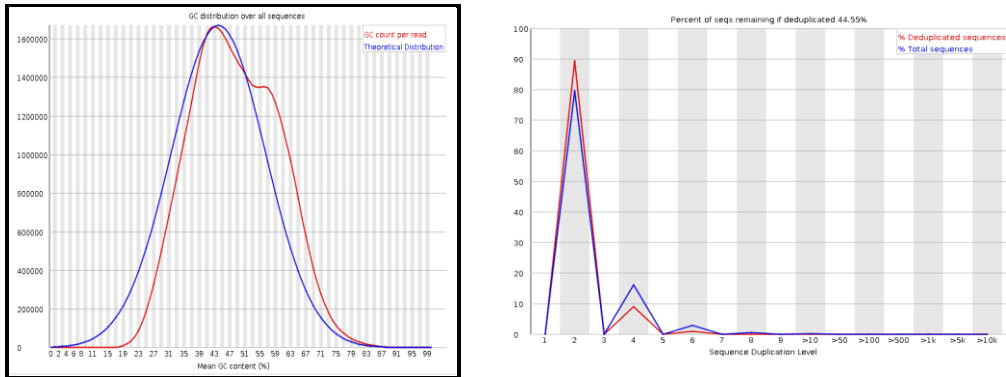
Figure 53 (g): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). T2 _ 2. FastQc

S3_ 1.FastQc

Measure	Value
Filename	S3_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	40041016
Sequences flagged as poor quality	0
Sequence length	100
%GC	47
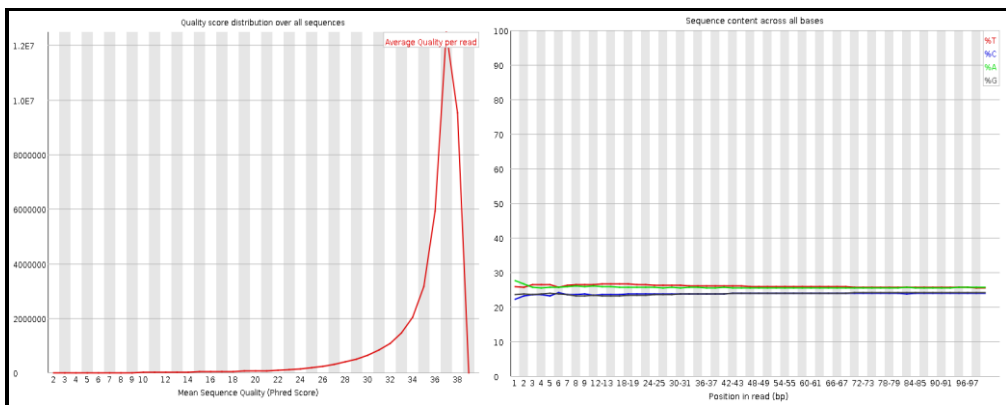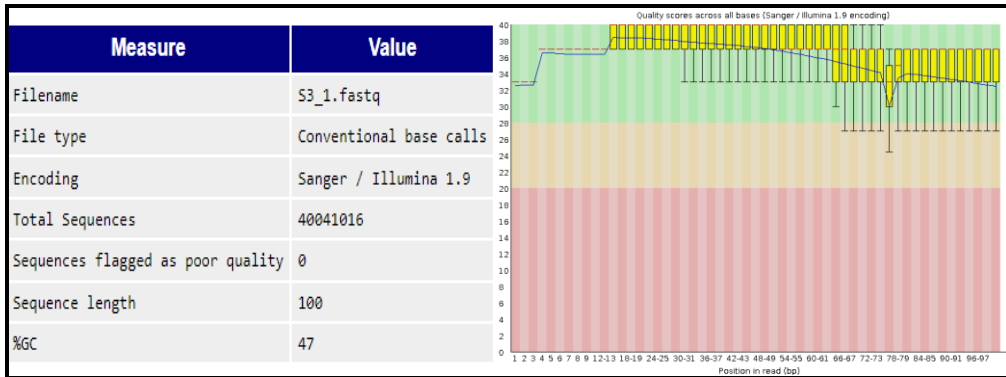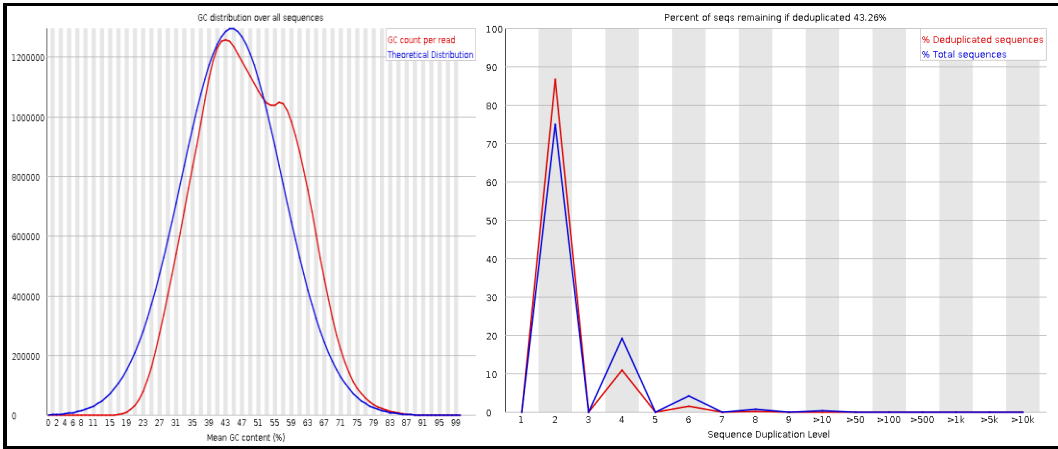
Figure 53 (h): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). S3 _ 1. FastQc

T3_1.FastQc

Measure	Value
Filename	T3_1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	58774594
Sequences flagged as poor quality	0
Sequence length	100
%GC	47
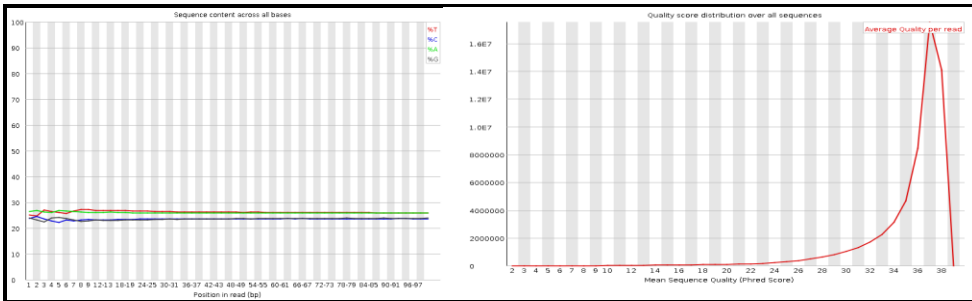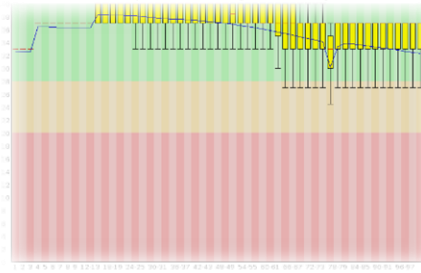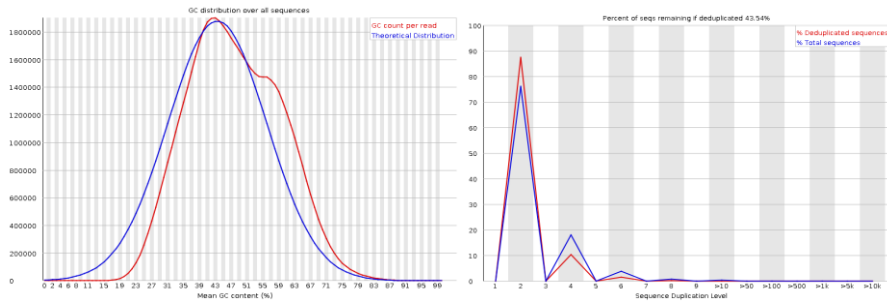
Figure 53 (i): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). T3 _ 1. FastQc
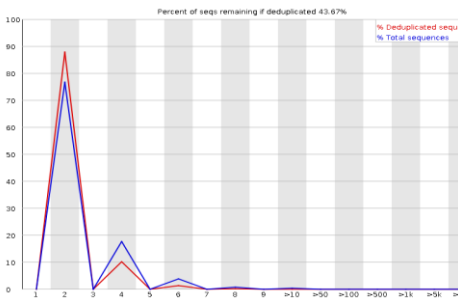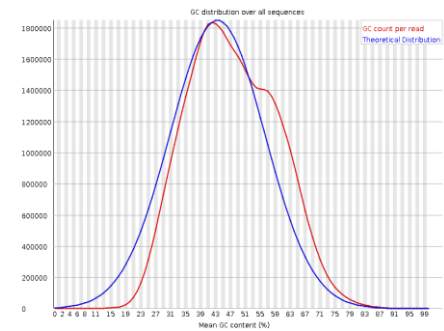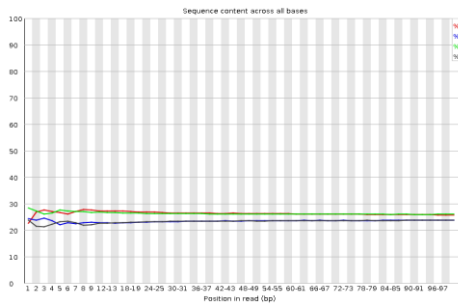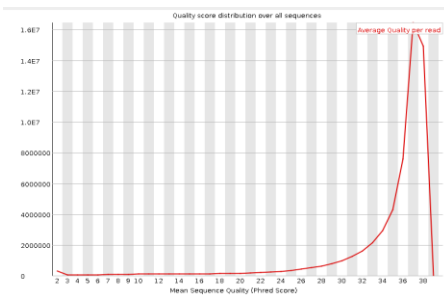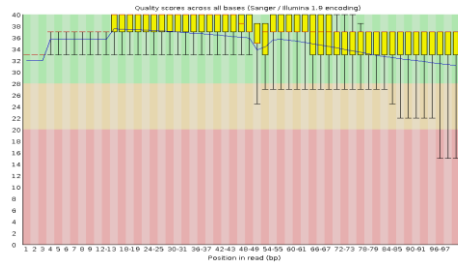
T3_2.FastQc

Figure 53 (j): FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). T3 _ 2. FastQc

T5_1.FastQc
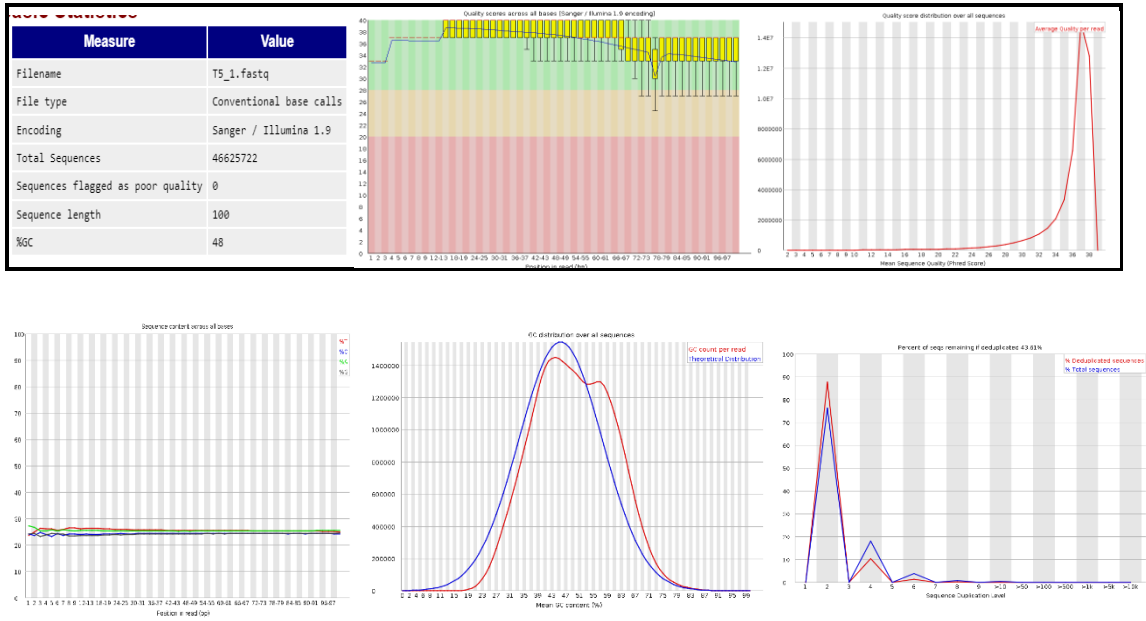
Figure 53 (k): Graphs of FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). T5 _ 1. FastQc

T5_2.FastQc
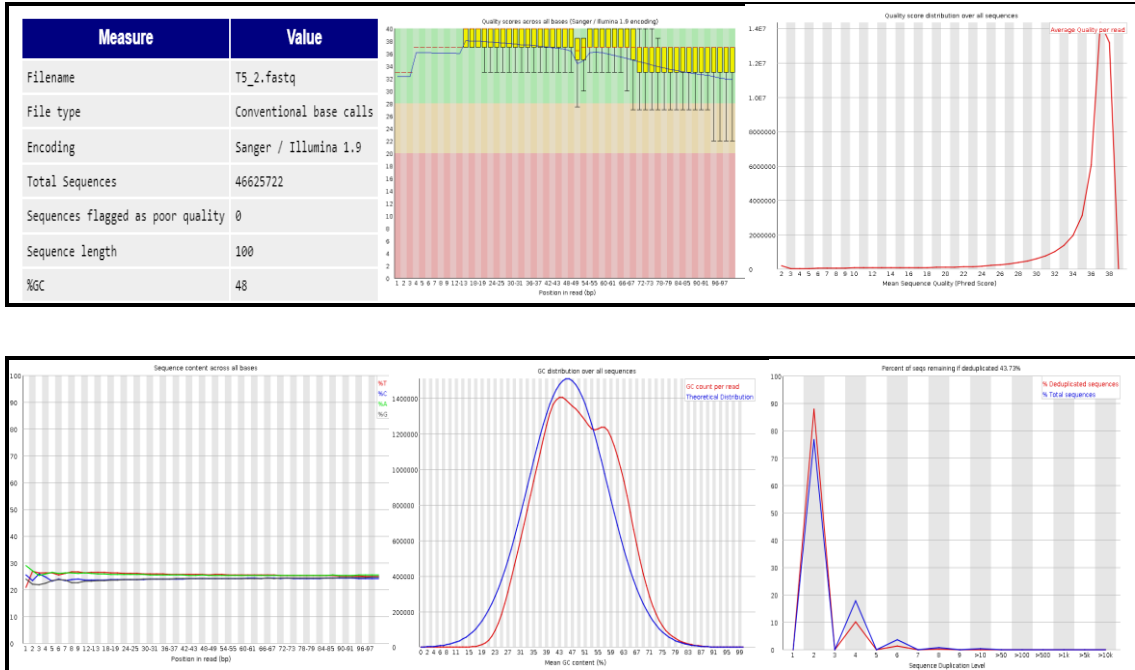
Measure	Value
Filename	T5_2.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	46625722
Sequences flagged as poor quality	0
Sequence length	100
%GC	48

Figure 53 (l): Graphs of FastQC reads of samples determining per base quality scores, per base sequence quality, per base GC GC content, sequence duplication levels of all the tumor and control samples (T- tumor and S- control). T5 _ 2. FastQc

FASTQ files were analyzed using the Linux command line on Ubuntu by giving FastQC and sample numbers. All files were run for downloading 100% in the computer, and then Html files of fastqc were retrieved and downloaded. **(Figure 53 a, b, c, d. e, f, g, h, i, j, k, l)** checked for GC content, contamination, per base sequence quality, etc. **(Table 15)** (Meena et al.2018).

Sample Name	GC percent	Sequence Length	Per base sequence quality	Per sequence Quality scores	Per base sequence content	Per sequence GC content	Sequence Duplication Levels
S1_1.fastq	47	100	>20 good -26	1.2 E 7	all are straight (A, T,G,C) - good	47	43.75%
S1_2.fastq	47	100	few below 20	1.2 E 7	all are straight (A, T,G,C) - good	46	46.79%
S2_1.fastq	47	100	>20 very good - 25	1.2 E7	All are straight (A, T,G,C) - good 47	47	38.43%
S2_2.fastq	46	100	good but few <20	1.2 E7	All are straight ((A, T, G, C)	46	46.79%
S3_1.fastq	47	100	>20 and value 26	1.2 E 7	good all are straight	47	43.26%
S3_2.fastq	47	100	>20 value 22	1.2 E 7	all across the sequence	47	43.27%
S4_1.fastq	47	100	>20 value is 24	1.2 E 7	all across the sequence	47	46.46%
S4_2.fastq	47	100	few below 20	1.2 E 7	all across the sequence	47	46.55%
S5_1.fastq	44	100	gradual Dec. below 20 few	1.0 E 7	slight changes	44	41.91%
S5_2.fastq	44	100	gradual dec below 20 few	9000000	slight curves	44	42.31%
T1_1.fastq	47	100	>20 value 25	>1.6 E 7	all are straight	47	44.95%
T1_2.fastq	47	100	>20 value	>1.6 E7	all are straight	47	44.46%
T2_1.fastq	48	100	>20 good 25	>1.6 E 7	all are straight	48	44.55%
T2_2.fastq	48	100	>20 good 25	>1.6 E 7	all are straight	48	44.55%
T3_1.fastq	47	100	>20 good 25	>1.6 E E 7	all are straight	47	43.54%
T3_2.fastq	47	100	few are below 20	>1.6 E 7	slight curve but straight	47	43.67%
T5_1.fastq	48	100	>20 value 25	> 1.4 E 7	straight line	48	43.61%
T5_2.fastq	48	100	>20 value 22	>1.4 E 7	straight line	48	43.73%

Table 15: FastQc scores of Tumor and control samples and their GC content, per base sequence quality, per base Quality scores, Per sequence duplication levels.

6.7 Annotation by SNP Nexus Software

We observed that the total number of annotated variants of all samples is more than 1,40,062 after normalization from the SNP nexus. To remove false positives, we observed variation between the controls and inclusion of tumour samples to find the heterozygous variants and exclude the low coverage sites and deletions. We found 62,781 heterozygous variants after filtering, as per CADD scores >10, and Genomic evolutionary regulatory profiling GERP scores >2 was considered for pathogenicity and mutations. We observed a total of 54 SNPs, out of which 17 SNPs are confirmed that exhibited significant association with CC.

6.7.1 Variants for similarities using venny 2.1

139

Variants are filtered out by using CADD and GERP and filtered variants are run by venny 2.1 for similarities in samples, the common most variants filtered

6.7.2 Venn Diagrams of CADD Scores

Figure: 54a. Venn diagram of CADD scores showing surrounding tissue samples S1, S2, S3, S4, S5. **54b**. Venn diagram of CADD scores showing Cervical cancer tissue samples of T1, T3, T4, T5
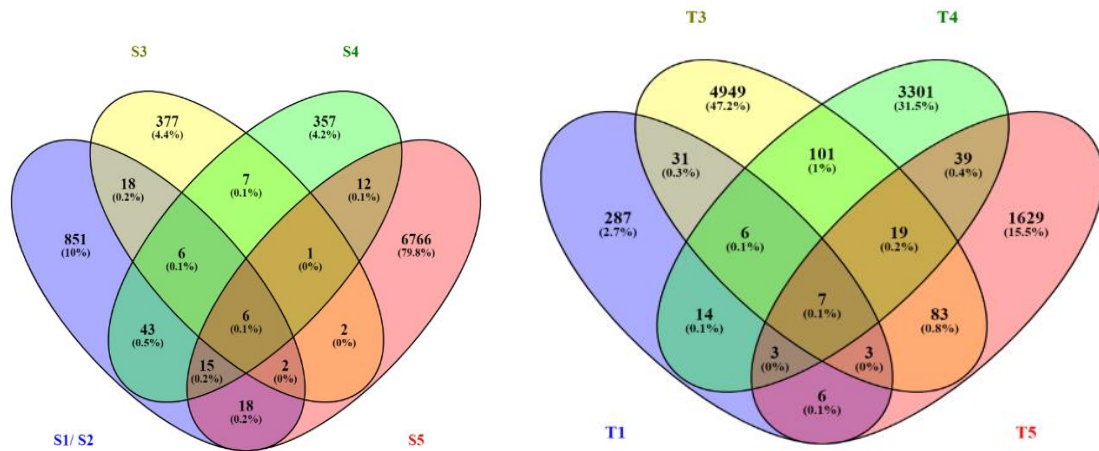
Figure 54a, 54b. Venn Diagram of CADD scores of CC tissue samples T1 & T3, d. Venn Diagram CADD scores of CC tissue samples T1, T3 & T4
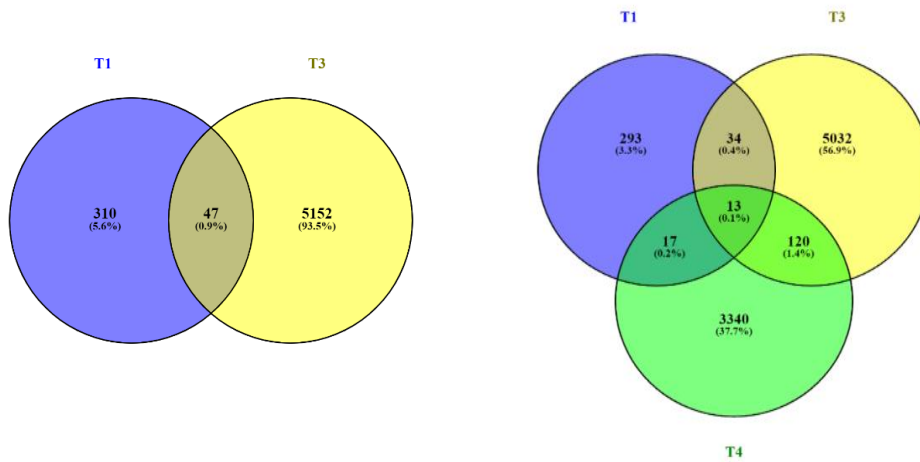
Figure 54c: Venn Diagram showing intersection of Surrounding tissue samples S1, S3, S4, S5 of GERP scores with respect to depth DP≥5 and ≤20

Figure 54d: Venn Diagram showing Intersection of CC tissue samples T1, T3, T4, T5 of GERP scores with respect to depth DP≥5 and ≤20
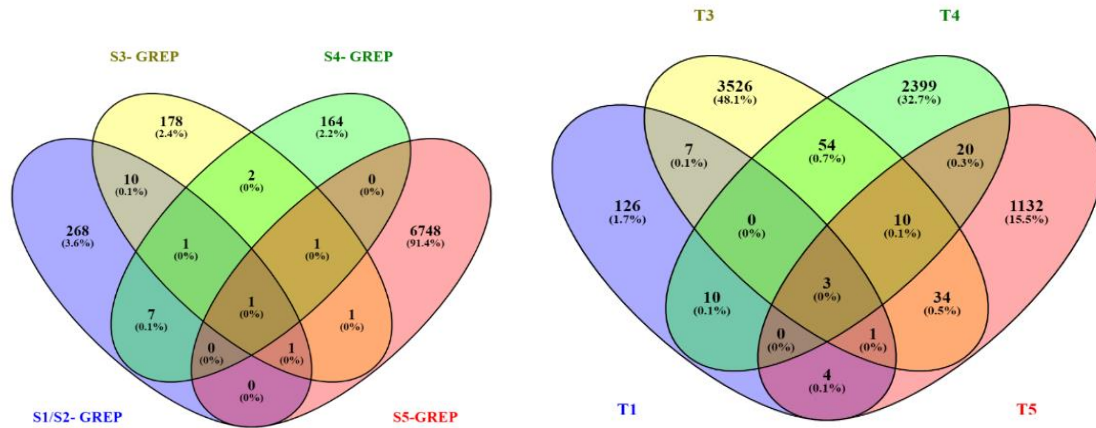
Figure 54e: Venn Diagram of T1, T3, T4 GERP scores of CC tissue samples

Figure 54f: Venn Diagram of T1, T3 GERP scores of CC tissue samples
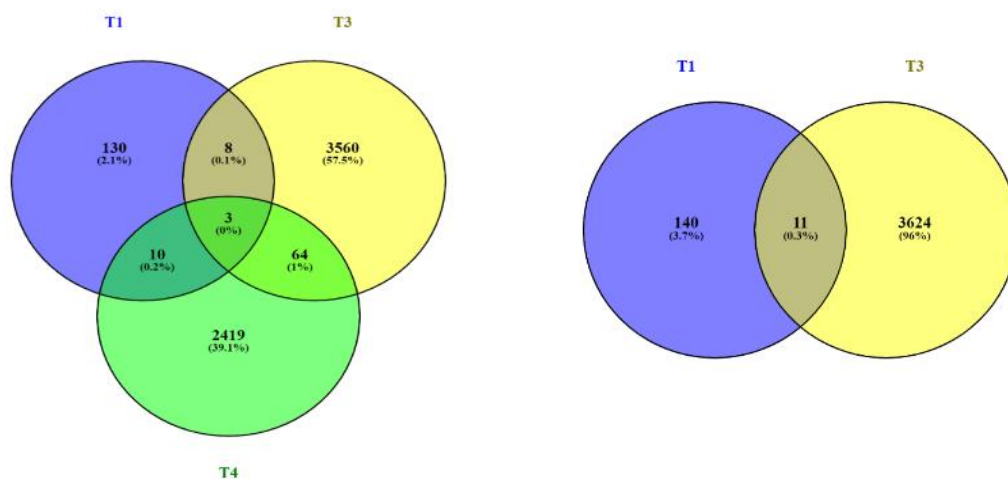
Figure: 54g. Venn Diagram of T1, T3, T4 GERP scores of CC tissue samples **54h**. Venn Diagram of T1, T3 GERP scores of CC tissue samples

There are several variants of which 15 SNPs are recognised and they are novel variants in the study. Among the variants showing mutations and pathogenicity were KMT2C, C1QTNF, EPBH1 etc shown in the given along with their minor allele frequency values **(Table 16).**

Chr. No	Position	Variation	Variant Type	SNP ID	Gene	Function Class	MAF value <0.05
7	152265083	C>A, T	SNV	138908625	KMT2C	Non coding transcript variant, Missense Variant, coding	0.001496

					sequence transcript	
7	1522650 91	T>A	SNV	62478356	KMT2C	Non-coding transcript variant, synonymous variant, upstream	0.03222 1
14	1978052 2	C>A, T	SNV	76136760 8	OR4M1	Missense Variant, Coding sequence variant	0.00001 2
16	7433490 8	G>C, T	SNV	13777396 12	PDPR2P	Non- coding transcript variant	0.00004
3	1351669 49	G>A	SNV	75459643 9	EPHB1	Missense variant, Coding sequence variant	0.00000 7
10	8900892 1	C>A	SNV	13853691 72	FAS	Non coding transcript variant, Missense variant	0.00000 08

7	1329430 85	A>C, T	SNV	74544852 2	OPCML	Missense variant, Coding sequence variant	0.00000 8
12	1636386 8	G>A	SNV	77508888 7	MGST1	Non- coding transcript variant, missense variant,3 prime UTR variant	0.00002 1
13	2432166 8	T>C	SNV	95740627 9	C1QTN F9	Coding sequence variant, Missense variant	0.00004 3
13	9678229 3	A>G	SNV	11405830 3	HS6ST3	Intron variant, Genic downstream transcript variant	0.01794 8
14	1987649 7	C>T	SNV	13419005 54	OR4K2	coding sequence variant, Missense variant	0.00001 4

17	1890222 7	G>A	SNV	99480144 3	PRPSAP 2	Intron variant	0.00012 1
17	2139950 95	C>T	SNV	14829676 92	KCNJ12	Genic upstream transcript, intron variant	0.00000 7
21	1046286 1	G>A, T	SNV	28571918	BAGE2	Non-coding transcript variant	0.00872
22	1997122 0	C>A, G	SNV	75584206 3	ARVCF	Missense variant, downstream transcript	0.00000 7
5	1754216 1	A>C, G	SNV	13809168 95	SFXN1	Missense variant, Coding sequence, ge nic downstream transcript variant	0.00000 7
7	5044692 3	C>T	SNV	76220999 8	FIGNL1	Missense variant, Coding sequence variant	0.00000 7
7	1062510 03	A>G	SNV	89680130 1	NAMPT	3 -Prime UTR variant	0.00005 1

7	152265184	C>T	SNV	770240282	KMT2C	Synonymous variant, non-coding transcript variant	0.002639

Table 16: Chromosome number, position, Variation and MAF values of Top hit Mutations

OR4M1(olfactory receptor 4 family) is not reported with CC, but according to scientists its expression is seen in ovarian cancer (Kerslake et al.2021) and also co-expression of peripheral olfactory receptors. EPHB1 is also known as Ephrin receptor B1, Eph receptors, which belong to the receptor tyrosine kinase (RTK) family, are the most extensive group of RTKs. They can be categorized into two subclasses: EphA and EphB, Eph receptors were initially recognized as regulators of axon guidance, but they are also involved in various processes, notably the genesis and advancement of cancer. FAS is a cell death receptor that leads to apoptosis in the cell causing Cervical Intraepithelial Neoplasia (CIN), In one of a recent study FAS and FASL variants showed an effect on CC prognosis (Karumazondo 2023). MGST1, known as Microsomal Glutathione S transferase 1, is mainly associated with chronic obstructive pulmonary disease (as per the GeneCards database). Still, it is also seen in cancer signaling pathways such as lung cancer, and few studies reported on cervical cancer when cell line studies performed in combination with other compounds. In one of the studies based on gene expression levels and clinicopathological studies of CC reported from the TCGA, researchers performed cox regression analysis where MGST1 didn't show any significant activity when compared to CSCC and normal cervical cells (Yu et al.2021). OPCML is called an opioid-binding protein, To, examine the methylation of CpG islands and the expression of the OPCML gene in patients with cervical carcinoma, scientists performed research and their findings indicate that the methylation of the OPCML gene promoter may have a significant impact on the development of CC. (Ye, Feng, et al. 2008). The OPCML gene could potentially be a candidate tumor suppressor gene linked with cervical carcinoma.

The clinical study of OPCML and its possible prognostic & and therapeutic uses in ovarian and breast malignancies as reported in earlier studies (Louis 2014 & Bhattacharjee 2022). SFXN1 – Sideroflexin 1, seen in several disorders, such as growth and developmental disorders, Anemia Neonatal etc. High levels of SFXN1 are associated with a negative outlook and contribute to the advancement of lung adenocarcinoma, and the association between cancer and SFXN1 shows elucidated (Chen, et al. 2022).

ARVCF Delta catenin family member, associated with Schizophrenia, and other disorders (McCrea, Pierre D.2007). It is not yet been reported in cervical cancer, and research found that the expression of ARVCF is sturdily connected with the malignant characteristics of non-small cell lung cancer (Zhang et al. 2015), also shown that it responds as a prognostic marker for prostate cancer (Penney et al.2015). Recent studies also show association and prognostic features in breast cancer (Huang et al.2023). FIGNL1 also called Fidgetin-like 1, study shows the gene is responsible for DNA damage, and loss of function in ovarian disease (Florsheim et al. 2023).

C1QTNF 9 tumor necrosis factor 9, is connected with late-onset autosomal degeneration, and according to several other studies C1QTNF 6 is seen in Head and neck squamous cell carcinoma (Huang et al 2023), and also in oral squamous cell carcinoma. But studies for C1QTNF 9 very few studies reported on other diseases and in CC report not yet reported. It is seen in systemic sclerosis-associated lung disease, whereas C1QTNF 9 shows elevated levels. (Korman, Benjamin, et al.2018). C1QTNF 9 is not yet seen in cervical cancer samples and in our study, it acts as a novel gene in CC.

In one of the clinical NGS studies, genetic landscape of mutations was identified in which dominant cancerous alterations observed in PIK3CA and dominant epigenetic suppressor alterations observed in KMT2C, KMT2D (Scholl, Suzy, et al. 2019). In our study, we focused on KMT2C, which is a lysine methyltransferase 2C gene and is associated with Myeloid or lymphoid lineage leukemia. It is one of the 10 ten mutated genes among the CC in our study. In one of the IB 1 to IV FIGO stage population studies progression-free survival rates at median follow-up by chemoradiation (87%) in CC were observed, While

they also show dominant alterations in KMT2C 16%, KMT2D 15% and PIK3CA 40% (Faundes et al.2018), It is mainly observed as a Kleefstra Syndrome-2 seen as heterozygous deletion or transversion in KMT2C gene (https://www.omim.org/entry/606833#2). KMT2C mediates the mono and trimethylation of H3 histone lysine 4 commonly called MLL 3 in WES of 3 patient samples (Figure 5a-5e) KMT2C is occupied in tumorigenesis as oncogenes and tumor suppressors in various neoplasia. It is linked to tumor sequencing studies in various cancers ascertaining lower KMT2C activity led to a deficiency in homologous recombination-mediated double-strand break DNA repair and the cells suffer from considerably higher endogenous DNA damage and genomic instability. Such cells rely heavily on PARP1/2 for DNA repair. Studies in bladder cancer have earlier reported that individuals with KM2TC variants respond better to PARP inhibitors like Olaparib than Cisplatin (Joshi et al. 2019; Mann et al. 2019).

6.8 Gene enrichment pathway analysis

The mutations encoded genes were then subjected to the interaction network analysis with all 17 genes **(Figure 55)** showing co-expression (75.79%) physical interactions (14.49%), Co-localisation (8.26%) and genetic association (1.46%) from GeneMania. Among them, associations show the majority of the interactions with uterus, CC, endometrial, oral, leukemia. The genes from GeneMania interactions show integrated biological pathway associations.
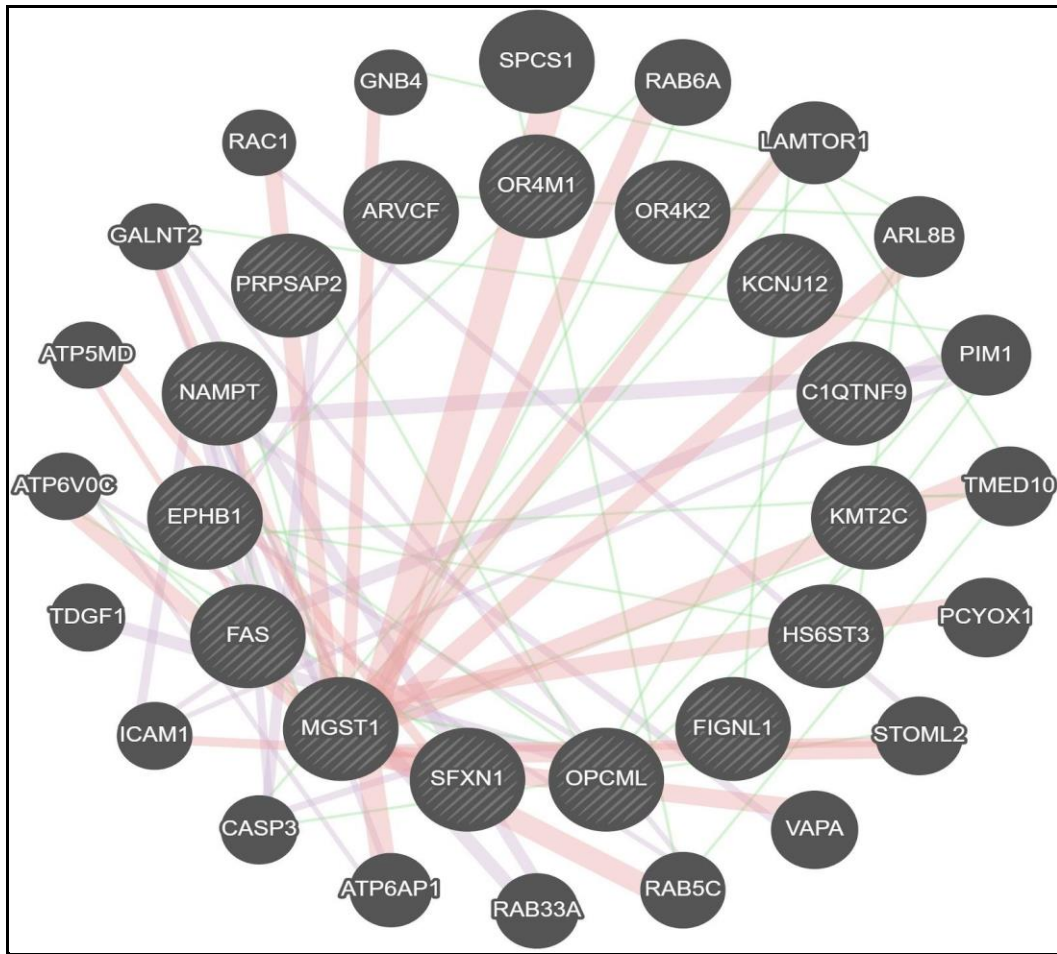
Figure 55: Interactions Network of exome variants in Gene Mania

6.9 Interaction network of KMT2C

The interaction network of KMT2C (also known as MML3) (Figure 56) includes Physical interactions- 77.64%, Co-expression- 8.01%, Genetic interaction: 2.87% from GeneMania. KMT2C is a chromatin remodeling gene, where histone acetylation and modulation of genes happen, KMT2C along with its lysine methyltransferase subfamily KMT2D, A, E cause leukemia or Mixed lineage leukemia. The other genes that interact with KMT2C include WDR5 causes Kabuki syndrome and its pathways PKN1 stimulates transcription factor androgen receptors the same as KMT2C while ASCL2 is mostly seen in breast cancer cell lines and frequently most of the genes are involved in transcription factors.
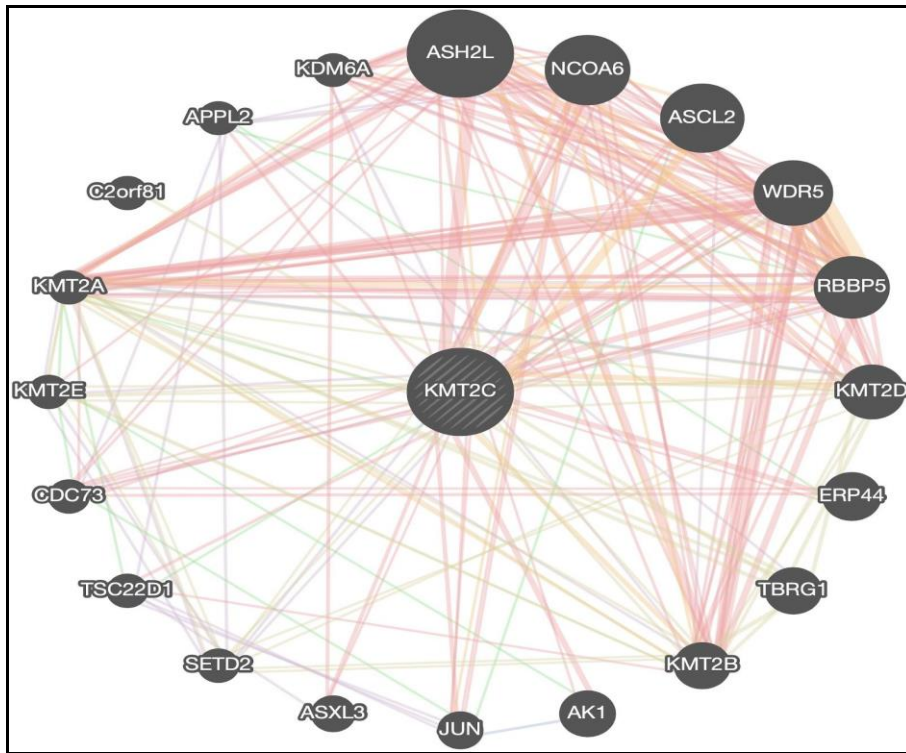
Figure 56: The interaction network of KMT2C using Gene Mania

6.10 Integrative Genome Visualisation (IGV)

In the IGV browser, there is a change in the allele of KMT2C and shown variation in the allele frequency. Control sample BAGE variant is found and in the tissue samples KMT2C pathogenic variants found**.** The IGV browser results of KMT2C are the most common SNV, chromosome 7:152265091, T>A, SNV 62478356, showing Novel Mutation and Pathogenic**. (Figures 57 A to I)**
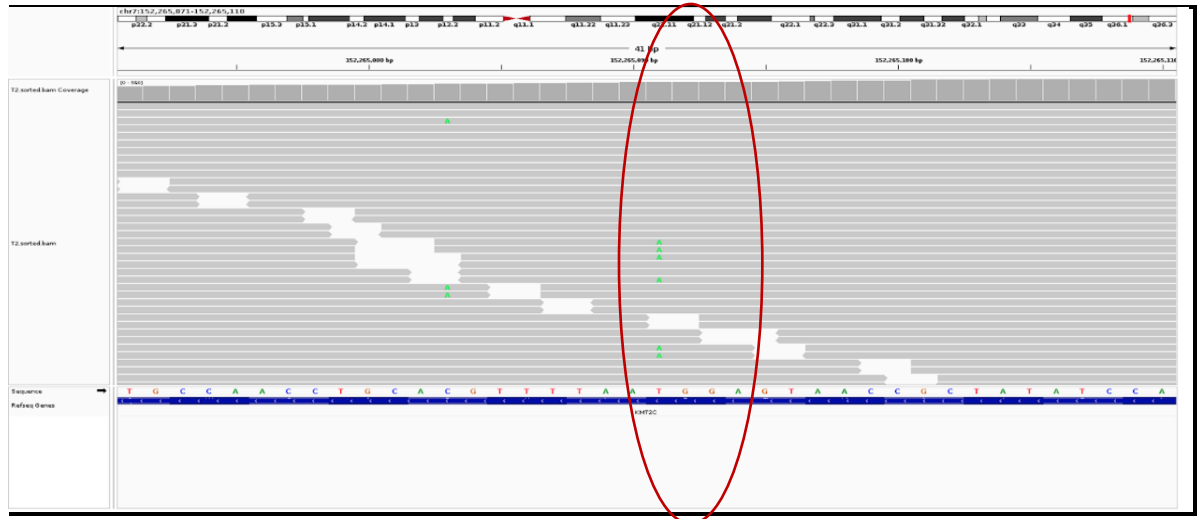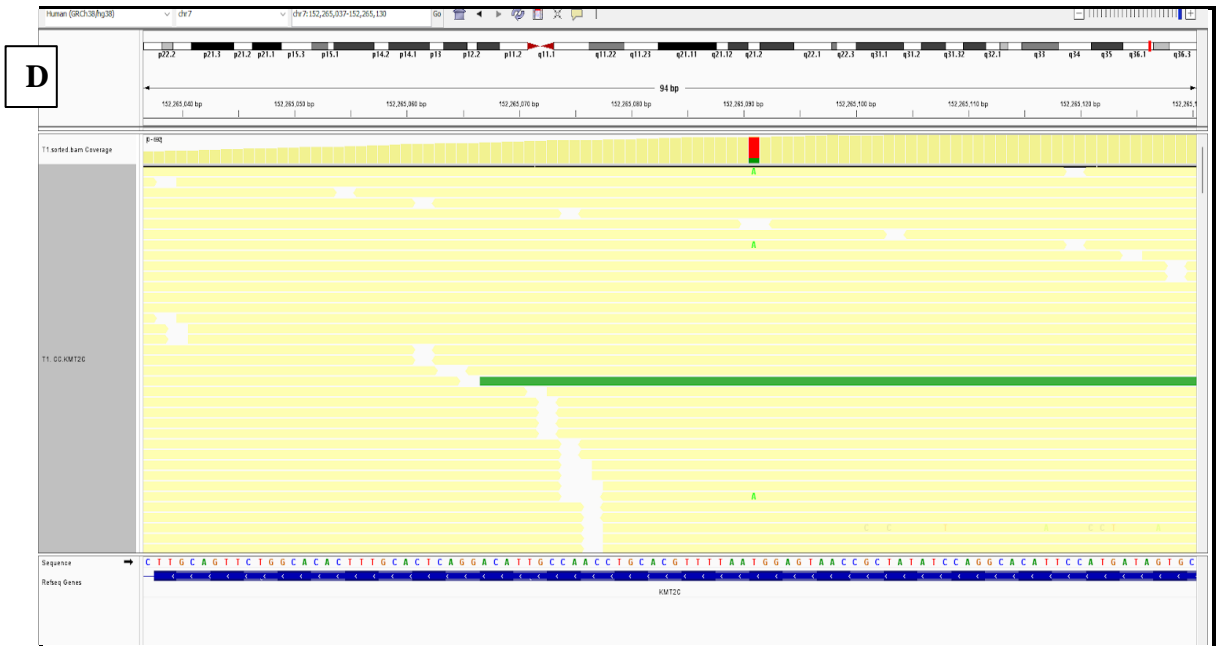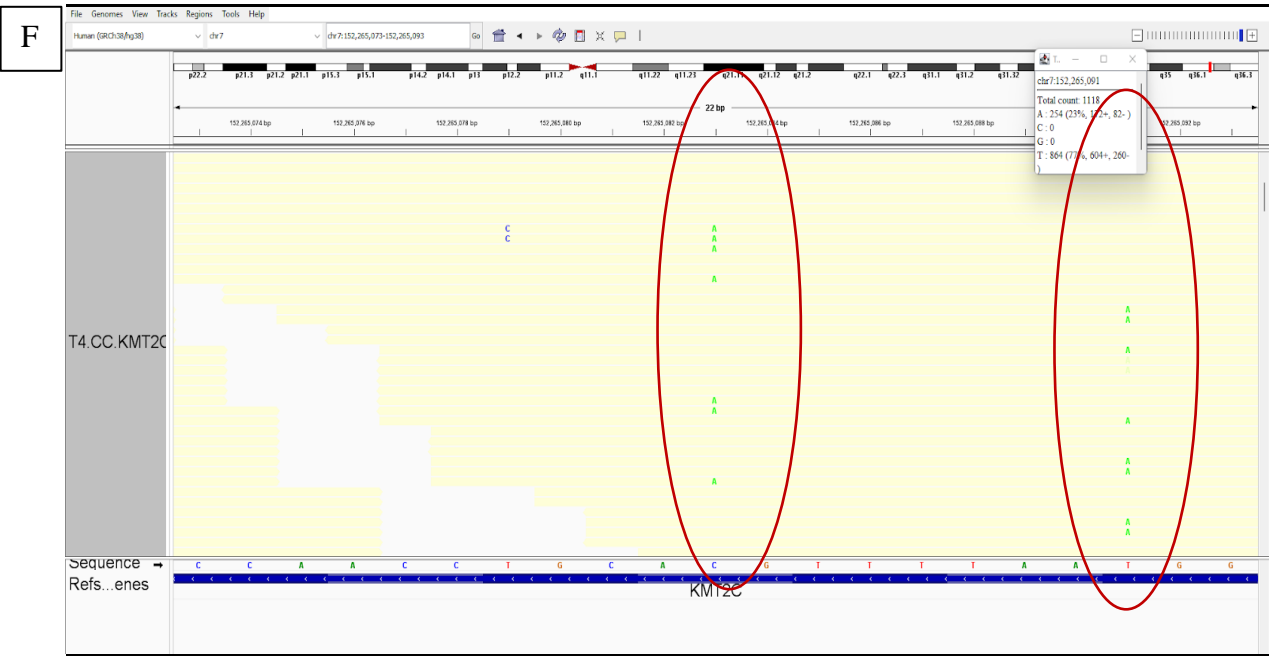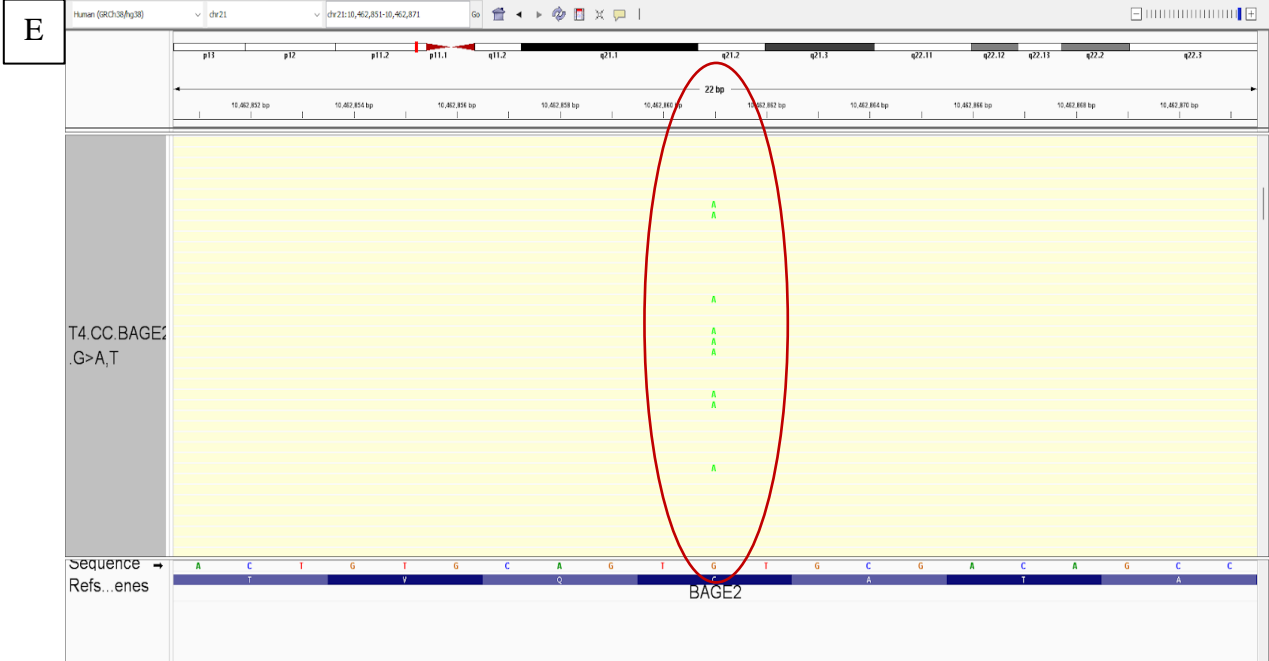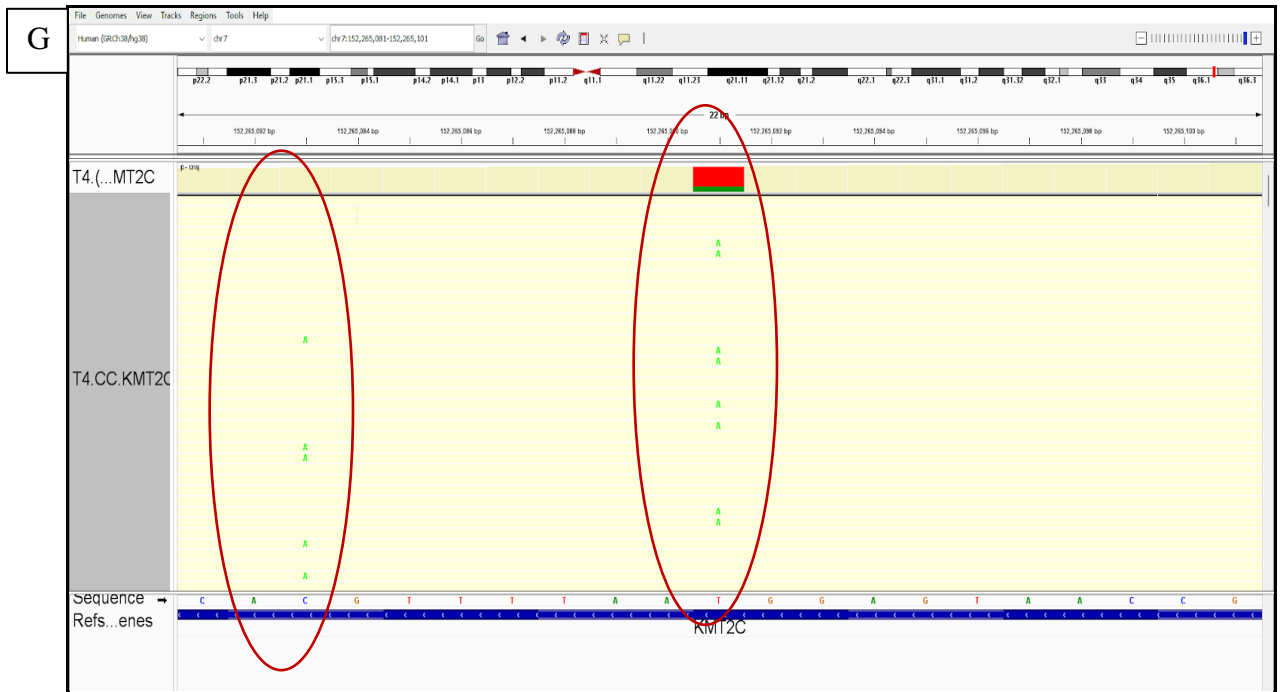
Figures 57 A: T1.CC.KMT2C Mutation. 57 B. IGV browser results of KMT2C of T1 CC tissue sample Chr7:152265083 C>A, T 57 C. KMT2C mutation of T5 CC tissue sample Chr: 7:152265083. 57 D. KMT2C mutation of T4 CC tissue sample Chr7:152265083 C>A, T. 57 E. IGV Browser results of BAGE2 of T4 CC tissue sample Chr21:10862861, G>A, T. 57 F. KMT2C mutation of T4 CC tissue sample Chr7:152265083 C>A, T. 57 G. KMT2C mutation of T4 CC tissue sample Chr7:152265091 T>A

6.11 Single Nucleotide Conformation Polymorphism (SSCP)

SSCP has been performed after the whole exome sequencing of variants to identify and validate the mutations with the help of bands (**Figure 58**). In this present investigation, to validate the samples for KMT2C mutations we have taken 64 samples of which 15 samples control and other are tumor samples of grade 2,3 and 4 for comparison and validation for the variants.By comparing electrophoretic mobility, SSCP analysis can identify sequence variants, such as single-point mutations and modifications. When DNA is exposed to non-

154

denaturing or partially breaking down environments, there is a noticeable difference in mobility between wild type DNA and DNA with sequence mutations, even if they are only one base pair long.13% acrylamide gel is used to check the mutations in KMT2C and C1QTNF 9, we got the few wild and few are homozygous in the samples.
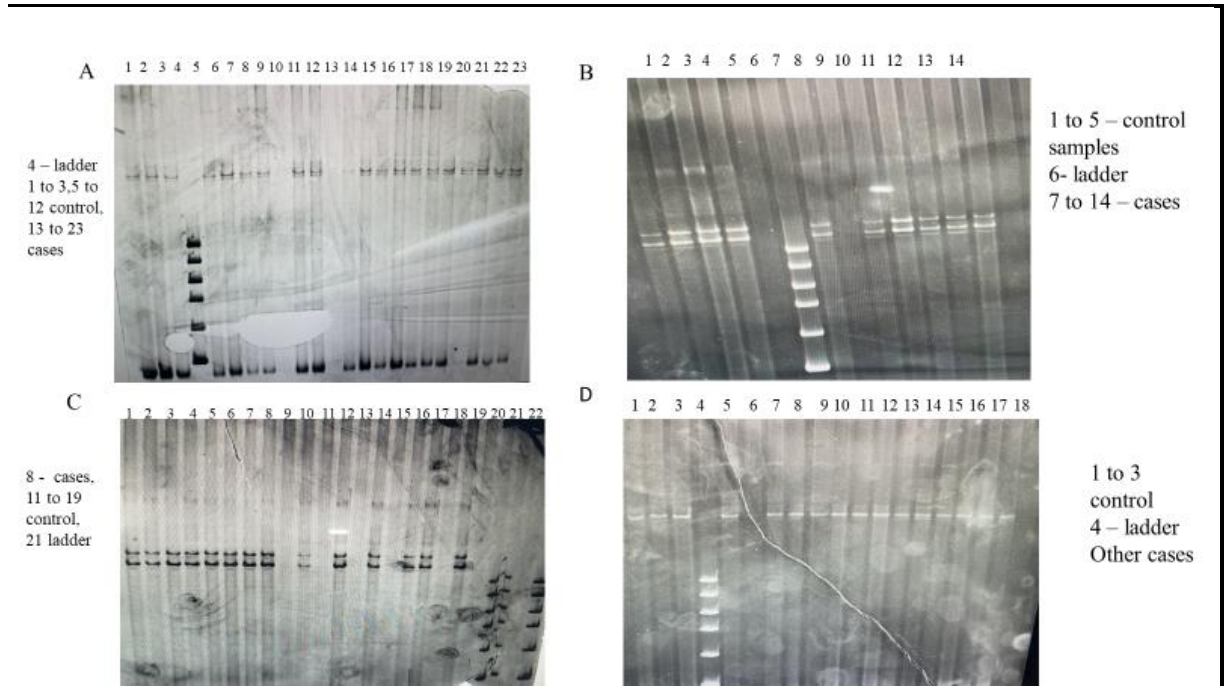
Figure 58: SSCP Polyacrylamide gel electrophoresis of KMT2C and C1QTNF 9

6.12 Sanger Sequencing of KMT2C

In our study, NGS data of sanger sequencing shows false positive and negative results and reported discrepancy results between NGS whole exome sequencing and sanger sequencing in 10 samples and control 5 samples. NGS sequencing with Discrepancies in sanger sequencing results show that the process successfully produced an electropherogram that displayed the variation in its homozygous form (**Figure 59A, 59B**). We assessed the primer again utilized for the first amplification to determine their capability to bind to SNVs, and the outcome was negative. Consequently, we examined the experimental procedures to definitively determine the stage at which the loss of alleles

occurred and we have also performed SSCP (Single Strand Conformation Polymorphism) for identifying mutation using 14% SSCP gel (non-denaturing Polyacrylamide Gel) When observed on 13% polyacrylamide gel, Silver staining method is a straightforward, expeditious, and economical technique that can be regularly executed in clinical laboratories. So, we have used silver staining to see the clear demarcation of band patterns, resulting in two distinct band patterns that indicate the presence of the mutation in the wild or homozygous form. (**Fig: 59C, D)**

KMT2C: **rs62478356 T>A ; rs138908625 C>A**

Figure (A, B): The software used to analyse the results is Finch T.V

Figure (C, D): The software used to analyse the results is Finch T.V
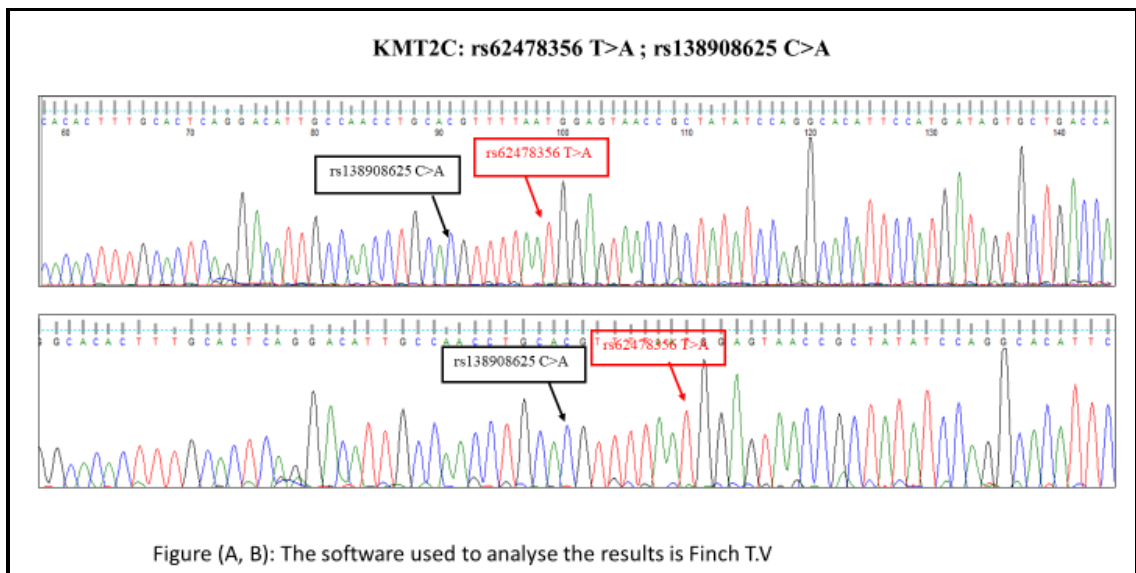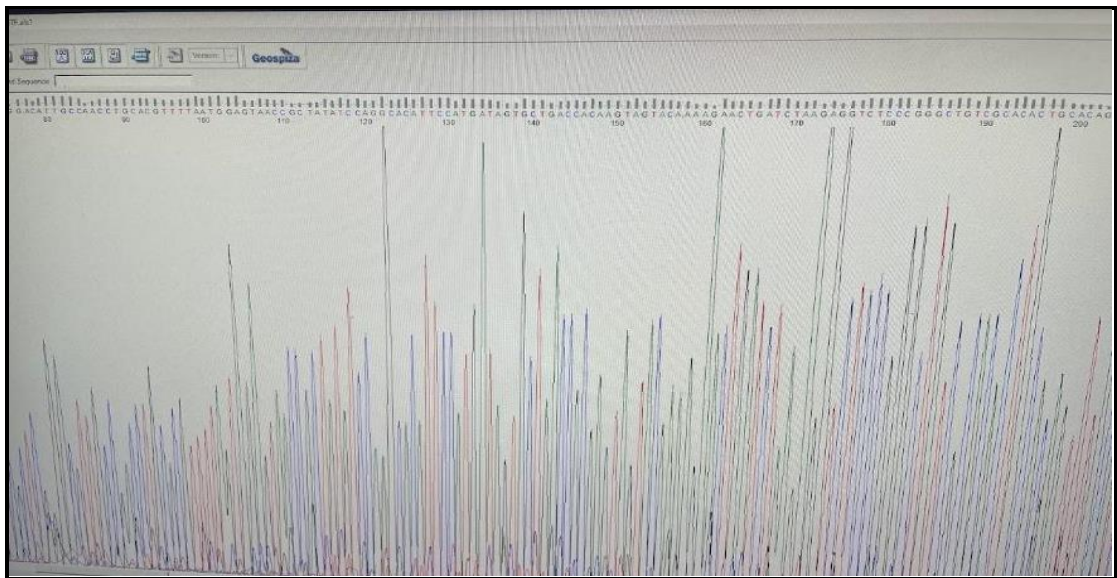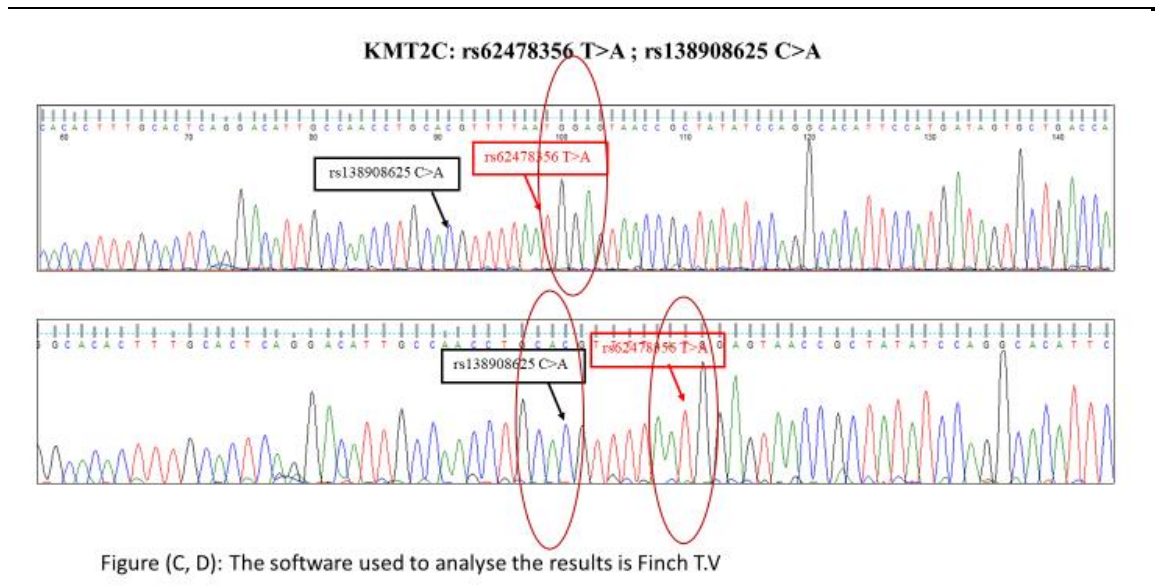
Figure 59: A, B, C, D: Sanger Sequencing results of KMT2C rs 62478356 T>A, rs138908625 C>A

After the WES of the variants, we validated the data by performing SSCP, Sanger sequencing, and transcriptomics data. PCR -SSCP is a simple way to identify mutations by using Acrylamide gel and silver staining methods, However, we have seen some homozygous and wild mutations.

To Confirm again the experiment have set for sanger validation in which the study found the mutations with discrepancy wild or homozygous. Hence, there, might be several reasons for the discrepancy, including non-primer binding sites of SNVs, Additionally, there are sections of the genome that are challenging to amplify and may not produce satisfactory results, Sanger continues to produce the same outcomes despite repeated tries, which could lead to Sanger testing on an excellent NGS variant, rather than a sufficient backing, however, errors can still occur in NGS, based on other findings (De Cario, et al. 2020) (**Figure 59A to D).**

6.13 COSMIC and cBioportal Database

Through the Genome Browser COSMIC database, the SNVs having rs ids were annotated and 138908625, 62478356 respectively **(Figure 60).** Cosmic database shows KMT2C mutations which were missense, frameshift, deletions and substitutions. The WES SNVs of CC were compared with data from PAN cancer studies of 278 samples of CSCC from the cBioportal database, KMT2C plot of overall survival against the probability of overall survival that denotes the altered and unaltered groups (**Figure 61**).
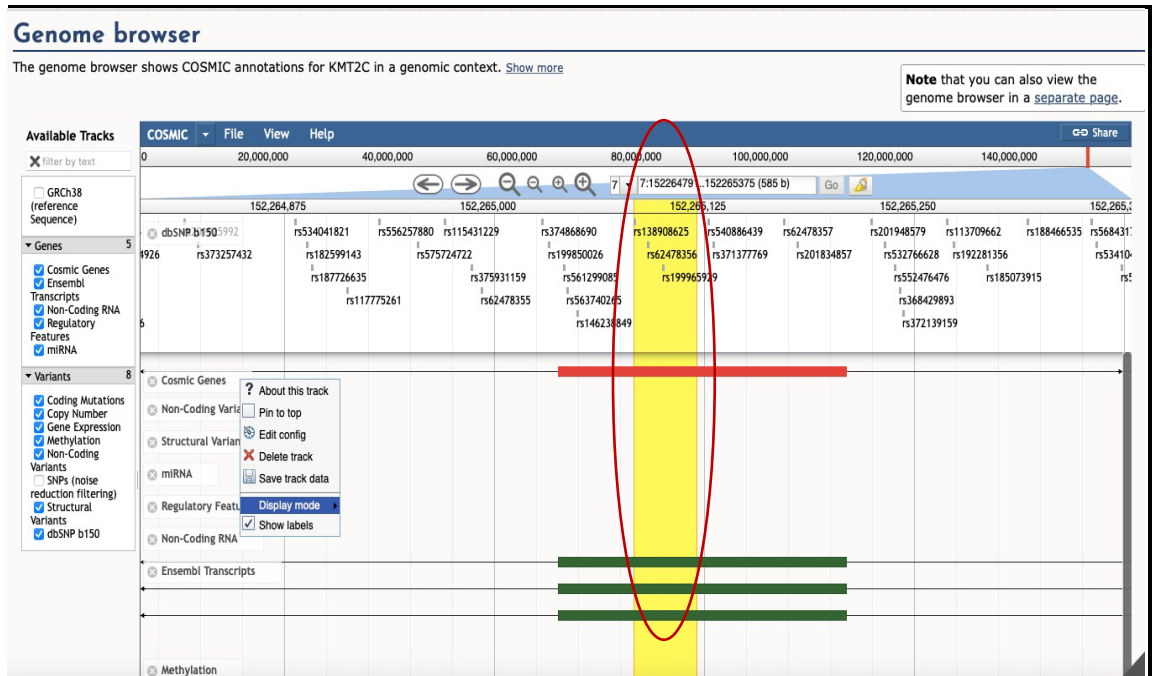
Figure 60: KMT2C rs ID 138908625, 624 78356 in COSMIC database
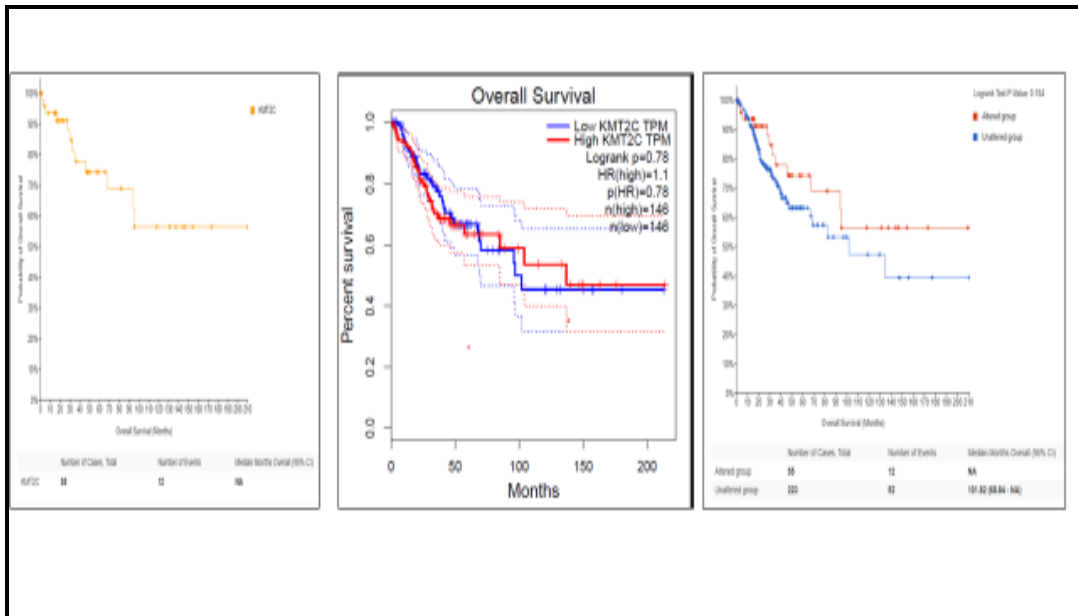
Figure 61: KMT2C survival plot of disease-specific and disease-free from the cBioportal database (cbioportal.org)

6.14 Transcriptome data analysis of KMT2C gene expression

(RNA-Seq) has become most influential tool for analyzing cancer transcriptomes, detecting items such as alternative splicing and novel transcripts with greater precision and efficiency as next generation sequencing technologies have evolved (Ozsolak, 2011 & Wang, et al. 2009). RNA-Seq is broadly used in the analysis of transcriptome profiles in various cancers (Stephens et al 2012 & Wu, et al. 2012). A heatmap was developed based on the transcriptomic analysis of cervical cancer tissues. Affymetrix's heatmap in Clariom D analysis reveals differentially expressed genes/mRNAs in CC tissues. KMT2C expression found to be having lower ($p < 0.05$) in CC tissue than in normal tissue, as shown in **Figure 62**. Cancer cells acquire Cisplatin resistance through the DNA repair mechanism; hence, cancer cells with low KMT2C expression are attractive targets for therapies with PARP1/2 inhibitors (Mann, Minakshi. et al. 2019). Therefore, studies focused on how KMT2C mutations alter gene expression and ways to target these mutations therapeutically are very important in the current scenario.
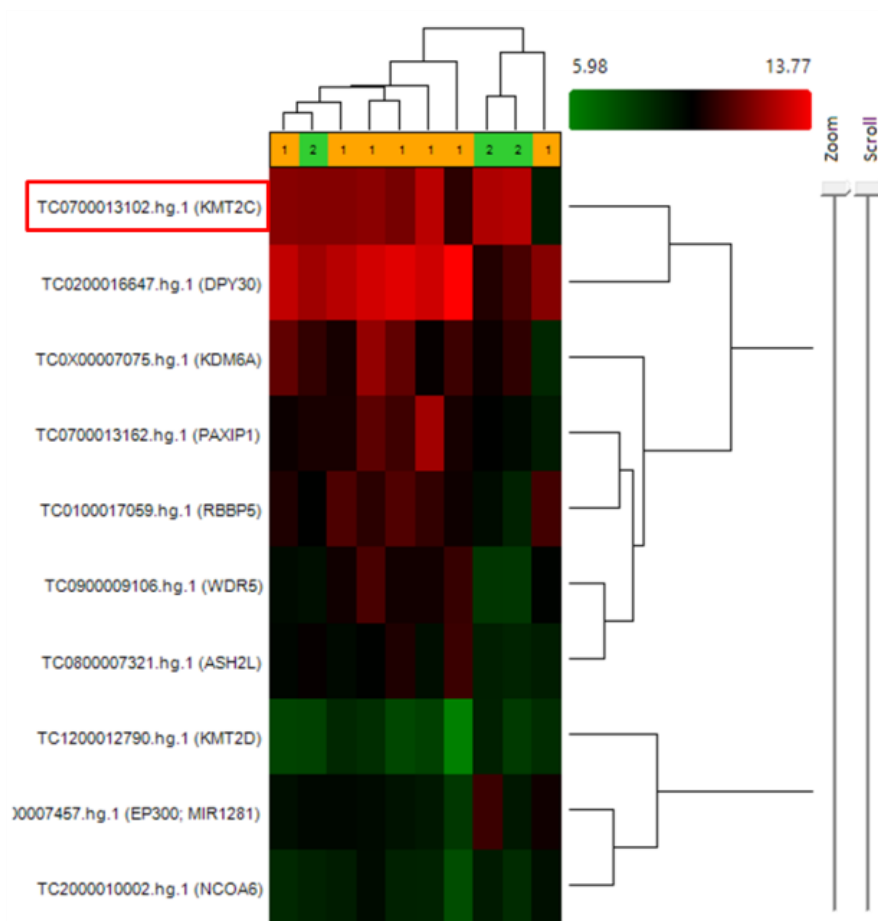
Figure 62. Heat map illustrating the expression levels of KMT2C in tumor samples against non-tumor samples. Red color represents genes that are up-regulated (fold change > 2) and green color represents genes that are down-regulated (fold change < 2) in tumor samples (p-value < 0.05).

6.15 Validation of KMT2C gene in several other DBs

The KMT2C gene was validated in normal and cervical tissue using the NCBI-GEO and GEPIA databases (DBs). Three GEO datasets, GSE6791, GSE29570, GSE63514, GSE67522, and GSE138080, were downloaded from the (GEO) database and analyzed using Affymetrix's Expression console and TAC software. GSE6791 has 20 cancerous and 8 normal cervical tissues, GSE29570 has 45 cancerous and 17 normal cervical tissues, GSE63514 has 28 CC and 23 normal, GSE67522 has 20 cancerous and 22 normal cervical

tissues, and GSE138080 has 10 CC and 10 normal samples. **Table 17** depicts the KMT2C gene expression profile derived from these data sets.

Dataset	Gene Symbol	Log FC	ANNOVA P-Value
GSE127265	KMT2C	-1.61	0.2408
GSE6791	KMT2C	1.01225239	4.24E-05
GSE29570	KMT2C	-2.71E-01	1.48E-02
GSE63514	KMT2C	-0.00962845	3.87E-01
GSE67522	KMT2C	-1.25E-01	2.32E-01
GSE138080	KMT2C	-0.57377958	2.06E-03

Table 17: KMT2C gene expression values in different datasets

The gene KMT2C in various GSE datasets predicts that log fold change and P value less than 0.05 which means it is significant.

Datasets		KMT 2C	KMT 2D	ASH 2L	NCO A6	PAXI P1	EP3 00	DPY 30	RBB P5	KDM 6A	WD R5
						Gene names					
GSE12 7265	log FC	-1.61	-1.03	1.71	1.07	1.42	-1.19	3.72	2.28	1.15	3.78
	P-Val ue	0.24	0.48	0.20	0.99	0.30	0.30	0.01	0.00	0.76	0.00
GSE67 91	log FC	1.01	-0.77	0.45	-0.24	2.14	0.57	1.60	0.69	0.57	0.17
	P-Val ue	0.00	0.00	0.01	0.06	0.00	0.01	0.00	0.01	0.02	0.32
GSE29 570	log FC	-0.27	0.17	0.13	0.24	0.39	0.05	0.11	0.27	0.14	0.62

	P-Value	0.01	0.03	0.17	0.00	0.00	0.47	0.44	0.00	0.29	0.00
GSE63514	log FC	-0.01	-0.29	0.36	0.00	0.99	0.71	0.40	0.08	-0.04	0.94
	P-Value	0.39	0.00	0.07	0.96	0.00	0.03	0.22	0.71	0.93	0.00
GSE67522	log FC	-0.13	0.03	0.19	0.58	0.53	-0.40	0.16	0.06	0.21	0.34
	P-Value	0.23	0.24	0.09	0.00	0.00	0.11	0.15	0.76	0.39	0.12
GSE138080	log FC	-0.57	-0.29	0.16	0.28	0.19	-0.09	0.16	0.08	0.18	-0.01
	P-Value	0.00	0.04	0.33	0.23	0.33	0.39	0.36	0.39	0.45	0.97

Table 18: GEO datasets of five sample sets to validate the KMT2C gene, and their p value, log FC value.

Understanding the molecular mechanisms of CC may be aided by the documentation of the dysregulated genes. This study used transcriptome-based gene expression analysis, and the findings showed that compared to normal cervix controls, the expression of about 4560 dysregulated genes increased or decreased by more than 2-fold in CSCC tissue (SD Annapurna, et al 2021; Deepti Pasumarthi, et al.2022). In our analysis, KMT2C was downregulated in CC tissue compared to normal tissues. The use of five GEO datasets i.e. GSE6791, GSE29570, GSE63514, GSE67522, and GSE138080 to validate the KMT2C expression in cervical cancer were illustrated in the given **(Table 18).**

6.16 KMT2C expression levels beyond cancers

In the present study the GEPIA database is used to conduct a comparative analysis of KMT2C mRNA expression levels in cancerous tissues versus normal tissues. **Figure 63**

(A & B) illustrates the presence of KMT2C in the proximity of cancerous cells. The expression of KMT2C was observed to be downregulated in various cancer types, such as Bladder, Cervical, Colon, Lung, Ovary, Rectum, and Uterine, compared to control. The expression of KMT2C mRNA, however, was dramatically increased in the brain, pancreas, and stomach. In addition, we employed TIMER 2.0 to evaluate the expression of KMT2C in TCGA. **Figure 63 (C)** displays the specifics of KMT2C expression in several types of cancer.
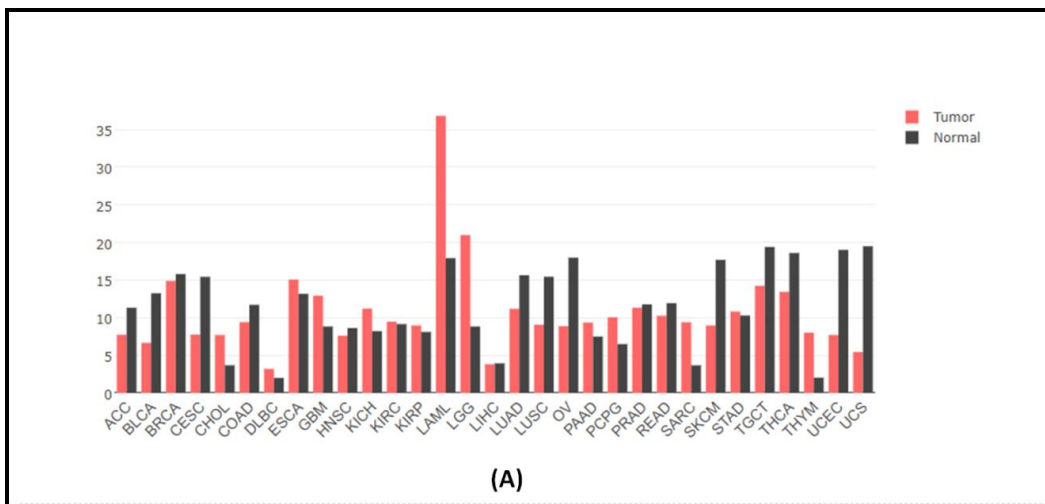
Figure 63: KMT2C proximity with other cancers
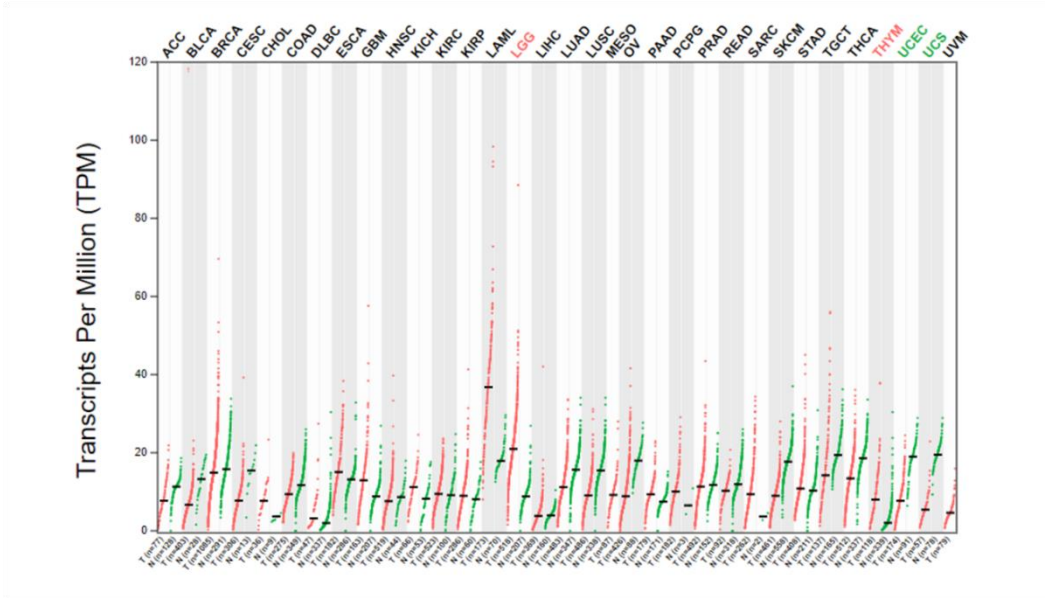
(B)
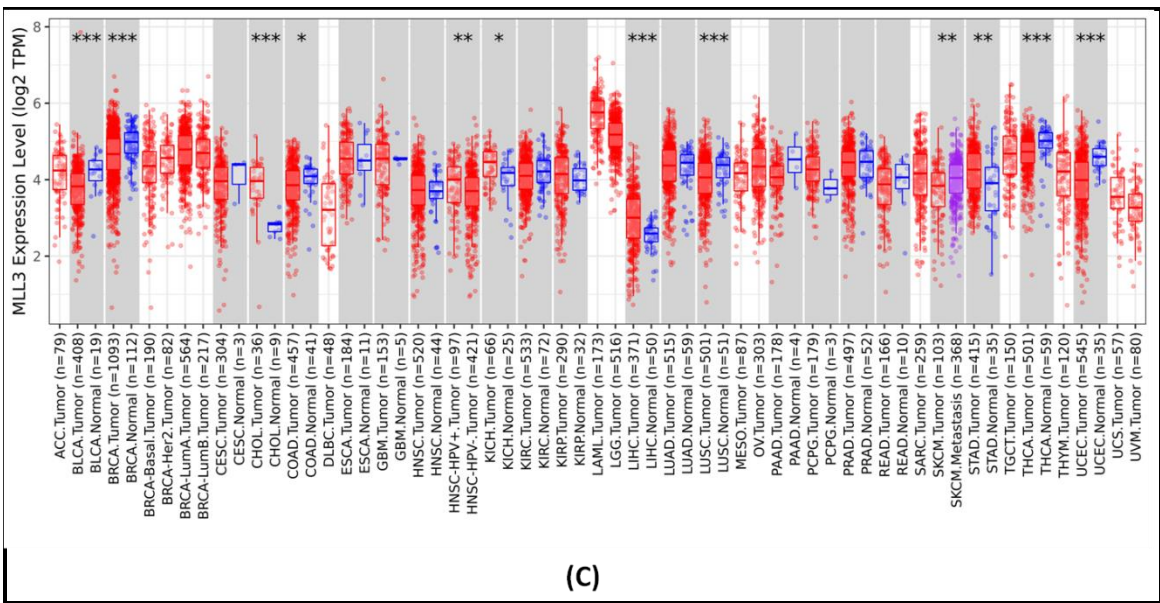
(C)

Figure 63: The levels of KMT2C expression across tumors

63(A) The KMT2C gene expression profile across all tumor samples and associated normal tissues (Bar Plot) is created using the GEPIA database. Figure 63 (B) has the

GEPIA algorithm was used to investigate the levels of human KMT2C expression in various tumor types that were taken from the TCGA database. It is the median expression of a tumor type or normal tissue that is represented by the height of the bar. 63(C) The levels of KMT2C expression across tumors, as determined by the TCGA data in TIMER2.0, are statistically substantial at $p<0.01$ and $p<0.001$.

6.17 Expression levels analysis using GEPIA Dataset

The box plot depicted in **Figure 64 (A**) illustrates the disparity in KMT2C expression between CC and normal. The data for this graphic is sourced from the GEPIA database. The expression of KMT2C was significantly modified in CC tissues. In the present study the Kaplan-Meier Plotter was analyzed to perform a comprehensive study of KMT2C in CC. The decreased expression of KMT2C mRNA did not exhibit a substantial association with the overall survival (OS) of all patients diagnosed with CC (**Figure.64 (B**). In contrast, there was a notable correlation between extended relapse-free survival (RFS) and decreased levels of KMT2C mRNA ($p<0.05$) (**Figure 64 (C)).**

CESC
(num(T)=306; num(N)=13)

(A)
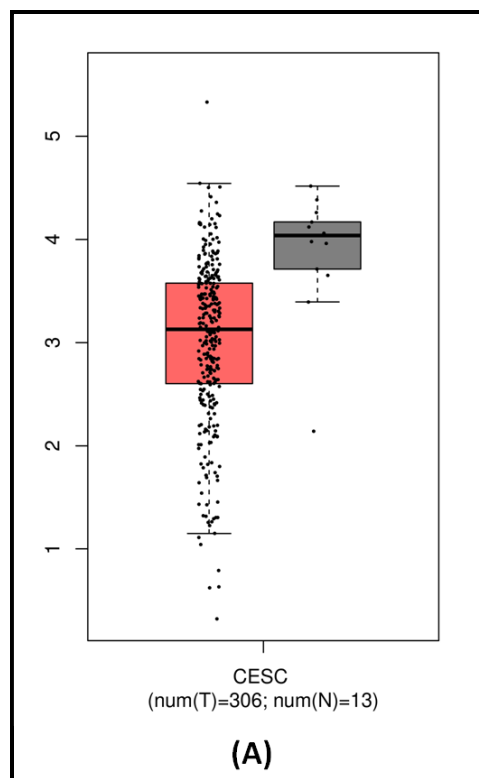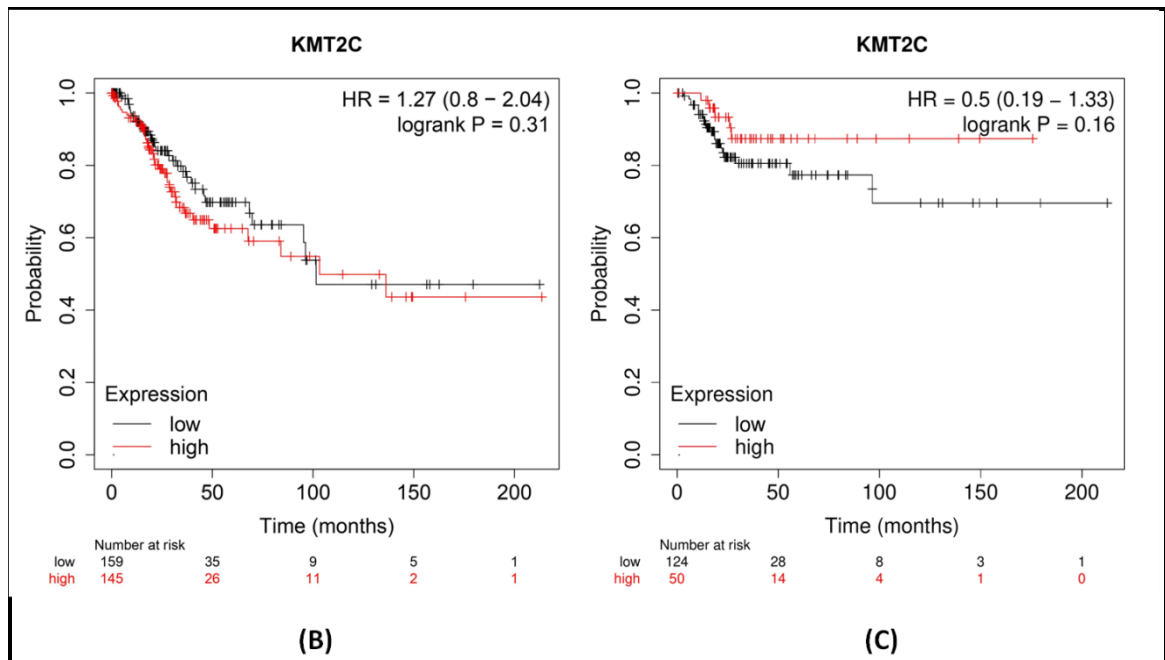
Figure 64. The expression of KMT2C in cervical cancer. Figure 64(A) GEPIA box plot. Kaplan-Meier plotter analysis of KMT2C gene in cervical cancer Figure 64 (B) OS (overall survival) and Figure 64 (C) RFS (relapse-free survival)

6.18 Immune Infiltration and KMT2C Expression Correlations in CC

The presence of resistant cells into the cancer microenvironment can knowingly affect the survival of patients. In the investigated study the connection between KMT2C and the invasion of immune resistant cells in CC utilizing the TIMER2.0 database. The results demonstrated a strong correlation between KMT2C and the purity of cervical cancer (R = 0.043, p = 4.80e − 01). The expression of KMT2C showed a helpful correlation in CD8+ T cells (R = 0.069, p = 2.52e − 01), B cells (R = 0.007, p = 9.08e − 01), and macrophages (R = 0.15, p = 1.27e − 02) in CC. Conversely, the expression of KMT2C exhibited a significant negative correlation with the infiltration level of CD4+ T cells (R = -0.067, p = 2.63e − 01), neutrophils (R = -0.016, p = 7.94e − 01), and dendritic cells (R = -0.087, p = 1.47e − 01) in the same context **(Figure 65).**
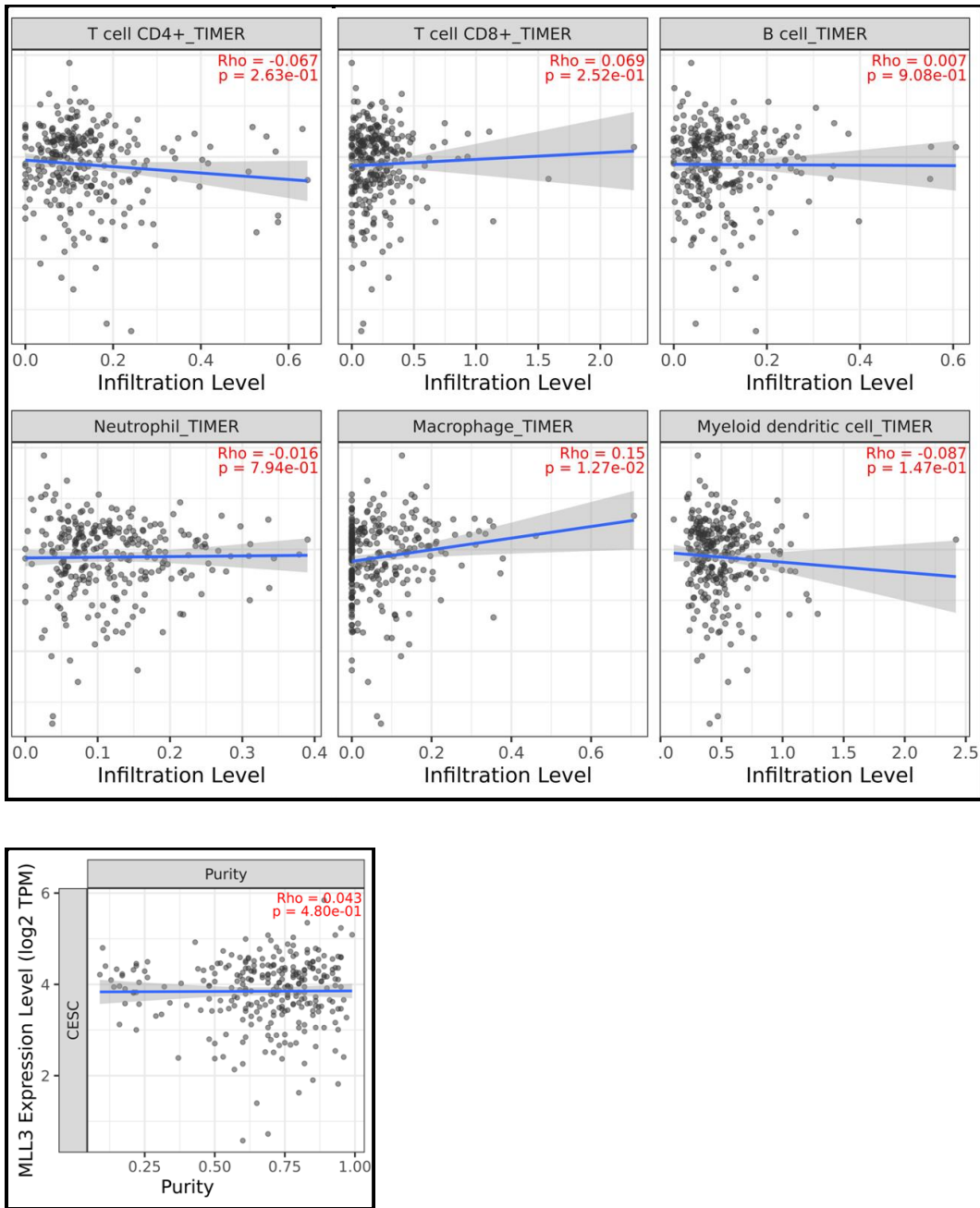
Figure 65: This study investigates the connection between the expression of KMT2C and the degree of immune infiltration. In cervical cancer, there is a correlation between the expression of KMT2C and the purity of the tumor, as well as the infiltration levels of CD4+ T cells, CD8+ T cells, B cells, neutrophils, macrophages, and dendritic cells. In order to be considered statistically significant, a p-value that is lower than 0.05

6.19 KMT2C connected pathways and PPI Interactions

According to the findings of the KEGG pathway, the KMT2C gene, which is involved in major signaling pathways such as Lysine degradation (hsa00310) and metabolic pathways (hs01100), plays an important role. In the study, the STRING online database tool was utilized to regulate the top ten proteins that are associated with the KMT2C gene. The following proteins were found to be associated with the KMT2C gene: PA1 (PAXIP1 associated glutamate rich protein 1), KMT2D (Lysine Methyltransferase 2D), ASH2L (Set1/Ash2 histone methyltransferase complex subunit), NCOA6 (Nuclear receptor coactivator 6), EP300 (Histone acetyltransferase p300), DPY30, PAXIP1, RBBP5 (Retinoblastoma-binding protein 5), KDM6A (Lysine Demethylase 6A), and WDR5 (WD repeat-containing protein 5). This figure illustrates the results of the STRING and KEGG analysis of KMT2C **(Figure 66).**
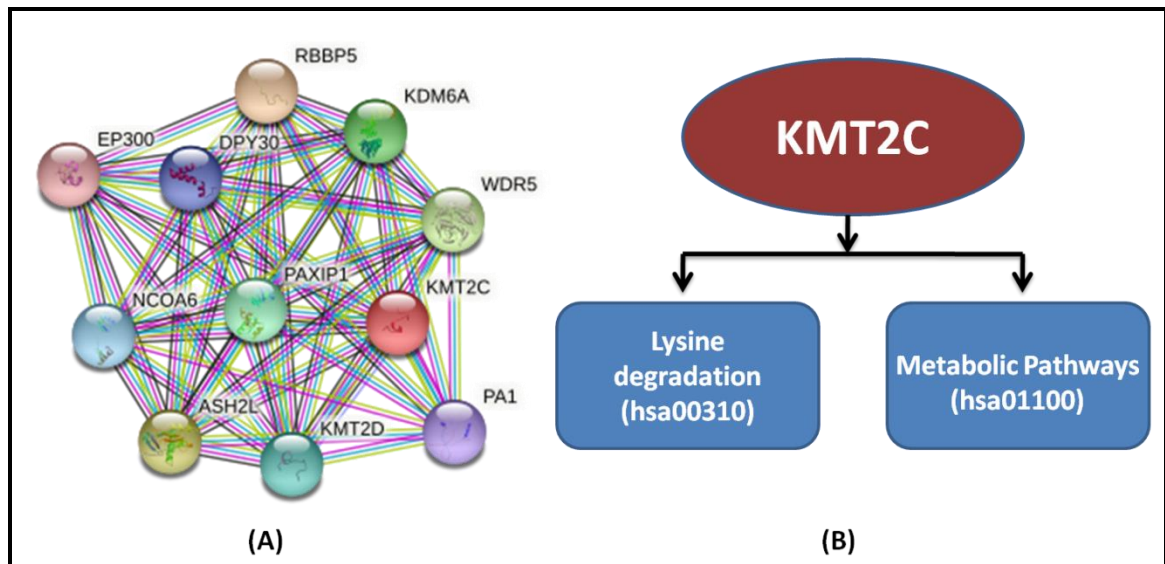
Figure 66: Analysis of the KMT2C gene reveals its association with STRING networks and KEGG pathways. The online tool visually represents interacting nodes as colored circles (B) KEGG pathway analysis elucidates the functional pathways associated with the KMT2C gene.

6.20 Transcription factor binding site analysis

The KMT2C gene exon variant T>A (SNP ID: rs 62478356) is more common in the group of individuals with the illness. An in-silico analysis performed using the alibaba TRANSFAC tool reveals that this variant is linked to an alteration in the MEB-1 transcription factor binding site, leading to the decrease in activity of the KMT2C gene. The KMT2C gene exon variation C>T (SNP ID: rs 62478357) is more common in the group of individuals with the illness. An in-silico analysis performed using the alibaba TRANSFAC tool reveals that this variant is linked to a change in the SP-1 transcription factor binding region, leading to the decreased expression of the KMT2C gene (**Figure 67 and 68**).
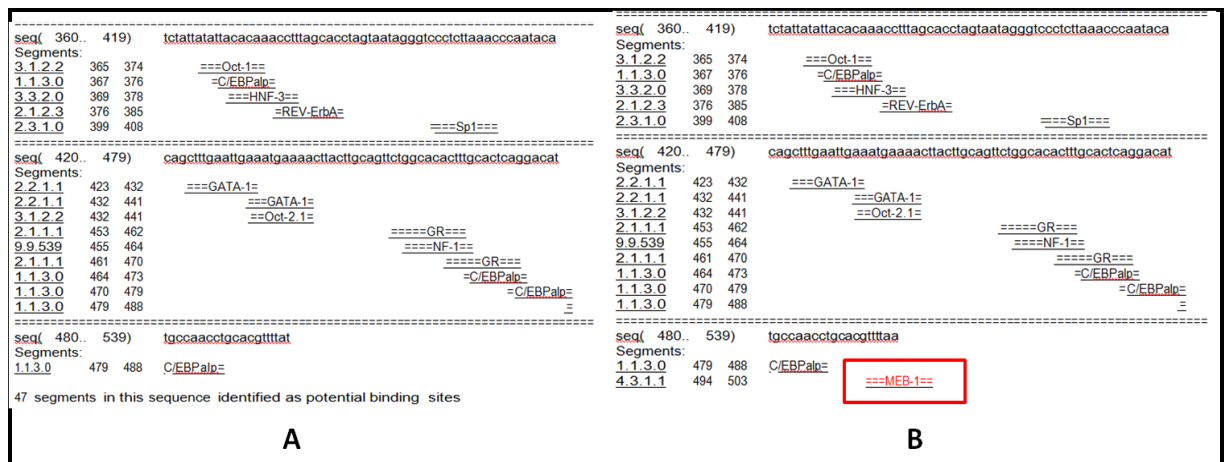
Figure 67: Transcription Factor binding (TFB) site prediction at rs 62478356 (T>A) in the KMT2C exon region (A) Wild type Sequence (B) Target exon sequence.
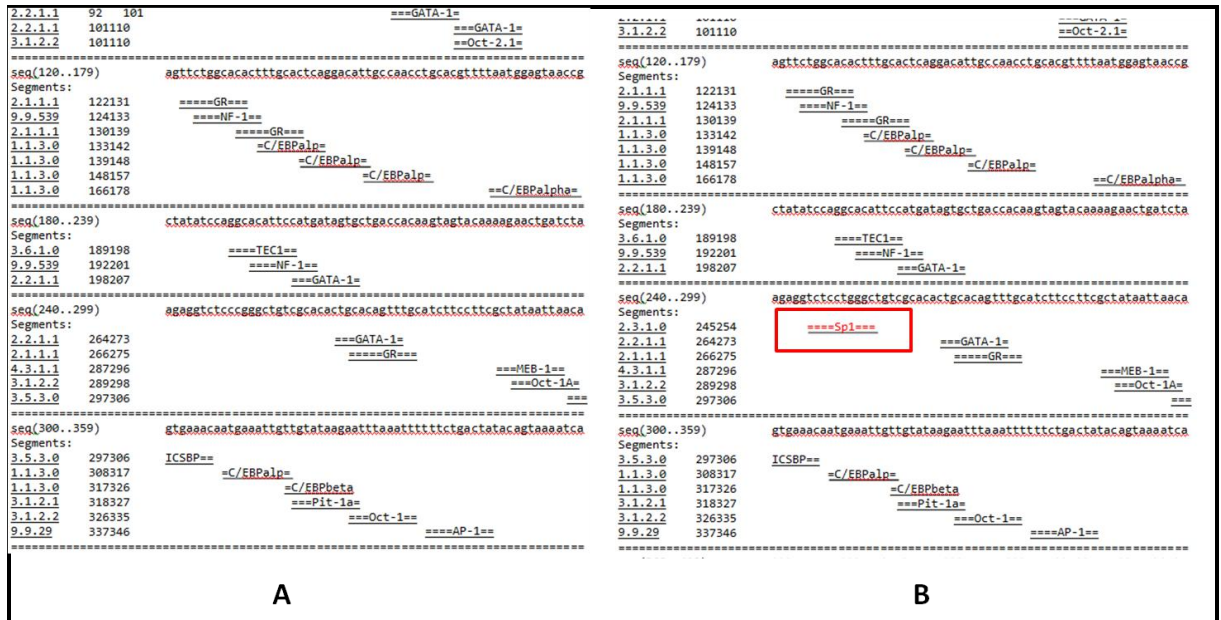
Figure 68: TFB site prediction at rs 62478357 (C>T) in the KMT2C exon region (A) Wild type Sequence (B) Target exon sequence

6.21 Hallmarks of cancer in KMT2C

Cancer mutation of KMT2C among all cancers is promoted in proliferative signaling pathways and suppression of growth in the cancers. And it suppresses cell replicative immortality, invasion and metastasis and genome instability and mutations. KMT2C mutation types show Nonsense substitution 966, Missense substitution 3034, Synonymous substitution 760, frameshift insertion 217 etc., among all the cancers.
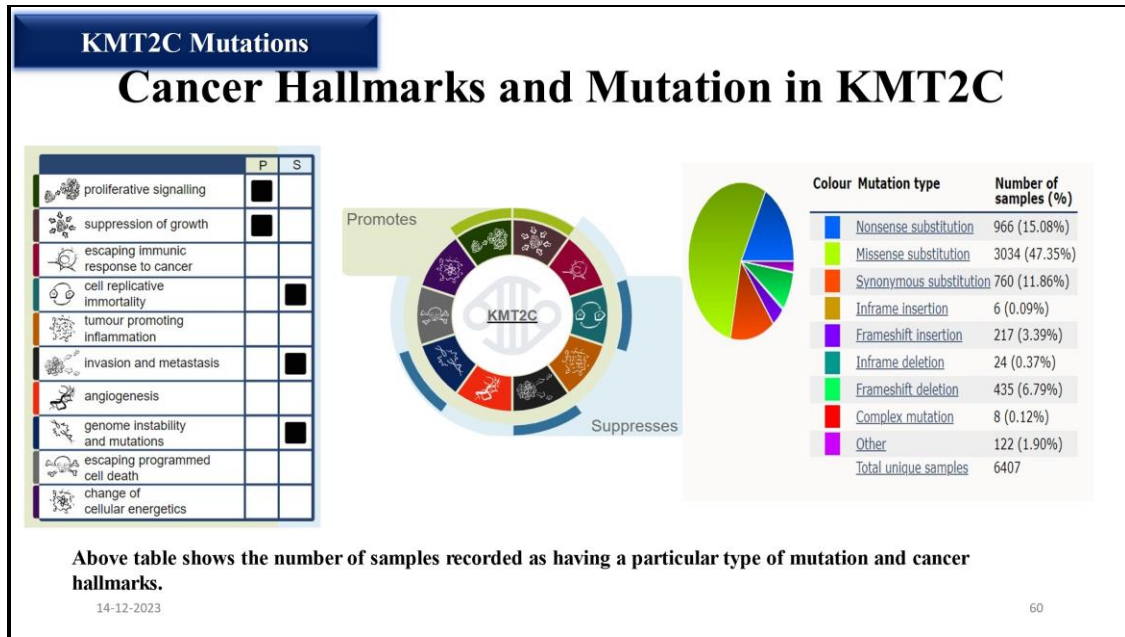
Figure 69: Cancer mutations and Hallmarks in KMT2C

6.22 Ualcan Database

6.22.1 *In-silico* Hypo and Hyper Methylation levels by Ualcan database

University of Alabama at Birmingham (also known as UAB) Cancer data analysis is a portal for cancer data analysis, and UALCAN is an all-inclusive, user-friendly, and interactive web resource for cancer OMICS data analysis. Cancer data analysis is for cancer data analysis. PERL-CGI is used to construct it, and JavaScript and CSS are used to create graphics of a high quality. The following is what UALCAN is intended to accomplish: a) make it simple for users to obtain cancer OMICS data that is publically available (TCGA, MET500); b) give users the ability to find biomarkers or undertake *in silico* validation of genes that may be important; and c) make it easier for users to analyze these datasets.
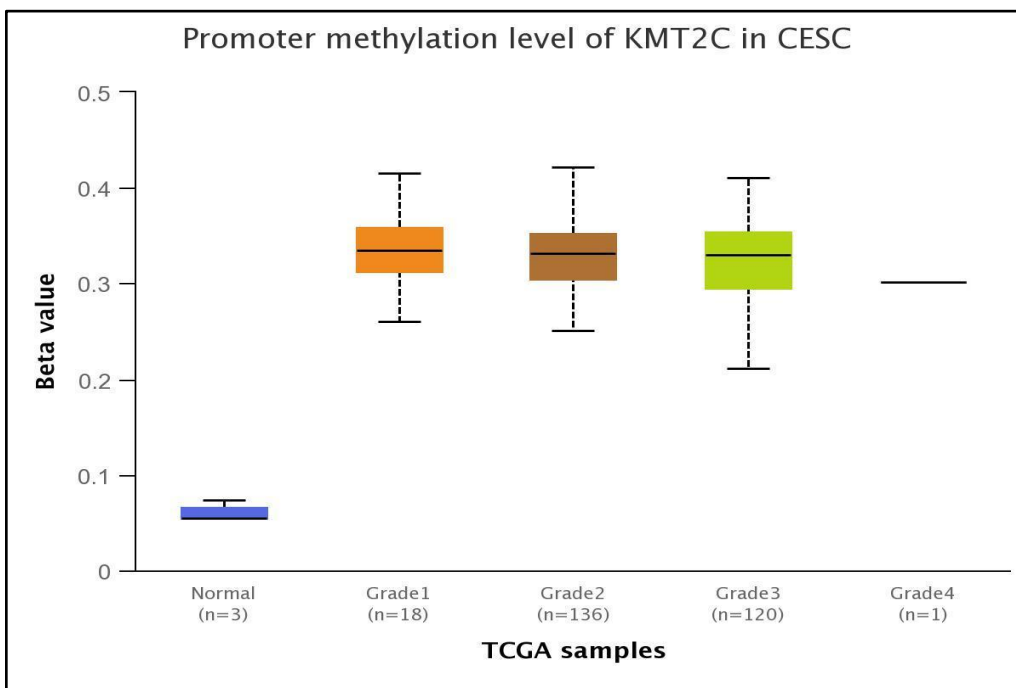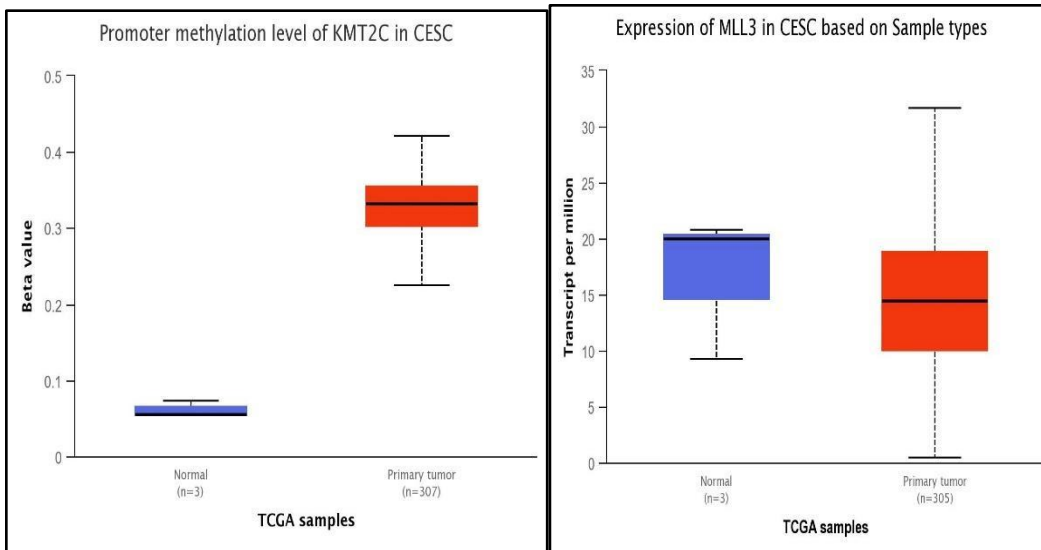
Figures 70 (A): KMT2C mutation normal vs Tumor. Figure 70 (B): Methylation levels of control and tumor tissue shows beta methylation. Figure 70 (C): TCGA Datasets of normal and Grade 1, grade 2, grade 3 and grade 4 tumor samples showing hypomethylation

Comparison	Statistical significance
Normal-vs-Grade1	2.51220044766853E-10
Normal-vs-Grade2	1.62447832963153E-12
Normal-vs-Grade3	2.85980128467145E-11
Normal-vs-Grade4	N/A
Grade1-vs-Grade2	6.407600E-01
Grade1-vs-Grade3	1.142810E-01
Grade2-vs-Grade3	3.470400E-02

Table 19: Methylation statistical significance of normal and tumor tissue samples
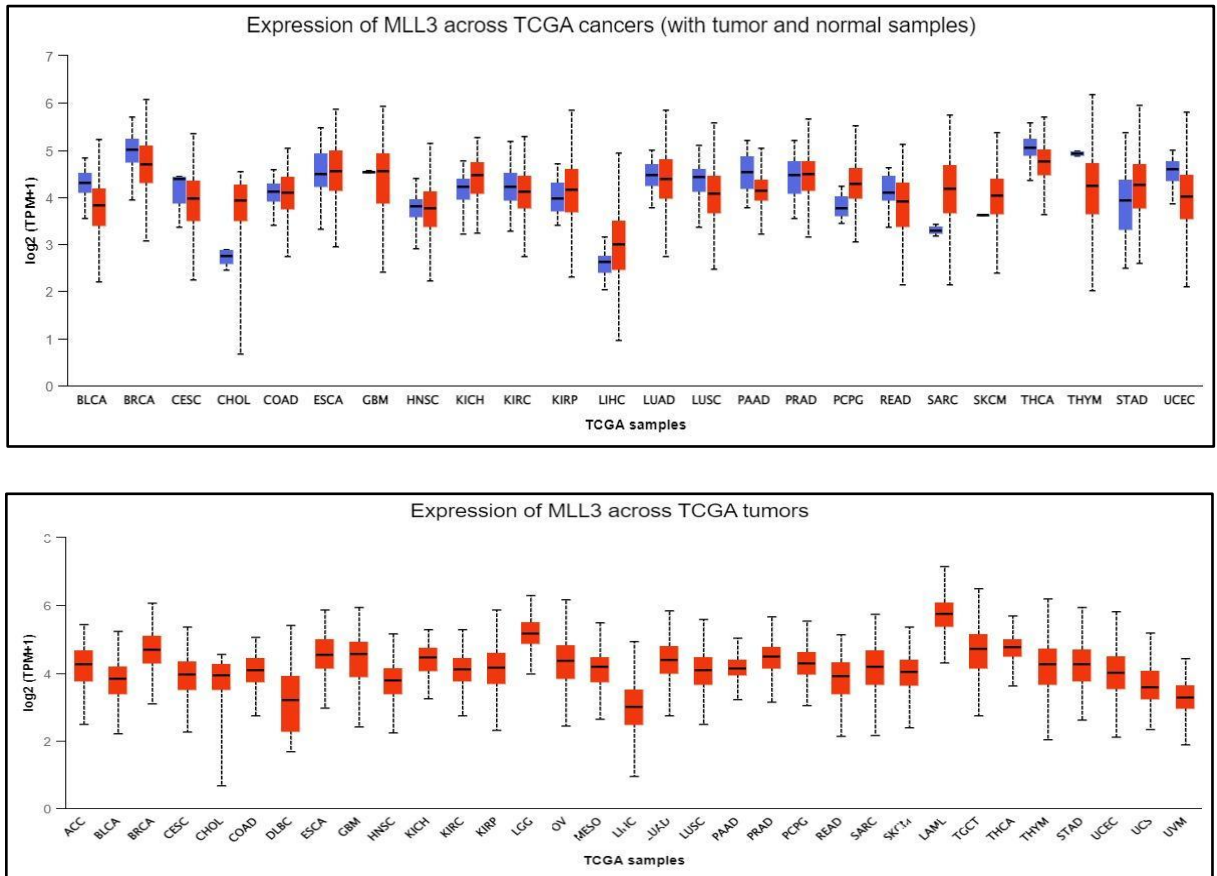
Figure 71: (A) Expression of KMT2C/MLL 3 in cervical cancer methylated genes and (B) expression of KMT2C/MLL3 in all the cancers
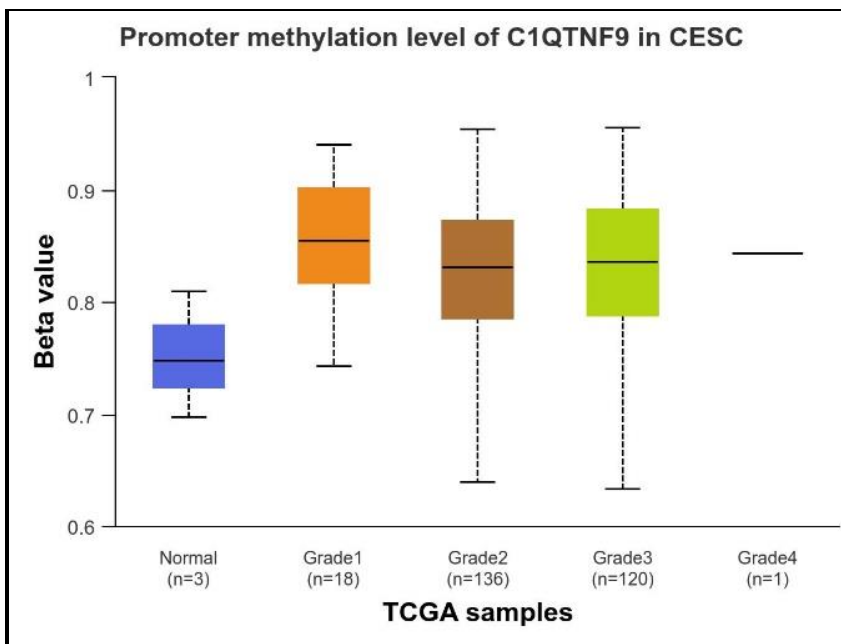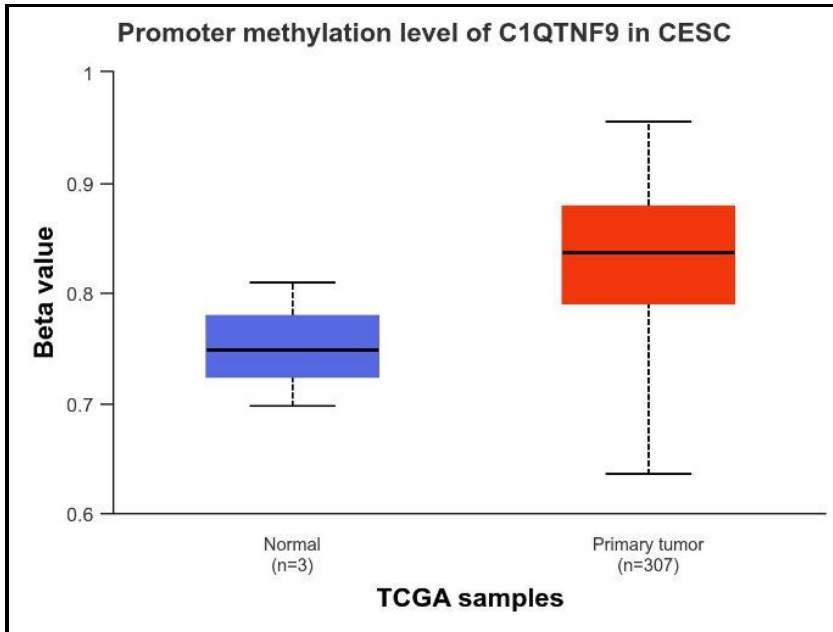
Figure 72 (A): C1QTNF 9 mutation with CESC (normal with primary tumor) (B): C1QTNF 9 mutation with normal tissue with grades of CSCE

174

The Beta Value indicates level of DNA methylation ranging from 0 that is unmethylated to 1 which is fully methylated. Different beta value cut off has considered to indicate hyper methylation or hypomethylation if the value is between (0.7-0.5) it is hyper methylated and if the beta value is between 0.3 to 0.25 it is hypomethylated. Hence, C1QTNF 9 it is hypermethylated and fully methylated as the value is between 0.7.

6.23 I – TASSER (Iterative Threading Assembly Refinement)

6.23.1 I – TASSER results

Protein structures of KMT2C exon 8 region Chr 7:152265083 variation C>A, T, rs ID 138908625 (missense and frameshift mutation). The results of I TASSER show the wild-type sequence results with 5 predicted models and query 1 with 5 models and Query 2 with 3 predicted models and their secondary structures, these models with different c scores based on their TM value and energy categorized. The best score with a positive value can be considered for comparison. PDB hits which are identical to KMT2C wild type and mutated sequences, Ligands, and Gene ontology PDB hits that are identical, showing biological consensus, and molecular function of the PDB images. I TASSER is an in silico computational work, and predictions of in silico or any bioinformatic tools will have some limitations too as they may vary according to their accuracy.

I TASSER having in built construction of 3D models relies on the alignment of numerous threads using the Local Meta-Threading Server (LOMETS) is a web-based service that predicts protein structure. The software creates 3D models by gathering top-scoring alignments between target and template structures from nine threading applications that are installed locally several tools. The trustworthiness of the structure estimate results is quantified by a template modeling (TM)-score, which is a numerical value ranging from 0 to 1. A TM score less than 0.2 means there is no similarity between structures and a TM-score above 0.5 implies two structures with comparable topology (Wang, Dan, et al.2022).

6.23.2 Predicted Secondary structure

Sequence	20 40
	`	|`
	KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKRAG
	WQCPECKVCQNC
Prediction	CCCCCCCCCCCCCCCCCCCSSSCCCCCCCCCCHHHCCCCCCCCCCCCCCSS
	CCCCCCCCCC
Conf.Score	98988778899799767678799892804121688888888987899588677659
	H: Helix; S: Strand; C: Coil

Figure 73: Wild type KMT2C Predicted structure and Conf. score

If the configuration score is more than 5, the secondary structure prediction is strong and confident. Here we got a higher score prediction of the secondary structure.
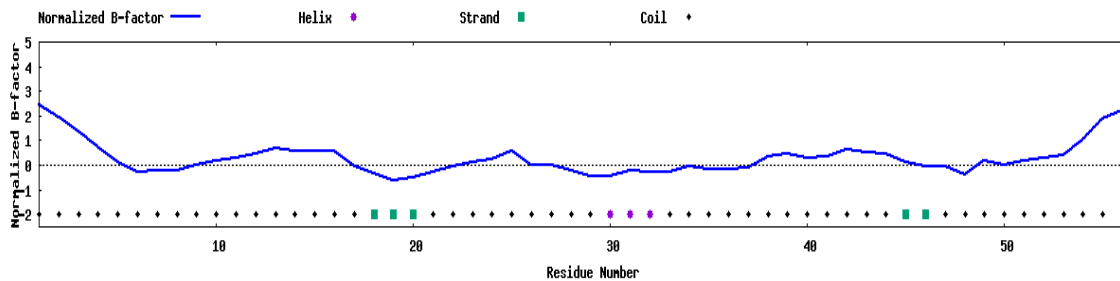
Figure 74: Predicted normalized B- factor of first sequence

Rank	PDB Hit	Iden 1	Iden 2	Cov	Norm. Z-score	Download Align.			
							20 40		
							`	|`		

						Sec.Str Seq	
						CCCCCCCCCCCCCCCCCCSSSCCCCCCCCCHHHCCCCCCCCCCCCCCSSCCCCCCCCCC KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKRAGWQCPECKVCQNC	
1	1f62A	0.31	0.29	0.91	0.90	Download	--- ARCKVCRKKGEDDKLILCDECNKAFHLFCLRPALYEVPDGEWQCPACQPAT--
2	7ebjA	0.27	0.36	0.91	4.35	Download	--EDFCAVCINGGE--- LLCCDRCPKVYHLSCHVPALLSFPGGEWVCTLCRSLTQP
3	6ryr	0.28	0.38	0.95	1.35	Download	HHMEFCRVCKDGGE--- LLCCDTCPSSYHIHCLNPPLPEIPNGEWLCPRCTCPALK
4	6ryr	0.28	0.38	0.95	1.03	Download	HHMEFCRVCKDGGE--- LLCCDTCPSSYHIHCLNPPLPEIPNGEWLCPRCTCPALK
5	4q6fA	0.20	0.27	0.96	4.67	Download	- NKVTCLVCRKGDNDEFLLLCDGCDRGCHIYCHRPKMEAVPEGDWFCTVCLAQQV-
6	2ysm	0.96	0.95	0.96	1.33	Download	- SGANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKRAGWQCPECKVCQN-
7	2miqA	0.25	0.32	1.000	0.81	Download	LMKYICHICNRGDVEESMLLCDGCDDSYHTFCLLPPLTSIPKGEWLCPRCVVEEVS

8	1f62A	0.31	0.29	0.991	3.95	Download		--- ARCKVCRKKGEDDKLILCDECNKAFHLFCLRPALYEVPDGEWQCPACQPAT--
9	7ebjA	0.27	0.36	0.991	0.96	Download		EDF--CAVCINGGE--- LLCCDRCPKVYHLSCHVPALLSFPGGEWVCTLCRSMEYD
10	5t8rA	0.20	0.27	0.996	3.69	Download		- NKVTCLVCRKGDNDEFLLLCDGCDRGCHIYCHRPKMEAVPEGDWFCTVCLAQQV -

Table 20: Predicted top structures with PDB hits, Coverage and sequences

Rank	PDB Hit	Iden1	Iden2	Cov
1	1f62A	0.31	0.29	0.91
2	7ebjA	0.27	0.36	0.91
3	6ryr	0.28	0.38	0.95
4	6ryr	0.28	0.38	0.95
5	4q6fA	0.20	0.27	0.96
6	2ysm	0.96	0.95	0.96
7	2miqA	0.25	0.32	1.00
8	1f62A	0.31	0.29	0.91
9	7ebjA	0.27	0.36	0.91
10	5t8rA	0.20	0.27	0.96

Table 21: PDB hits of KMT2C wild and Query 1 and query 2 sequences coverage.

The Coverage of the top 10 PDB hits of KMT2C are almost the same between 0 and 1 values.

Top 5 structures prediction of KMT2C

Name	C-score	Exp.TM-Score	Exp. RMSD	No. of decoys	Cluster density
KMT2C					
Model1:	0.22	0.74+-0.11	2.4+-1.8	9865	0.5228
Model2:	-0.40			4916	0.2786
Model3:	-0.56			3892	0.2385
Model4:	-1.59			1610	0.0853
Model5:	-1.39			1716	0.1038

Table 22: Top 5 models of KMT2C and their C score, TM score, and RMSD

C- score is used to estimate the quality of projected models. If the c- score is in between (-5, 2) the model signifies with high confidence and higher value. It is calculated based on template alignment and structure parameters. TM score and RMSD are measured based on the structure similarity between two structures it is used to evaluate the accurateness of the structural model with the native structure.
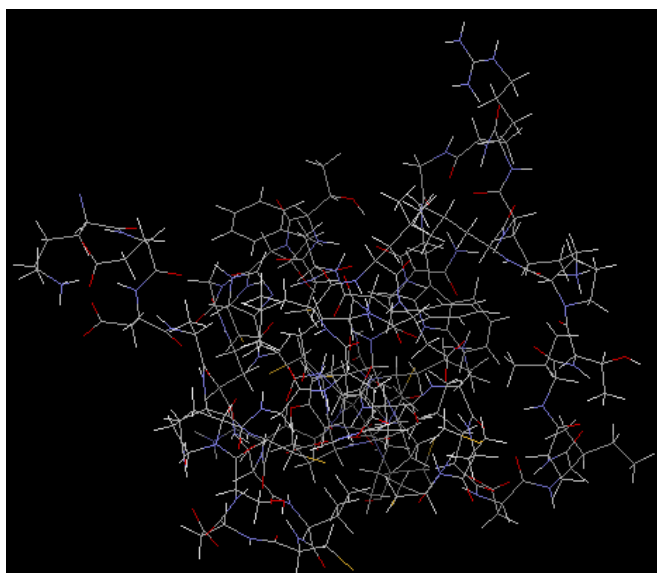
Figure 75: Best predicted image among 5 models. The c-score value is 0.22 and showing cluster range 0.5228

The structures with structural similarity have highest TM value, and Cov is the coverage of the alignment of TM and structural aligned residues.

PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov	Alignment
4	2puyB	0.805	1.25	0.302	0.946
5	3souA	0.791	1.69	0.278	0.911
6	5vabA	0.783	1.65	0.231	0.929
7	4q6fA	0.782	1.65	0.200	0.982
8	6guuA	0.766	1.55	0.264	0.946
9	6ryrW1	0.754	1.60	0.283	0.946
10	4gndA	0.749	1.54	0.255	0.911

Table 23: The 10 PDB structures showing close structural similarity

These are 10 Gene Ontology PDB structures

Rank	Cscore^{GO}	TM-score	RMSD^a	IDEN^a	Cov	PDB Hit
1	0.35	0.8281	1.27	0.23	1.00	2kwjA
2	0.34	0.8053	1.25	0.30	0.95	2puyB
3	0.33	0.8098	1.64	0.27	1.00	2e6rA
4	0.32	0.7281	1.78	0.29	0.91	2e6sA
5	0.32	0.7201	2.40	0.23	0.98	1wevA
6	0.31	0.6496	2.25	0.21	0.84	2qicA
7	0.30	0.6864	2.11	0.24	0.91	3o36A
8	0.30	0.6072	2.45	0.34	0.88	1f62A
9	0.29	0.7053	1.85	0.23	0.95	1xwhA
10	0.29	0.6624	2.24	0.22	0.93	2xb1C

Table 24: Gene ontology template structures similarity with the query protein and template.

Consensus Prediction of GO terms

Molecular Function	GO:0008270	GO:0005515
GO-Score	0.87	0.87
Biological Process	GO:0035556	
GO-Score	0.32	
Cellular Component	None was predicted	

Table 25: GO ontology aand consensus prediction of Query

Query -1 KMT2C   G> T (R to L)

sequence submitted

>protein

KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKLAGWQCPECKVC
QNC

Predicted secondary structures

Sequence	20 40
	| | KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKLAG WQCPECKVCQNC
Prediction	CCCCCCCCCCCCCCCCCCSSSCCCCCCCCCCCCCCCCCCCCCCCCCSS CCCCCCCCCC
Conf.Score	988756668798799774777288996804011598888889972599588677759
	H: Helix; S: Strand; C: Coil

Figure 76: Predicted structure and configuration score of KMT2C (Query 1)

As the configuration score in most of the strands are above 5. There is a chance of higher confidence.

Predicted Solvent Solubility

Sequence	20 40
	| |
Sequence	KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKLAGW QCPECKVCQNC
Prediction	86754043464674643002113474321233162634634747443464441666

The values range from 0 to 9, with 0 representing buried residue and 9 representing widely exposed residue.

Predicted Normalized factor
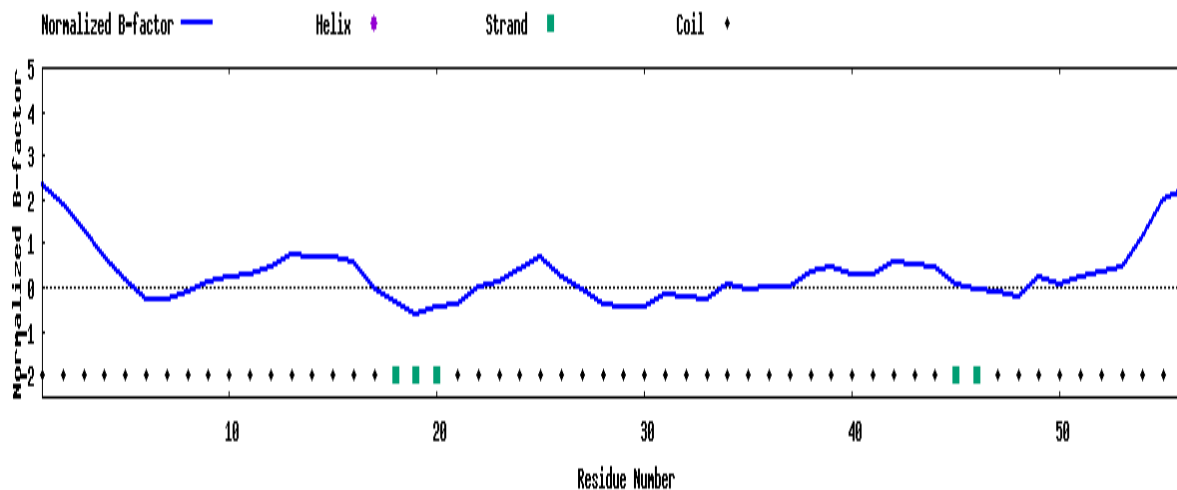
Figure 77, A & B: Predicted Normalized factor of KMT2C sequence. The negative values mean the residue is relatively more stable in the structure.

The top 5 models scores of KMT2C (R to H)

Name	C-score	Exp.TM-Score	Exp. RMSD	No. of decoys	Cluster density
KMT2C					
G>T					
Model1:	0.15	0.73+-0.11	2.6+-1.9	9811	0.4907
Model2:	-0.62		4472	0.2272	
Model3:	-0.54		4357	0.2455	
Model4:	-0.03		6792	0.4085	
Model5:	-1.74		1205	0.0738	

Table 26: Top 5 models of KMT2C of Query 1 showing C score and Tm score
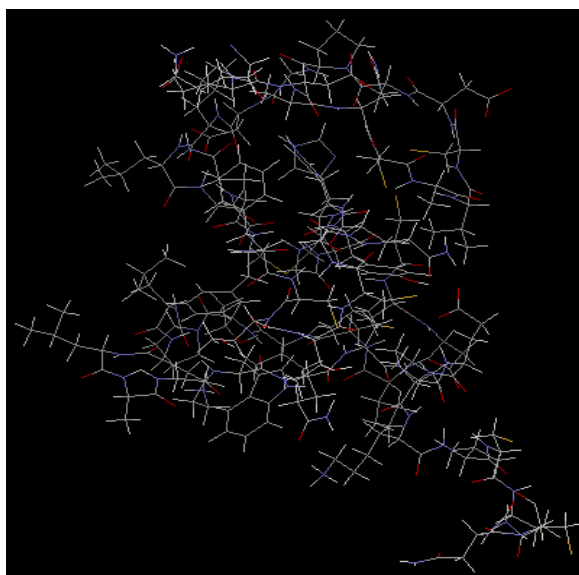
Figure 78: Structural prediction showing the sight variation from the wild type as the c score value of the best confidence level shows 0.15, Tm value 0.73+_0.11 and cluster density with 0.490

Rank	PDB Hit	TM-score	RMSD[a]	IDEN[a]	Cov
1	2ysmA	0.829	1.26	0.255	0.982
2	2e6rA	0.827	1.38	0.250	1.000
3	5vabA	0.814	1.51	0.231	0.929
4	3souA	0.797	1.87	0.278	0.929
5	2kwjA	0.789	1.53	0.232	1.000
6	4qf3A	0.773	1.72	0.200	0.982
7	5xfrA	0.765	1.85	0.214	0.946
8	1wevA	0.756	2.42	0.232	1.000
9	6guuA	0.755	1.98	0.264	0.946
10	2puyB	0.748	1.78	0.302	0.946

Table 27: Top 10 identified structural analogs in PDB

Go score is the evaluation of global and query protein similarity and it ranges from 0- 1 and higher values indicate high confidence predictions. The GO scores of the predictions

of all top hits show from the range of 0 to 1 and these are best identified gene ontology analogues.

Rank	Cscore^{GO}	TM-score	RMSD^a	IDEN^a	Cov	PDB Hit
1	0.33	0.8271	1.38	0.25	1.00	2e6rA
2	0.33	0.7888	1.53	0.23	1.00	2kwjA
3	0.33	0.7563	2.42	0.23	1.00	1wevA
4	0.32	0.7205	1.99	0.28	0.93	2e6sA
5	0.31	0.7482	1.78	0.30	0.95	2puyB
6	0.31	0.7205	1.92	0.23	0.91	3o36A
7	0.30	0.7156	2.41	0.21	1.00	2yt5A
8	0.30	0.6238	2.26	0.33	0.86	1f62A
9	0.29	0.7115	1.93	0.24	0.95	1xwhA
10	0.29	0.6494	2.10	0.23	0.91	1mm2A

Table 28: Top 10 Gene ontology templates and predicted PDB hits

Consensus prediction of GO terms		
Molecular Function	GO:0008270	GO:0005515
GO-Score	0.86	0.86
Biological Process	GO:0035556	
GO-Score	0.33	
Cellular Component	None was predicted	

Table 29: GO predictions and score of Query 1

Query 2 - KMT2C (Query 2) G to A (R to H)

>protein

KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKHAGWQCPECKVCQN

Predicted secondary structure

Sequence	20 40
	<pre>	|</pre> KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKHAG WQCPECKVCQNC
Prediction	CCCCCCCCCCCCCCCCCCCSSSCCCCCCCCCCHHHCCCCCCCCCCCCCSS CCCCCCCCCC
Conf.Score	988887788997996646786998938030205988788899788995877877 69
	H: Helix; S: Strand; C: Coil

Figure 79: Predicted structure and Conf. score of KMT2C (Query -2)

The above configuration score is more than 5 so the predicted secondary structure is highly confident. The helical structure H is a good configuration structure, strands of S also denote with all values above 5 but in coil region the structure varies as few coils show high confidence level and above 5 and few irregulars below the range.

Predicted solvent Accessibility

Sequence   20   40
                   | |
KEDANCAVCDSPGDLLDQFFCTTCGQHYHGMCLDIAVTPLKHAGWQCPECKVC
QNC

Prediction 877444634646745440121253443212431726246347564434644441667

The values range from 0 to 9, with 0 representing buried residue and 9 representing widely exposed residue.

The values range from 0 to 9, with 0 representing buried residue and 9 representing widely exposed residue.

Projected normalized B factor

Higher the negative values the prediction score and B –factor is more confident
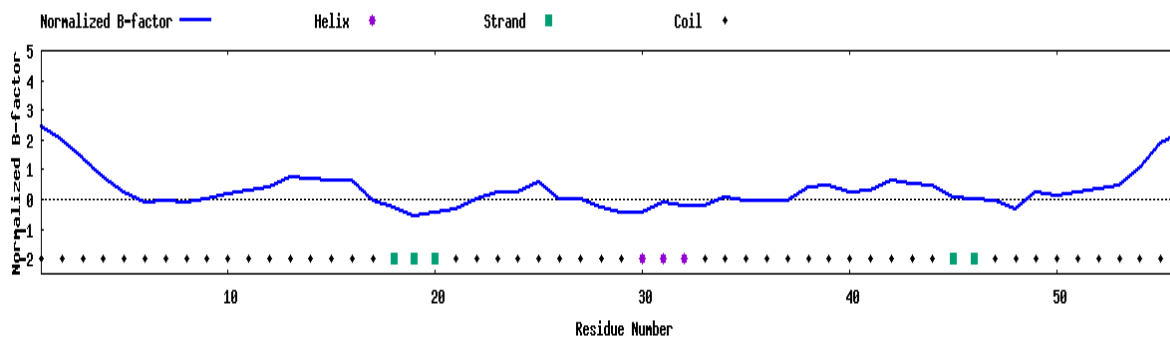
Figure 80: Projected normalised B factor

Name	C-score	Exp.TM-Score	Exp.RMSD	No.of decoys	Cluster density
Model1:	0.22	0.74+-0.11	2.4+-1.8	9992	0.5267
Model2:	-0.20		5719	0.3440	
Model3:	-1.35		2081	0.1091	

Table 30: The values of c score should be from -5 to 2 and there is positive result the higher confidence level and accuracy. Model 1 shows higher accuracy and confidence and the other two models fall below the range and show lower accuracy.
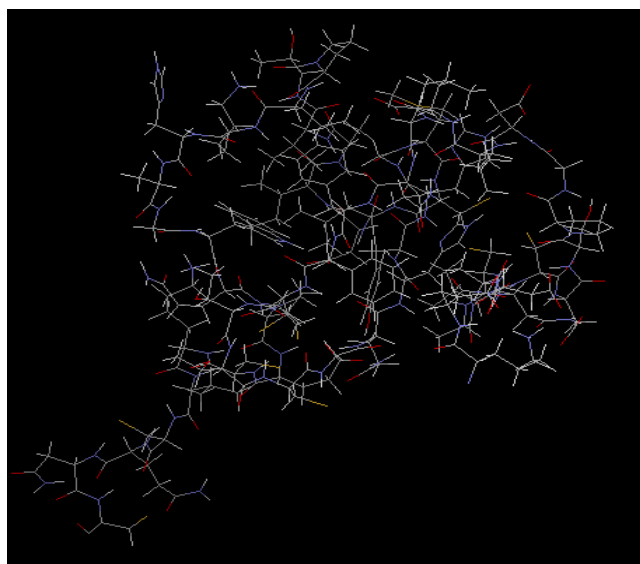
Figure 81: Homology modelling shows best predicted image among 5 models from I Tasser results. The c score 0.22, TM score 0.74+_0.11 and showing cluster density of 0.5267.

These are 10 ten Gene ontology homology structures that are predicted.

Rank	Cscore^{GO}	TM-score	RMSD^a	IDEN^a	Cov	PDB Hit
1	0.35	0.8425	1.36	0.25	1.00	2e6rA
2	0.34	0.7991	1.53	0.23	1.00	2kwjA
3	0.33	0.7312	2.20	0.28	0.95	2e6sA
4	0.32	0.7373	2.55	0.23	1.00	1wevA
5	0.31	0.7465	1.85	0.30	0.95	2puyB
6	0.31	0.7111	2.07	0.23	0.91	3o36A
7	0.30	0.7081	2.39	0.22	0.98	2yt5A
8	0.30	0.7204	1.96	0.24	0.95	1xwhA
9	0.30	0.6079	2.45	0.33	0.86	1f62A
10	0.29	0.8373	1.27	0.26	0.98	2ysmA

Table 31: Query 2 Top 10 gene ontology and homology structures

The GO scores should be from 0 to 1, and if higher the range the more accuracy and confidence level. Cov is the coverage, Tm align and structurally aligned residues of query protein. IDEN is the sequence identity with structural alignment.

Consensus prediction of GO terms		
Molecular Function	GO:0008270	GO:0005515
GO-Score	0.86	0.86
Biological Process	GO:0035556	
GO-Score	0.32	
Cellular Component	None was predicted	

Table 32: Query 2 KMT2C GO score and their molecular function and biological process shown the activity and no cellular activity

GO in terms pf molecular function, biological process and cellular component reported till now and the GO score ranges from 0 to 1

Results of KMT2C wild type and mutated type

These are the three KMT2C structures of wild type and mutation sequences and their Tm score binding affinities and C value scores results.

Name	C-score	Exp.TM-Score	Exp. RMSD	No. of decoys	Cluster density
KMT2C					
KMT2C	0.22	0.74+-0.11	2.4+-1.8	9865	0.5228
KMT2C 1	0.15	0.73+-0.11	2.6+-1.9	9811	0.4907
KMT2C 2	0.22	0.74+-0.11	2.4+-1.8	9992	0.5267

Table 33: Wild type and variations of query 1 and query 2 predictions of the c score, Tm score and cluster density

It is calculated on the basis of template alignment and structure parameters. Expected Tm score and RMSD are measured based on their structure similarity between two structures it is used to evaluate the accuracy of the structural model with the native structure. Decoys

are low-temperature structures generated during monte Carlo simulations. Higher Cluster density is to find the higher simulations with good quality.

Therefore, from our studies we observe KMT2C projects higher similarity with KMT2C (Query 2) followed by Query 1

6.24 Protein – Protein Interactions of KMT2C

KMT2C interacts with the following lysine methyltransferase group genes and also other genes such as KMT2D, ASHL2, SETD1A, RBBP5, KDM6A, PAXP1, PAGR1, NCOA6,DPY30, WDR5. The network node represent post-translational modifications from the gene and all the nodes represent from a single gene.
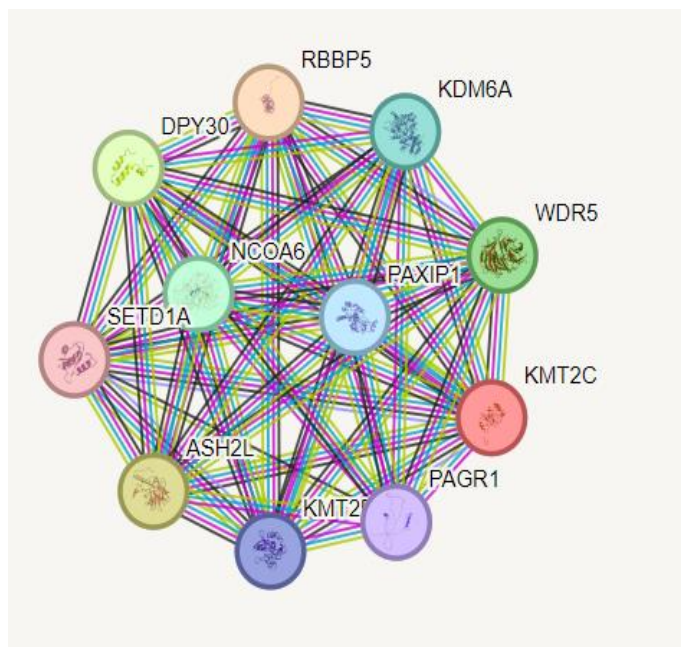
Figure 82: Protein-Protein Interactions of KMT2C and their interacted network of genes

6.25 KMT2C docking using Cluspro 2.0 software

In our study a few compounds that were having good therapeutic value on cancer cell lines. According to the literature search, these were shown very good activity on cancer cells as we are focusing on molecular docking with KMT2C we have used Cluspro 2.0 software

for docking. Hence**,** Protein- Protein Docking of KMT2C binding with the compounds that show activity on cervical cancer such as Artrmisinim, Shikonin, Sitoindoside IX, Bucidarasin A, Betulin. The binding energies of Betulin shows good activity with KMT2C with -12.5 Kcal/mol then further Bucidarasin A with -12.3 Kcal/Mol and followed by – 11.5 Kcal/mol in Sitoindoside IX (Table 34). The best docking structures of KMT2C with Betulin and Bucidarasin A are as shown in the **(Figure: 83 & 84)**
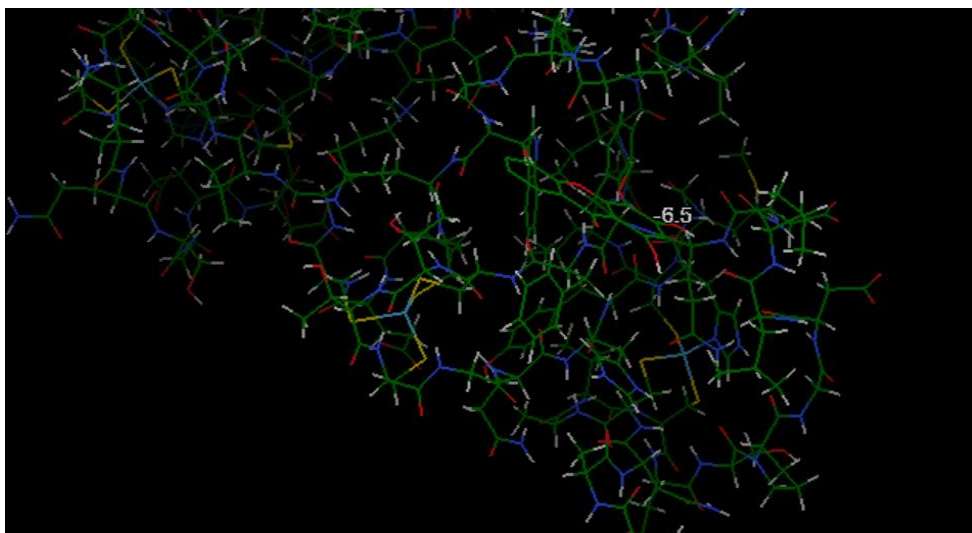
Figure 83: Protein -Protein (P - P) docking and binding affinity of KMT2C with Betulin is -12.5
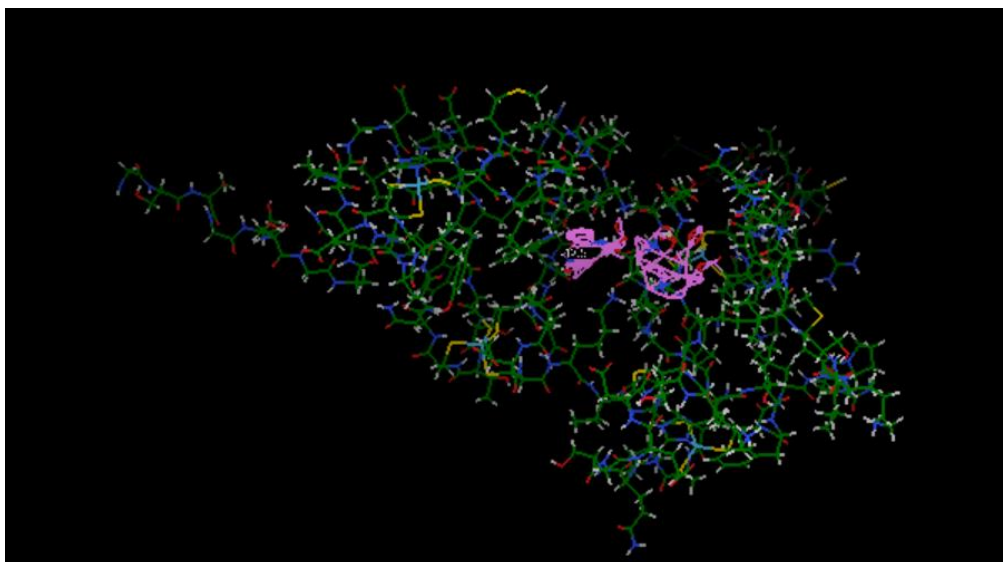
Figures 84: P – P docking and binding affinity of KMT2Cwith Bucidarasin A shows - 12.3

Compound name	Binding energy (Kcal/ mol)
Artemisinim	-9.9
Shikonin	-10.6
Sitoindoside IX	-11.5
Bucidarasin A	-12.3
Betulin	-12.5

Table 34: Protein - Protein Docking of KMT2C with few compounds useful in the treatment in Cervical cancer

KMT2C/MLL3 is a complex and is involved in Acute leukemogenesis, Bladder, Breast, Colon, Cervical cancer and developmental disorder; Lysine methyltransferases. Understanding the gene KMT2C mutation and its involvement in genome integrity, DNA damage, cellcycle, deciphers tumor mutation burden or tumorigenesis (Sinha, Vimlendu Bhushan, et al 2016).

In the evaluation the best predicted structures of the original sequence of KMT2C and the two variations with highest energy and confidence levels of all structures. SNPs show a

vital role in the disease phenotypes. It is clear that changing a protein binding and slight variation changes the secondary structure and their binding affinities. Here KMT2C of wild type sequences are compared with mutated sequences, both the sequences differ in their nucleotide base pair or protein structure showing mutation.

Protein structures of KMT2C of original or wild type sequence Arginine changes to Histidine (query 1) with respect to Missense mutation, variation of the nucleotide (G>A) and protein from (R to H) modifies. Arginine is involved in synthesis of Amines, nitric oxide and multiple cellular processes, Arg directly activates mTOR signaling pathway. It is an epigenetic regulator and plays a key role in cancer metabolism, cell proliferation, activates as a therapeutic target when cells undergo deprive condition (Chen et al.2021). Arginine expression is increased in several cancers such as colorectal cancer, gastric cancer, breast cancer, AML, Prostate cancer, endometrial cancer (Wei et al.2021). So, in our study KMT2C protein modeling reveals that Arginine is the amino acid where the mutation is shown to cause cervical cancer. I TASSER is a software app used for protein structural and 3D conformation prediction to analyse the structures of KMT2(Zhang et al. 2008). Leucine is a branched chain Amino acid, and it is not present in human cells but through food supplements and dietary intake it is present. Expression of androgen receptors in prostate cancer moderates LAT 3 and leucine uptake can be inhibited by anti- androgen treatment (Wei, Zhen, et al.2021). Histidine is an amino acid having neutral PH and basic side chain substitution of arginine to Histidine in exon 8 region might impair the post translational changes and DNA damage, and pH value might alter and cause cell proliferation and metabolism. In, the next sequence (Query 2), Arginine modifies to Leucine with respect to frameshift mutation, Variation from (G>T) and protein from R to L. However, KMT2C is a potent mutation showing both missense and frameshift variations on the same chromosome number 7 and position.

In Figure73, the predicted secondary structure is having more than 5 configuration scores in the sequence so it means the predicted KMT2C structure poses good configuration. Our findings indicate that the cluster size is a more resilient metric than the C-score when it comes to rating the projected models. These final models are scored based on size of the

cluster rather than C-score in the result. However, the C-score exhibits a robust association with the final models' quality, enabling a quantitative estimation of the RMSD and TM-score of these models about the native structure. The limitation of the study, this robust link is only present in the initial anticipated model derived from the most extensive cluster. Therefore, the C-scores of the lower-ranked models (models 2-5) are provided solely for reference, and it is not recommended to compare them. Put simply, while lower-ranking models may occasionally have a better C-score than the first model, on average, the first model is the most dependable and should be chosen unless there are specific reasons not to choose.

In our findings from the **(Table 22, 26, 30)** as the top predicted models of wild KMT2C and KMT2C 1(Query 1) and KMT2C 2 (Query 2) only first models of each table were considered as c score of the first model is mostly considered in the I TASSER and in the three top models with c score cluster density TM values are considered. When compared with wild type c score is 0.22, TM value is 0.74+-0.11 in the wild type and cluster density is also 0.5228 and RSMD is also almost similar to Query 2 KMT2C 2 showing c score and TM value same as wild type and RSMD, cluster density was almost similar. So, Query 2 is similar to wild type than Query 1.

The protein models of KMT2C in table 33 determines the wild type KMT2C and Mutated Query sequence 1 and Query 2 with their secondary structure predictions are shown by graphical model and the homology modeling of KMT2C structure and top predicted GO ontology PDB and binding scores are evaluated. The Tm score and Cluster density of wild type KMT2C when compared with mutated sequences Query 1 and Query 2, the Query 2 sequence of KMT2C mutation shows higher TM score and Cluster density. The homology modeling of the KMT2C variant shows the same structural similarity with wild type.

The Gene ontology of top 10 PDB structures are analysed based on the values 0 and 1, and in the **(See Tables.22,28,31)** shows the coverage values are almost similar ranging from 0 to 1 in wild type, then Query 2 and query 1. I TASSER is in silico-based software where it produces inside several thousands of PDB IDs but outputs only top ten GO PDB hits.

Protein -Protein Interaction network of KMT2C and their interactions shows (**Figure 82**) that all the genes of the interactome are involved directly or indirectly to DNA damage, Epigenetic modifications, and post translational changes and apoptosis of the cell. ASHL2 is histone methyltransferase complex subunit, and it may function as a transcriptional factor and may produce a role in hematopoiesis (Lüscher-Firzlaff, et al.2019). In connotation with RBBP5 and WDR5 ignites KMT2C, KMT2D and SETD1A (Lüscher-Firzlaff, et al.2019). PAX-interacting protein 1 (PAXP1) is implicated in the DNA damage and transcriptional control via histone methyltransferase. Contributes to the initial stages of development. Cell survival following ionizing radiation necessitates DNA damage response. In vitro studies have demonstrated the involvement of homologous recombination processes in repairing double-stranded breaks. The localization of it to DNA damage foci necessitates the presence of TP53BP1 and is recruited to DNA impairment foci and, in certain repair pathways, its interaction with TP is necessary for an efficient DNA damage response. Retinoblastoma-binding protein 5 (RBBP5) is essential for the differentiation potential of embryonic stem (ES) cells, especially in the brain lineage (Yang, Zhiqin, et al.2023). Serves as a distinct marker for epigenetic mechanisms. PAGR1 is also involved in DNA damage. Nuclear receptor coactivator 6 (NCOA6) is a protein that directly interacts with nuclear receptors and enhances their transcriptional activity in a hormone-dependent manner. Enhances gene expression via acting in a manner that relies on the presence of the AF2 domain (Coulthard, 2003). Participating in the simultaneous activation of various nuclear receptors, including those for steroids, thyroid hormone, vitamin D3. It is also involved in the action of NF-kB pathway.

Further, KMT2C Protein- Protein docking performed with the compounds, whereas Artemisinin is helpful in treating cervical cancer and it shows higher cytotoxicity in cervical cancer cell lines and cells (Goodrich et al.2014). Arteminisin also reported in leukemia, Colon cancers, and even in other others by several scientists. Shikonin is useful in the treatment of CC, as Shikonin exerted a substantial inhibitory effect on cell migration and led to a decrease in the production of migration-related proteins, such as MTA1, PTEN, AKT, TGFβ1, and VEGF (Xu et al.2022). Shikonin also inhibit viability of the cell,

invasion cell arrest on Hela cell line (Tang Q et al 2020). Betulin is an anti-cancer compound, it can induce apoptosis in CC cell lines such as Hela and induces apoptosis om PIK3/AKT signalling pathway and molecular mechanism (George et al. 2015). Even Bucidarisin A induces the cytotoxicity in cervical cancer cell lines Ic5 (Efdi et al.2007). Furthermore, increased doses successfully reduced the growth of tumors in mice with CC. The results unequivocally validated the therapeutic capacity of BA for cervical cancer (Wang, Dan, et al.2022). Arteminisin in several previous studies seen in cervical cancer radiotherapy treatment as a drug it is used (Huang et al.2021) and Bucidarasin is a diterpenoid and it is used in the drug therapeutic management (Huang et al.2021). The impact of Betulin or betulinic acid on cervical cancer merits attention. Several investigations have demonstrated in a previous study that a concentration of 30 μM Betulin resulted in an inhibitory impact of 60 % to 100% on HeLa cells, as reported ((Efdi et al.2007)). In our study the binding energy of KMT2C shows greater binding energy with Betulin (-12.5 Kcal/mol) followed by Bucidarisin A (-12.3 Kcal/mol) **(Table 34).**

Chapter 7
Summary and conclusion

Summary:

CC is one of the top cancers among women, and early diagnosis and screening tests play a vital role in the detection of CC. Surgery is an effective treatment for early-stage cervical cancer, but concomitant chemoradiation is the preferred option for cervical cancer that has advanced locally. Nevertheless, a significant proportion (21-30%) of individuals diagnosed with cervical cancer encounter a relapse or the spread of the disease to distant sites following initial treatment. Unfortunately, the options for managing these patients are restricted. Hence, the primary reasons for mortality in advanced cervical cancer are local recurrence and distant metastases.

7.1 To catalogue the Human papillomavirus (HPV) through genotyping and phylogenetic analysis.

The samples considered for the present study are 25 samples of which 20 samples are tumor tissue samples and 5 surrounding tissue or control samples. We first isolated DNA with a mini Quigen kit and further HPV detection for the presence or absence of HPV was done with Gp+5 and GP +6 primers, however after confirming HPV genotyping was performed to check the type of HPV (low or high-risk). HPV genotyping with proper PCR conditions maintained and performed PAGE gel electrophoresis using *RsaI* restriction enzyme. In our study we found the bands and base pairs of HPV 6, 16, 18, and 33 strains of HPV in 6 samples. After genotyping we have performed *in silico* analysis for the construction of Phylogenetic tree of LR HPV and HR- HPV to see the evolutionary relatedness of HPV strains. HPV is double stranded circular and its membrane consists of proteins such as E1 and E2 are early proteins, E4 and E5 are replicative stages of the proteins and E6 and E7 are oncoproteins and L1 and L2 capsid proteins of HPV. The sequences of HPV retrieved from the PAVE database of E6 and E7 HPV LR and HR strains of HPV and sequences are first arranged in Fasta format with > symbol and saved in fasta format all files. These sequences are uploaded in MEGA X software and performed for multiple sequence alignment and then performed for phylogenetic analysis, maximum likelihood of the

sequences to see the closeness and relatedness of the tree. Later protein protein interaction network was performed with P^{53} and P^{Rb} as these are tumor suppressor proteins of cancer, if it gets activated with the viral oncoproteins and the interactions network shows ATM gene as one of the interacted genes with P53 and PRb. ATM gene is present in all cancers as it causes DNA damage and apoptosis. Later, L1 and L2 capsid proteins of HPV strains PDB files downloaded as ligands and receptors of PDB files of endothelin A, Immunoglobulins, Heparan sulfate, EGFR, VEGF A, laminin etc from the literature as receptors for docking studies using softwares Hex 8.0 and Cluspro 2.0. The results from both evaluations gives that EGFR, Laminin shows better binding affinity and energy then Heparan Sulfate and followed by immunoglobulins and VEGF A. The HPV genotyping and through docking studies, the targeted therapies and treatment modalities can be used to treat a patient.

7.2 To evaluate the genetic variants through whole exome sequencing.

The samples for whole exome sequencing taken are ten. Five tumor tissue samples and five surrounding tissue samples or control samples were sent for Exome sequencing to DNA xperts company, DNA extraction with FFPE kit based and library preparation performed with sequential steps was performed and later raw reads were retrieved from the samples. We have analysed the raw reads using inhouse built pipeline, NGS Pre-processing first step where raw reads were analysed with FASTQc to check for contamination, per base sequence quality, GC content, per base duplication sequence, etc., and further alignment was performed using bowtie2 command and later conversion of SAM files to Bam files a second step to perform variant discovery in which mpile up, indel and variant calling was performed. VCF files are uploaded in SNP Nexus to gather the results of variants. Whole exome sequencing reveals potent variants from this study and further validation steps are processed to prioritize the variants.

7.3 Prioritize variants, validate specific markers (SSCP, Sanger sequencing, *in silico* Methylation and RNA-Seq analysis), and establish their associations with disease causation per se

Variant Prioritization was performed after filtering out given the parameters for depth >5, and <20. After downloading the data from SNP nexus, further annotation and analysis performed by using Clinvar, dbSNP, and entrez etc., software tools and databases were used to know the variants type, chromosome number, pathogenic, benign, likely pathogenic or unknown. dbSNP used to see the minor allele frequency and rs ID of the variant. The CADD and GERP cut off values were taken as 10 and 2 to filter out the top variants, and venny 2.1 is used to see the similarity between the variants. In our study the top filtered variants were KMT2C, OR4M, C1QTNF 9, MGST1, FAS, EPHB1, etc.

Single Nucleotide Conformation polymorphism were studied for see the mutations, we have correlated our samples data with same grade cervical cancer samples 45 and 15 control samples to check for polymorphism, Primer designing is done for top two variants KMT2C and C1QTNF 9 and PCR was performed for the samples after that polyacrylamide gel (PAGE) electrophoresis done and the gel made to run at 110V for 16 hours, after gel run silver staining performed (AgNo3) and washed with ethanol and checked the bands in gel doc, we have got few bands in few samples and those samples were again validated with sanger sequencing results whether the variants are homozygous, heterozygous or wild type.

in silico methylation analysis done by using the Ualcan databse to check for the hypomethylation or hypermethylation of the top variants seeing the statistical significance of Beta value. KMT2C is hypo methylated and C1QTNF 9 is hyper methylated according to in silico analysis.

RNA Seq analysis of transcriptomics data of 10 samples correlated the data of our whole exome sequencing variants in which heat map is generated for the top variants of KMT2C, it shows down regulation in the heatmap, as KMT2C variant is missense, Frameshift, and non-coding, synonymous and seen in several cancers.

7.4 To perform integrative analysis using system biology-based approaches.

Integrative systems biology approaches refer to methodologies that combine diverse biological data types and computational techniques to gain a comprehensive understanding of biological systems at multiple levels of organization. These approaches aim to integrate data from various sources such as genomics, transcriptomics in our study and other high throughput technologies such as software tools and databases. COSMIC database, cBioportal, GEO datasets, TCGA data. Excel PAN data, we have done integrative analysis and correlated the WES data with transcriptomics data and generated heat maps. Several bioinformatics tools were used to analyse the variants from the WES study. Through these databases and analysis of WES and transcriptomics it lays the foundation for identifying potent and novel mutations from our study and correlation between the CC with all the databases.

Conclusion

Genotyping of HPV in addition to normal or HSIL cytology holds the potential to expedite care for a substantial proportion of individuals with abnormal test findings and may be especially helpful in groups for further longitudinal follow-up. This study constructed the phylogenetic evolutionary tree between cancers causing low and high-risk HPV types. The PPI, GeneMania, and Cystoscope interactions evaluate the top 5 interactions of these genes, which cause cancers, disease progression, subcellular localization, and pathway enrichment analysis. Hence, they may be used as better vaccine components for the control of HPV. The Protein-Protein docking results from hex 8.0 and CluPro are based on the highest binding score (energy) and its interactions with the molecules, as discussed. Of which ClusPro gives the best with laminin 5, followed by EGFR and Heparan Sulphate, and Hex 8.0 shows the best binding energy and interaction with EGFR, followed by Heparan Sulfate and interferon that might be targeted in the treatment modalities.

The interaction between HPV infection, particularly strains 16 and 18 and genetic variables such as mutations in the KMT2C variant, can contribute to the expansion and progression of CC. The KMT2C gene, commonly referred to as MLL3 (Mixed Lineage Leukemia 3), encodes a protein that acts a vital role in the regulation of genes. The KMT2C gene

mutations have been linked to different types of malignancies, including cervical cancer. Cervical cancer is frequently linked to chronic infection caused by high-risk (HPV) types. However, the presence of additional genetic abnormalities can also play a role in the development and advancement of the illness. Nevertheless, the occurrence and precise characteristics of mutations can differ among various subcategories of cancer and populations.

KMT2C plays an important role in chromatin remodeling, as it is important for gene modulation and it can lead to alterations in the gene and cause cancer. The tumour mutation burden of KMT2C is high in CC samples of our study and the investigation shows HPV and KMT2C have an association that leads to cervical cancer. Comprehending the genetic modifications, such as KMT2C mutations in cervical cancer can have therapeutic ramifications. Oncology research is being focused on developing targeted medicines that address specific genetic abnormalities, including those caused by KMT2C mutations.

Other top 10 mutations from WES from this study were C1QTNF 9, EPHB1, FAS, MGST1, OR4M1, etc also play important roles in several other cancers and these were novel mutations found in cervical cancer. Complement C1Q tumor necrosis factor 9 (C1QTNF 9) is seen in oral cancers in few studies and also head and neck cancers but very few studies or extensive studies regarding this variant have not yet been studied for cervical cancer.

Mutations in KMT2C may alter its normal function in regulating gene expression and could potentially influence the cellular processes that contribute to cancer formation. Understanding the relationship between HPV infection and genetic variables such as mutations in genes like KMT2C is critical for deciphering the mechanisms underlying CC development and discovering possible therapeutic targets or biomarkers for prognosis and treatment. However, the particular mechanisms by which KMT2C mutations contribute to CC development in the context of HPV infection may differ with various conditions.

As KMT2C is a histone methyltransferase it can target the pathways dysregulated by KMT2C mutations, drugs targeting epigenetic modifications that may offer potential therapeutic avenues. Patients with KMT2C mutations in our WES study reveal that they might exhibit distinct clinical characteristics or responses to treatment, which could inform personalized treatment strategies. Understanding the role of C1QTNF 9 in CC pathogenesis could lead to the development of targeted therapies aimed at modulating its function or the downstream pathways it regulates. Similar to KMT2C, the expression levels or mutations of C1QTNF could have prognostic significance, treatment response in CC. Both KMT2C and C1QTNF hold promise as potential therapeutic targets and prognostic markers in CC. Further research into their specific roles in the disease, as well as clinical studies evaluating their utility in guiding treatment decisions and predicting patient outcomes, is needed to fully harness their therapeutic and prognostic value.

Significance of the study:

- From our study phylogenetic analysis enriches the understanding of genes E6 and E7 involved in HPV replication that cause metastasis, and the HPV types and sequences reveal the molecular and evolutionary change.

- Genotyping is the association between gene and gene variants of a disease. Here, DNA HPV genotyping shows us effective results of HPV 6, 16, 18, 33 and this method is more effective in preventing cervical cancer than visual inspection or cytology results.

- Our Methylation studies of KMT2C gene from TCGA data play an important role as promoter region of this tumor suppressor gene causes silencing and induce mutation, KMT2C is a missense mutation, and it is called as lysine 3 methyl transferase

- CCs potentially have an increased rate of oncogenic mutations. The WES is used to find out the clinical phenotypes, genetic variants and cause of a disease, and to understand the molecular mechanisms of the genes to find the new diagnostic and therapeutic targets of CC.

- Our comprehensive WES analysis through bioinformatics lays the foundation for understanding the novel pathogenic mutations from the Indian perspective.

- In our study of 5 tumors and 5 surrounding tissue samples, we have attempted to identify mutations that are potent pathogenic, and novel and correlated the data with transcriptomics, and also performed sanger validation. Among them, KMT2C is observed as one of the 10 ten mutated genes in CC specific to the Indian phenotype of CC.

- In the South East Indian cohort, it is the first study about KMT2C in cervical cancer.

- All the Top genes are involved in transcription factors
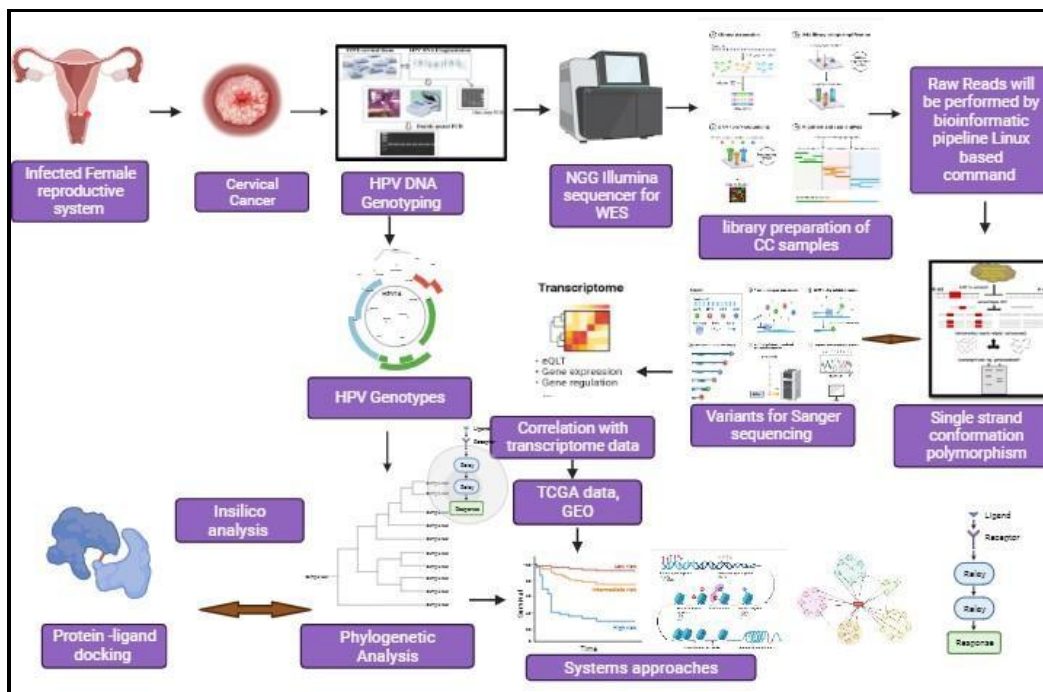
Figure 85: Schematic representation of overall work flow of WES, *In silico* and Transcriptomics

Future Perspectives:

- Continued research may focus on understanding the prognostic implications of KMT2C mutations in cervical cancer. Identifying whether the mutation correlates with the specific clinical outcomes could help in tailoring treatment plans for patients.

- Efforts might be directed towards developing targeted therapies that specifically address the molecular pathways affected by *KMT2C, C1QTNF, EPHB1, MGST1*, etc mutations Precision medicine approaches could become more prominent this can aim to develop treatments tailored to the genetic profile of individual patients.

- We can explore the possible role of immunotherapies in the context of these mutations in another avenue. Understanding how these mutations interact with the immune system could lead to the development of novel immunotherapeutic strategies

- Participation in clinical trials is crucial for advancing practice and could aim to identify WES mutations from this study KMT2C mutation early on, this may facilitate personalized treatment approaches based on the precise genetic makeup of each patient's tumor.

Bibliography

Acharya, R., Mahapatra, A., Verma, H. K., & Bhaskar, L. V. K. S. (2023). Unveiling Therapeutic Targets for Esophageal Cancer: A Comprehensive Review. *Current Oncology*, *30*(11), 9542-9568. https://doi.org/10.3390/curroncol30110691

Aggarwal, R., Mathe, P., & Sharma, P. (2023). A cross-sectional study to assess the knowledge, attitudes, and practices of cervical cancer screening/pap smear among health care professionals in a tertiary care hospital in north India. *International Journal of Reproduction, Contraception, Obstetrics and Gynecology*, *12*(12), 3484-3491.

Anton, M., Horký, M., Kuchtícková, S., Vojtěsek, B., & Bláha, O. (2004). Immunohistochemical detection of acetylation and phosphorylation of histone H3 in cervical smears. Ceska Gynekologie, 69(1), 3–6. PMID: 15112379

Apgar, B. S., Zoschnick, L., & Wright Jr, T. C. (2003). The 2001 Bethesda system terminology. *American family physician*, *68*(10), 1992-1999.

Arney, Ashley, and Katie M. Bennett. "Molecular diagnostics of human papillomavirus." Laboratory Medicine 41.9 (2010): 523-530.

Azimi, T., Paryan, M., Mondanizadeh, M., Sarmadian, H., & Zamani, A. (2021). Pap Smear miR-92a-5p and miR-155-5p as potential diagnostic biomarkers of squamous intraepithelial cervical cancer. Asian Pacific Journal of Cancer Prevention, 22(4), 1271–1277. https://doi.org/10.31557/APJCP.2021.22.4.1271

Badillo Almaraz, J. I. (2014). *Infecciones por papilomavirus humanos (VPH) 16/18 en pacientes con cáncer pulmonar en México* (Doctoral dissertation, Universidad Autónoma de Nuevo León).

Bao, C., An, N., Xie, H., Xu, L., Zhou, B., Luo, J., ... & Huang, J. (2021). Identifying potential neoantigens for cervical cancer immunotherapy using comprehensive genomic variation profiling of cervical intraepithelial neoplasia and cervical

cancer. *Frontiers in Oncology*, *11*, 672386. https://doi.org/10.3389/fonc.2021.672386

Bartha, Á., & Győrffy, B. (2019). Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology. Cancers, 11(11). https://doi.org/10.3390/cancers11111725

Baylin, S. B. (2004, April). Reversal of gene silencing as a therapeutic target for cancer—roles for DNA methylation and its interdigitation with chromatin. In *Reversible Protein Acetylation: Novartis Foundation Symposium 259* (Vol. 259, pp. 226-237). Chichester, UK: John Wiley & Sons, Ltd.

Behrouzkia, Z., Joveini, Z., Keshavarzi, B., Eyvazzadeh, N., & Aghdam, R. Z. (2016). Hyperthermia: How Can It Be Used? Oman Medical Journal, 31(2), 89–97. https://doi.org/10.5001/omj.2016.19

Beniczky, S., Karoly, P., Nurse, E., Ryvlin, P., & Cook, M. (2021). Machine learning and wearable devices of the future. Epilepsia, 62(S2), S116–S124. https://doi.org/https://doi.org/10.1111/epi.16555

Bettadapura, A. D. S., Munivenkatappa, V. K., & Madhunapantula, S. R. V. (2018). Transcriptome Analysis Identified Elevated Expression of Bone Gamma-carboxyglutamic acid-containing Protein (BGLAP) in Human Cervical Cancer Tissues. *Biomedical and Pharmacology Journal*, *11*(2), 1119-1126.https://doi.org/10.13005/bpj/1472

Bhatla, N., Berek, J. S., Cuello Fredes, M., Denny, L. A., Grenman, S., Karunaratne, K., ... & Sankaranarayanan, R. (2019). Revised FIGO staging for carcinoma of the cervix uteri. *International Journal of Gynecology & Obstetrics*, *145*(1), 129-135. https://doi.org/10.1002/ijgo.12749

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & Development, 16(1), 6–21. https://doi.org/10.1101/gad.947102

Bornelöv, S., Reynolds, N., Xenophontos, M., Gharbi, S., Johnstone, E., Floyd, R., Ralser, M., Signolet, J., Loos, R., Dietmann, S., Bertone, P., & Hendrich, B. (2018). The Nucleosome Remodeling and Deacetylation Complex Modulates Chromatin Structure at Sites of Active Transcription to Fine-Tune Gene Expression. Molecular Cell, 71(1), 56-72. https://doi.org/10.1016/j.molcel.2018.06.003

Bossler, F., Hoppe-Seyler, K., & Hoppe-Seyler, F. (2019). PI3K/AKT/mTOR signaling regulates the virus/host cell crosstalk in HPV-positive cervical cancer cells. *International journal of molecular sciences*, *20*(9), 2188. https://doi.org/10.3390/ijms20092188

Bruni, L., Albero, G., Serrano, B., Mena, M., Collado, J., Gómez, D., Muñoz, J., Bosch, F., & de Sanjosé, S. (2023). Human Papillomavirus and Related Diseases in the World. Summary Report 10 March 2023ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). March. www.hpvcentre.net

Bruno, M. T., Cassaro, N., Bica, F., & Boemi, S. (2021). Progression of CIN1/LSIL HPV Persistent of the Cervix: Actual Progression or CIN3 Coexistence. Infectious Diseases in Obstetrics and Gynecology, 2021, 6627531. https://doi.org/10.1155/2021/6627531

Bu, Q., Wu, H. Y., Wang, S. F., Ma, J., Lei, W., Li, Y. H., ... & Luo, X. P. (2020). Methylated Septin9: biomarker for diagnosis of cervical cancer in cervical scrapings and for prediction of pelvic lymphatic metastasis in plasma. https://doi.org/10.21203/rs.3.rs-22142/v1

Burd, E. M. (2003). Human papillomavirus and cervical cancer. Clinical Microbiology Reviews, 16(1), 1–17. https://doi.org/10.1128/CMR.16.1.1-17.2003

Burk, R. D., Chen, Z., Saller, C., Tarvin, K., Carvalho, A. L., Scapulatempo-Neto, C., Silveira, H. C., Fregnani, J. H., Creighton, C. J., Anderson, M. L., Castro, P., Wang, S. S., Yau, C., Benz, C., Robertson, A. G., Mungall, K., Lim, L., Bowlby, R.,

Sadeghi, S., … Hill, U. of N. C. at C. (2017). Integrated genomic and molecular characterization of cervical cancer. Nature, 543(7645), 378–384. https://doi.org/10.1038/nature21386

Byun, J. M., Jeong, D. H., Kim, Y. N., Jung, E. J., Lee, K. B., Sung, M. S., & Kim, K. T. (2018). Persistent HPV-16 infection leads to recurrence of high-grade cervical intraepithelial neoplasia. *Medicine*, *97*(51), e13606. *DOI:* 10.1097/MD.0000000000013606

Carbone, D. P., Reck, M., Paz-Ares, L., Creelan, B., Horn, L., Steins, M., Felip, E., van den Heuvel, M. M., Ciuleanu, T.-E., Badin, F., Ready, N., Hiltermann, T. J. N., Nair, S., Juergens, R., Peters, S., Minenza, E., Wrangle, J. M., Rodriguez-Abreu, D., Borghaei, H., … Socinski, M. A. (2017). First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. The New England Journal of Medicine, 376(25), 2415–2426. https://doi.org/10.1056/NEJMoa1613493

Castle, P. E., Murokora, D., Perez, C., Alvarez, M., Quek, S. C., & Campbell, C. (2017). Treatment of cervical intraepithelial lesions. *International Journal of Gynecology & Obstetrics*, *138*, 20-25.https://doi.org/10.1002/ijgo.12191

Chan, T.-F., Su, T.-H., Yeh, K.-T., Chang, J.-Y., Lin, T.-H., Chen, J.-C., Yuang, S.-S. F., & Chang, J.-G. (2003). Mutational, epigenetic and expressional analyses of caveolin-1 gene in cervical cancers. International Journal of Oncology, 23(3), 599–604. https://doi.org/10.3892/ijo.23.3.599

Chen, C. L., Liu, S. S., Ip, S. M., Wong, L. C., Ng, T. Y., & Ngan, H. Y. S. (2003). E-cadherin expression is silenced by DNA methylation in cervical cancer cell lines and tumours. European Journal of Cancer (Oxford, England :1990), 39(4), 517–523. https://doi.org/10.1016/s0959-8049(02)00175-2

Chen, C. L., Hsu, S. C., Ann, D. K., Yen, Y., & Kung, H. J. (2021). Arginine signaling and cancer metabolism. *Cancers*, *13*(14), 3541.https://doi.org/10.3390/cancers13143541

Chen, X., Wu, Q.-F., & Yan, G.-Y. (2017). RKNNMDA: Ranking-based KNN for MiRNA-Disease Association prediction. RNA Biology, 14(7), 952–962. https://doi.org/10.1080/15476286.2017.1312226

Cheung, T.-H., Lo, K. W.-K., Yim, S.-F., Chan, L. K.-Y., Heung, M.-S., Chan, C.-S., Cheung, A. Y.-K., Chung, T. K.-H., & Wong, Y.-F. (2004). Epigenetic and genetic alternation of PTEN in cervical neoplasm. Gynecologic Oncology, 93(3), 621–627. https://doi.org/10.1016/j.ygyno.2004.03.013

Choi, K. H., Park, S. M., Park, J. S., Park, J. H., Kim, K. H., & Kim, M. J. (2013). Prevalence of and factors associated with osteoporosis among Korean cancer survivors: a cross-sectional analysis of the Fourth and Fifth Korea National Health and Nutrition Examination Surveys. *Asian Pacific Journal of Cancer Prevention*, *14*(8), 4743-4750.

Chu, C. S., Lee, N. P., Adeoye, J., Thomson, P., & Choi, S.-W. (2020). Machine learning and treatment outcome prediction for oral cancer. Journal of Oral Pathology & Medicine : Official Publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology, 49(10), 977–985. https://doi.org/10.1111/jop.13089

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E. S., & Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology, 31(3), 213–219. https://doi.org/10.1038/nbt.2514

Cohen, Y., Singer, G., Lavie, O., Dong, S. M., Beller, U., & Sidransky, D. (2003). The RASSF1A tumor suppressor gene is commonly inactivated in adenocarcinoma of

the uterine cervix. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 9(8), 2981–2984.

Cornall, A. M., Roberts, J. M., Garland, S. M., Hillman, R. J., Grulich, A. E., & Tabrizi, S. N. (2013). Anal and perianal squamous carcinomas and high-grade intraepithelial lesions exclusively associated with "low-risk" HPV genotypes 6 and 11. *International Journal of Cancer*, *133*(9), 2253-2258.https://doi.org/10.1002/ijc.28228

Coulthard, V. H. (2003). *Nuclear hormone receptor-specific interactions of the coactivator trap220*. University of Leicester (United Kingdom).

Cristescu, R., Mogg, R., Ayers, M., Albright, A., Murphy, E., Yearley, J., ... & Kaufman, D. (2018). Pan-tumor genomic biomarkers for PD-1 checkpoint blockade–based immunotherapy. *Science*, *362*(6411),eaar3593.https://doi.org/10.1126/science.aar3593

Curtis, C. R., Dorell, C., Yankey, D., Jeyarajah, J., Chesson, H., Saraiya, M., ... & Centers for Disease Control and Prevention (CDC). (2014). National human papillomavirus vaccination coverage among adolescents aged 13–17 years—National Immunization Survey-Teen, United States, 2011. *MMWR Surveill Summ*, *63*(Suppl 2), 61-70. https://www.cdc.gov/mmwr/preview/mmwrhtml/su6302a10.htm

Das, P., Bansal, A., Rao, S. N., Deodhar, K., Mahantshetty, U., Shrivastava, S. K., Sivaraman, K., & Mulherkar, R. (2016). Somatic Variations in Cervical Cancers in Indian Patients. PLOS ONE, 11(11), e0165878. https://doi.org/10.1371/journal.pone.0165878

Deepti, P., Pasha, A., Kumbhakar, D. V., Doneti, R., Heena, S. K., Bhanoth, S., ... & Pawar, S. C. (2022). Overexpression of Secreted Phosphoprotein 1 (SPP1) predicts poor survival in HPV positive cervical cancer. *Gene*, *824*, 146381. https://doi.org/10.1016/j.gene.2022.146381

De Tomasi, J. B., Opata, M. M., & Mowa, C. N. (2019). Immunity in the cervix: interphase between immune and cervical epithelial cells. *Journal of immunology research*, *2019*. https://doi.org/10.1155/2019/7693183

Dong, S. M., Kim, H. S., Rha, S. H., & Sidransky, D. (2001). Promoter hypermethylation of multiple genes in carcinoma of the uterine cervix. Clinical Cancer Research : An Official Journal of the American Association for Cancer Research, 7(7), 1982–1986.

Drolet, M., Laprise, J. F., Martin, D., Jit, M., Bénard, É., Gingras, G., Boily, M. C., Alary, M., Baussano, I., Hutubessy, R., & Brisson, M. (2021). Optimal human papillomavirus vaccination strategies to prevent cervical cancer in low-income and middle-income countries in the context of limited resources: a mathematical modelling analysis. The Lancet Infectious Diseases, 21(11), 1598–1610. https://doi.org/10.1016/S1473-3099(20)30860-4

Drolet, M., Laprise, J. F., Martin, D., Jit, M., Bénard, É., Gingras, G., ... & Brisson, M. (2021). Optimal human papillomavirus vaccination strategies to prevent cervical cancer in low-income and middle-income countries in the context of limited resources: a mathematical modelling analysis. *The Lancet Infectious Diseases*, *21*(11), 1598-1610.https://doi.org/10.1016/S1473-3099(20)30860-4.

Duppala, S. K., Yadala, R., Velingkar, A., Suravajhala, P., Pawar, S. C., & Vuree, S. (2022). Integrative Multi-Omics Approaches for Identifying Cervical Cancer Therapeutic Targets. *bioRxiv*, 2022-10. https://doi.org/10.1101/2022.10.07.511244

Dutta, S., Begum, R., Mazumder, D., Mandal, S. S., Mondal, R., Biswas, J., ... & Basu, P. (2012). Prevalence of human papillomavirus in women without cervical cancer: a population-based study in Eastern India. *International journal of gynecological pathology*, *31*(2), 178-183.doi: 10.1097/PGP.0b013e3182399391

Dy, G. K., Nesline, M. K., Papanicolau-Sengos, A., DePietro, P., LeVea, C. M., Early, A., ... & Morrison, C. (2019). Treatment recommendations to cancer patients in the context of FDA guidance for next generation sequencing. *BMC Medical Informatics and Decision Making*, *19*, 1-10. https://doi.org/10.1186/s12911-019-0743-x

Efdi, M., Itoh, T., Akao, Y., Nozawa, Y., Koketsu, M., & Ishihara, H. (2007). The isolation of secondary metabolites and in vitro potent anti-cancer activity of clerodermic acid from Enicosanthum membranifolium. *Bioorganic & medicinal chemistry*, *15*(11), 3667-3671.https://doi.org/10.1016/j.bmc.2007.03.051

Eifel, P. J., Ho, A., Khalid, N., Erickson, B., & Owen, J. (2014). Patterns of radiation therapy practice for patients treated for intact cervical cancer in 2005 to 2007: a quality research in radiation oncology study. *International Journal of Radiation Oncology* Biology* Physics*, *89*(2), 249-256.https://doi.org/10.1016/j.ijrobp.2013.11.228

El Husseini, Z. Full Title: Use of RNA-Seq to detect and characterise splicing abnormalities in BRCA1 and BRCA2 Short Title: RNA-Seq for BRCA1/BRCA2 splicing abnormalities.

Elit, L., Fyles, A. W., Oliver, T. K., Devries–Aboud, M. C., Fung-Kee-Fung, M., & Gynecology Cancer Disease Site Group of Cancer Care Ontario's Program in Evidence-Based Care. (2010). Follow-up for women after treatment for cervical cancer. *Current Oncology*, *17*(3), 65-69. https://doi.org/10.3747/co.v17i3.514

Elit, L., Fyles, A. W., Oliver, T. K., Devries–Aboud, M. C., Fung-Kee-Fung, M., & Gynecology Cancer Disease Site Group of Cancer Care Ontario's Program in Evidence-Based Care. (2010). Follow-up for women after treatment for cervical cancer. *Current Oncology*, *17*(3), 65-69. https://doi.org/10.3747/co.v17i3.514

El-Zein, M., Cheishvili, D., Gotlieb, W., Gilbert, L., Hemmings, R., Behr, M. A., Szyf, M., Franco, E. L., & group, for the M. study. (2020). Genome-wide DNA methylation profiling identifies two novel genes in cervical neoplasia. International Journal of Cancer, 147(5), 1264–1274. https://doi.org/https://doi.org/10.1002/ijc.32880

Feller, L., Khammissa, R. A., Wood, N. H., & Lemmer, J. (2009). Epithelial maturation and molecular biology of oral HPV. *Infectious agents and cancer*, *4*, 1-9. doi:10.1186/1750-9378-4-16.

Feng, Q., Balasubramanian, A., Hawes, S. E., Toure, P., Sow, P. S., Dem, A., Dembele, B., Critchlow, C. W., Xi, L., Lu, H., McIntosh, M. W., Young, A. M., & Kiviat, N. B. (2005). Detection of hypermethylated genes in women with and without cervical neoplasia. Journal of the National Cancer Institute, 97(4), 273–282. https://doi.org/10.1093/jnci/dji041

Ferlay, J. "Globocan. International Agency for Research on Cancer." World Health Organisation, Genf (2001). http://gco.iarc.fr/.

Fraga, M. F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., Iyer, N. G., Pérez-Rosado, A., Calvo, E., Lopez, J. A., Cano, A., Calasanz, M. J., Colomer, D., Piris, M. A., Ahn, N., … Esteller, M. (2005). Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nature Genetics, 37(4), 391–400. https://doi.org/10.1038/ng1531

Gaddi, M. S. (2018). *Diagnostic Accuracy of Papanicolaou's Smear in Detection of Cytological Abnormalities of Cervix* (Doctoral dissertation, Rajiv Gandhi University of Health Sciences (India)).

Gao, L., Zhang, L. J., Li, S. H., Wei, L. L., Luo, B., He, R. Q., & Xia, S. (2018). Role of miR-452-5p in the tumorigenesis of prostate cancer: a study based on the Cancer Genome Atl (TCGA), Gene Expression Omnibus (GEO), and bioinformatics

analysis. *Pathology-Research and Practice*, *214*(5), 732-749.https://doi.org/10.1016/j.prp.2018.03.002

Gao, Y., Ma, J. L., Gao, F., & Song, L. P. (2013). The evaluation of older patients with cervical cancer. *Clinical interventions in aging*, 783-788.https://doi.org/10.2147/CIA.S45613

Garrido-Martín, D., Palumbo, E., Guigó, R., & Breschi, A. (2018). ggsashimi: Sashimi plot revised for browser-and annotation-independent splicing visualization. *PLoS computational biology*, *14*(8), e1006360. https://doi.org/10.1371/journal.pcbi.1006360.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*. https://doi.org/10.48550/arXiv.1207.3907

Gasser, R. B., Hu, M., Chilton, N. B., Campbell, B. E., Jex, A. J., Otranto, D., ... & Zhu, X. (2006). Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. *Nature protocols*, *1*(6), 3121-3128.https://doi.org/10.1038/nprot.2006.485

George, J. T., & Batra, K. (2015). Major determinants and various preventive strategies of cervical cancer. *Asian Journal of Nursing Education and Research*, *5*(3), 420-424.Doi:10.5958/2349-2996.2015.00083.X

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International journal of Ayurveda research*, *1*(4), 274.doi:10.4103/0974-7788.76794.

Golding, B., Morton, D., & Haerty, W. (2000). Elementary Sequence Analysis. *Multiple Sequence Alignments*.

Goodrich, S. K., Schlegel, C. R., Wang, G., & Belinson, J. L. (2014). Use of artemisinin and its derivatives to treat HPV-infected/transformed cells and cervical cancer: a review. *Future Oncology*, *10*(4), 647-654.https://doi.org/10.2217/fon.13.228

Gupta, A., Shukla, N., Mishra, A. K., Batra, J., Lohiya, N. K., & Suravajhala, P. (2020). A pilot study on the whole exome sequencing of prostate cancer in the indian phenotype reveals distinct polymorphisms. *Frontiers in Genetics*, *11*, 544348. https://doi.org/10.3389/fgene.2020.00874

Harlfinger, J., Sommer, I., & Gartlehner, G. (2023). WHO Guideline for Screening and Treatment of Cervical Pre-Cancerous Lesions for Cervical Cancer Prevention (Second Edition). In Gesundheitswesen (Vol. 85, Issue 7). https://doi.org/10.1055/a-2052-6652

Harris, M., Wang, X. G., Jiang, Z., Phaeton, R., Koba, W., Goldberg, G. L., Casadevall, A., & Dadachova, E. (2013). Combined treatment of the experimental human papilloma virus-16-positive cervical and head and neck cancers with cisplatin and radioimmunotherapy targeting viral E6 oncoprotein. British Journal of Cancer, 108(4), 859–865. https://doi.org/10.1038/bjc.2013.43

He, S., Shi, J., Mao, J., Luo, X., Liu, W., Liu, R., & Yang, F. (2019). The expression of miR-375 in prostate cancer: A study based on GEO, TCGA data and bioinformatics analysis. *Pathology-Research and Practice*, *215*(6), 152375.https://doi.org/10.1016/j.prp.2019.03.004

Hellmann, M. D., Nathanson, T., Rizvi, H., Creelan, B. C., Sanchez-Vega, F., Ahuja, A., Ni, A., Novik, J. B., Mangarin, L. M. B., Abu-Akeel, M., Liu, C., Sauter, J. L., Rekhtman, N., Chang, E., Callahan, M. K., Chaft, J. E., Voss, M. H., Tenet, M., Li, X.-M., … Wolchok, J. D. (2018). Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. Cancer Cell, 33(5), 843-852.e4. https://doi.org/10.1016/j.ccell.2018.03.018

Herfs, M., Yamamoto, Y., Laury, A., Wang, X., Nucci, M. R., McLaughlin-Drubin, M. E., Münger, K., Feldman, S., McKeon, F. D., Xian, W., & Crum, C. P. (2012). A discrete population of squamocolumnar junction cells implicated in the pathogenesis of cervical cancer. Proceedings of the National Academy of Sciences, 109(26), 10516–10521. https://doi.org/10.1073/pnas.1202684109

Höckel, M., Hentschel, B., & Horn, L.-C. (2014). Association between developmental steps in the organogenesis of the uterine cervix and locoregional progression of cervical cancer: a prospective clinicopathological analysis. The Lancet Oncology, 15(4), 445–456. https://doi.org/10.1016/S1470-2045(14)70060-9

Hoots, B. E., Palefsky, J. M., Pimenta, J. M., & Smith, J. S. (2009). Human papillomavirus type distribution in anal cancer and anal intraepithelial lesions. *International journal of cancer*, *124*(10), 2375-2383..https://doi.org/10.1002/ijc.24215

Huang, J., Liu, T., Shang, C., Zhao, Y., Wang, W., Liang, Y., ... & Yao, S. (2018). Identification of lncRNAs by microarray analysis reveals the potential role of lncRNAs in cervical cancer pathogenesis. *Oncology letters*, *15*(4), 5584-5592.https://doi.org/10.3892/ol.2018.8037.

Huang, Y., Zhu, Q., Xue, L., Zhu, X., Chen, Y., & Wu, M. (2022). Machine learning-assisted ensemble analysis for the prediction of response to neoadjuvant chemotherapy in locally advanced cervical cancer. *Frontiers in Oncology*, *12*, 817250. https://doi.org/10.3389/fonc.2022.817250

Huang, Y., Hou, Y., Qu, P., & Cai, Y. (2021). Combating Cancer With Natural Products: What Would Non-Coding RNAs Bring? *Frontiers in Oncology*, *11*, 747586.https://doi.org/10.3389/fonc.2021.747586

Hugo, W., Zaretsky, J. M., Sun, L. U., Song, C., Moreno, B. H., Hu-Lieskovan, S., ... & Lo, R. S. (2016). Genomic and transcriptomic features of response to anti-PD-1

therapy in metastatic melanoma. *Cell*, *165*(1), 35-44. https://doi.org/10.1016/j.cell.2016.02.065

Humphrey, P., Dures, E., Hoskin, P., & Cramp, F. (2024). Patient experiences of brachytherapy for locally advanced cervical cancer: hearing the patient voice through qualitative interviews. *International Journal of Radiation Oncology* Biology* Physics*, *119*(3), 902-911. https://doi.org/10.1016/j.ijrobp.2023.12.016

IJff, M., Crezee, J., Oei, A. L., Stalpers, L. J. A., & Westerveld, H. (2022). The role of hyperthermia in the treatment of locally advanced cervical cancer: a comprehensive review. International Journal of Gynecological Cancer: Official Journal of the International Gynecological Cancer Society, 32(3), 288–296. https://doi.org/10.1136/ijgc-2021-002473

Ivanova, T., Petrenko, A., Gritsko, T., Vinokourova, S., Eshilev, E., Kobzeva, V., Kisseljov, F., & Kisseljova, N. (2002). Methylation and silencing of the retinoic acid receptor-β2 gene in cervical cancer. BMC Cancer, 2(1), 4. https://doi.org/10.1186/1471-2407-2-4

Ivanova, T., Vinokurova, S., Petrenko, A., Eshilev, E., Solovyova, N., Kisseljov, F., & Kisseljova, N. (2004). Frequent hypermethylation of 5' flanking region of TIMP-2 gene in cervical cancer. International Journal of Cancer, 108(6), 882–886. https://doi.org/10.1002/ijc.11652

Jass, J. R. (2004). HNPCC and sporadic MSI-H colorectal cancer: a review of the morphological similarities and differences. Familial Cancer, 3(2), 93–100. https://doi.org/10.1023/B:FAME.0000039849.86008.b7

Jayasinghe, Y., Rangiah, C., Gorelik, A., Ogilvie, G., Wark, J. D., Hartley, S., & Garland, S. M. (2016). Primary HPV DNA based cervical cancer screening at 25 years: Views of young Australian women aged 16–28 years. *Journal of Clinical Virology*, *76*, S74-S80. https://doi.org/10.1016/j.jcv.2015.10.026

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. CA: A Cancer Journal for Clinicians, 61(2), 69–90. https://doi.org/https://doi.org/10.3322/caac.20107

Jiang, S., Wang, H. L., & Yang, J. (2015). Low expression of long non-coding RNA LET inhibits carcinogenesis of cervical cancer. *International journal of clinical and experimental pathology*, *8*(1), 806.

Jing, L., Yuan, W., Ruofan, D., Jinjin, Y., & Haifeng, Q. (2015). HOTAIR enhanced aggressive biological behaviors and induced radio-resistance via inhibiting p21 in cervical cancer. *Tumor Biology*, *36*, 3611-3619.https://doi.org/10.1007/s13277-014-2998-2

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30. doi:10.1093/NAR/28.1.27.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, *45*(D1), D353-D361. https://doi.org/10.1093/nar/gkw1092

Kang, S., Kim, J. W., Kang, G. H., Park, N. H., Song, Y. S., Kang, S. B., & Lee, H. P. (2005). Polymorphism in folate- and methionine-metabolizing enzyme and aberrant CpG island hypermethylation in uterine cervical cancer. Gynecologic Oncology, 96(1), 173–180. https://doi.org/10.1016/j.ygyno.2004.09.031

Kausar Neyaz, M., Suresh Kumar, R., Hussain, S., Naqvi, S., Kohaar, I., Thakur, N., Kashyap, V., Das, B., Akhtar Husain, S., & Bharadwaj, M. (2008). Effect of aberrant promoter methylation of FHIT and RASSF1A genes on susceptibility to cervical cancer in a North Indian population. Biomarkers, 13(6), 597–606. https://doi.org/10.1080/13547500802078859

Khorsandi, D. (2023). Development of microfluidic platforms: human uterine cervix-on-a-chip. *Materials Chemistry Horizons*, *2*(1), 1-9. 10.22128/MCH.2022.605.1026.

Kloku, C. A. (2015). *Awareness and prevention of cervical cancer among female health professionals: A study of three health institutions in Winneba, Ghana* (Doctoral dissertation).

Knowles, L. L. (2009). Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Systematic biology*, *58*(5), 463-467.https://doi.org/10.1093/sysbio/syp061

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., & Wilson, R. K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research, 22(3), 568–576. https://doi.org/10.1101/gr.129684.111

Kundrod, K. A. (2020). *Point-of-Care Tests to Amplify and Detect High-Risk HPV DNA and mRNA* (Doctoral dissertation, Rice University).

Kuzmin, I., Liu, L., Dammann, R., Geil, L., Stanbridge, E. J., Wilczynski, S. P., ... & Pfeifer, G. P. (2003). Inactivation of RAS association domain family 1A gene in cervical carcinomas and the role of human papillomavirus infection. *Cancer research*, *63*(8), 1888-1893.

Lacroix, G., Gouyer, V., Gottrand, F., & Desseyn, J. L. (2020). The cervicovaginal mucus barrier. *International journal of molecular sciences*, *21*(21), 8266. https://doi.org/10.3390/ijms21218266

Larson, D. E., Harris, C. C., Chen, K., Koboldt, D. C., Abbott, T. E., Dooling, D. J., Ley, T. J., Mardis, E. R., Wilson, R. K., & Ding, L. (2012). SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics (Oxford, England), 28(3), 311–317. https://doi.org/10.1093/bioinformatics/btr665

Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., Skora, A. D., Luber, B. S., Azad, N. S., Laheru, D., Biedrzycki, B., Donehower, R. C., Zaheer, A., Fisher, G. A., Crocenzi, T. S., Lee, J. J., Duffy, S. M., Goldberg, R. M., de la

Chapelle, A., … Diaz, L. A. J. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. The New England Journal of Medicine, 372(26), 2509–2520. https://doi.org/10.1056/NEJMoa1500596

Lees, B. F., Erickson, B. K., & Huh, W. K. (2016). Cervical cancer screening: evidence behind the guidelines. *American journal of obstetrics and gynecology*, *214*(4), 438-443. https://doi.org/10.1016/j.ajog.2015.10.147

Leung, K. H., & Yip, S. P. (2008). Single Strand Conformation Polymorphism (SSCP) Analysis. *Molecular biomethods handbook*, 117-131.

Li, J., Zhang, Z., Bidder, M., Funk, M. C., Nguyen, L., Goodfellow, P. J., & Rader, J. S. (2005). IGSF4 promoter methylation and expression silencing in human cervical cancer. Gynecologic Oncology, 96(1), 150–158. https://doi.org/10.1016/j.ygyno.2004.08.050

Li, Q., Zhu, G., Zhang, L., Zeng, B., Cai, T., Wu, J., ... & Fan, S. (2023). Mutational landscape of head and neck cancer and cervical cancer in Chinese and Western population. *Head & Neck*.https://doi.org/10.1002/hed.27603

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., ... & Liu, X. S. (2017). TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer research*, *77*(21), e108-e110. https://doi.org/10.1158/0008-5472.CAN-17-0307

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., ... & Liu, X. S. (2020). TIMER2. 0 for analysis of tumor-infiltrating immune cells. *Nucleic acids research*, *48*(W1), W509-W514.https://doi.org/10.1093/nar/gkaa407

Li, X., Tian, R., Gao, H., Yang, Y., Williams, B. R. G., Gantier, M. P., McMillan, N. A. J., Xu, D., Hu, Y., & Gao, Y. (2017). Identification of a histone family gene signature for predicting the prognosis of cervical cancer patients. Scientific Reports, 7(1), 16495. https://doi.org/10.1038/s41598-017-16472-5

Lifson, A. R., Rhame, F. S., Belloso, W. H., Dragsted, U. B., El-Sadr, W. M., Gatell, J. M., Hoy, J. F., Krum, E. A., Nelson, R., Pedersen, C., Pett, S. L., & Davey, R. T. (2006). Reporting and Evaluation of HIV-Related Clinical Endpoints in Two Multicenter International Clinical Trials. HIV Clinical Trials, 7(3), 125–141. https://doi.org/10.1310/7MER-XFA7-1762-E2WR

Lin, H.-H., Zhang, Q.-R., Kong, X., Zhang, L., Zhang, Y., Tang, Y., & Xu, H. (2021). Machine learning prediction of antiviral-HPV protein interactions for anti-HPV pharmacotherapy. Scientific Reports, 11(1), 24367. https://doi.org/10.1038/s41598-021-03000-9

Liu, J., Li, Z., Lu, T., Pan, J., Li, L., Song, Y., ... & Xu, Q. (2022). Genomic landscape, immune characteristics and prognostic mutation signature of cervical cancer in China. *BMC Medical Genomics*, *15*(1), 231.https://doi.org/10.1186/s12920-022-01376-9

Lloyd, P. A., Briggs, E. V., Kane, N., Jeyarajah, A. R., & Shepherd, J. H. (2014). Women's experiences after a radical vaginal trachelectomy for early stage cervical cancer. A descriptive phenomenological study. *European Journal of Oncology Nursing*, *18*(4), 362-371. https://doi.org/10.1016/j.ejon.2014.03.014

Longworth, M. S., & Laimins, L. A. (2004). The binding of histone deacetylases and the integrity of zinc finger-like motifs of the E7 protein are essential for the life cycle of human papillomavirus type 31. *Journal of virology*, *78*(7), 3533-3541.https://doi.org/10.1128/jvi.78.7.3533-3541.2004

Lorenzi, A. T., Syrjänen, K. J., & Longatto-Filho, A. (2015). Human papillomavirus (HPV) screening and cervical cancer burden. A Brazilian perspective. *Virology journal*, *12*, 1-6. https://doi.org/10.1186/s12985-015-0342-0

Lüscher-Firzlaff, J., Chatain, N., Kuo, C. C., Braunschweig, T., Bochyńska, A., Ullius, A., ... & Lüscher, B. (2019). Hematopoietic stem and progenitor cell proliferation and

differentiation requires the trithorax protein Ash2l. *Scientific reports*, *9*(1), 8262.https://doi.org/10.1038/s41598-019-44720-3

Markovic, N., & Markovic, O. (2008). *What every woman should know about cervical cancer*. Dordrecht: Springer.

Martelli, A. M. (2016). Effects of Mutations in Wnt/β-Catenin, Hedgehog, Notch and PI3K Pathways on GSK-3 Activity—Diverse Effects on Cell Growth, Metabolism and Cancer. James A. McCubrey, Dariusz Rakus b, Agnieszka Gizak b, Linda S. Steelman, Steve L. Abrams, Kvin Lertpiriyapong c, Timothy L. Fitzgerald d, Li V. Yang e, Giuseppe Montalto f, g, Melchiorre Cervello g, Massimo Libra h, Ferdinando Nicoletti h, Aurora Scalisi i, Francesco.

Meister, M. T., Voss, S., & Schwabe, D. (2015). Treatment of EBV-associated nodular sclerosing Hodgkin lymphoma in a patient with ataxia telangiectasia with brentuximab vedotin and reduced COPP plus rituximab. Pediatric Blood & Cancer, 62(11), 2018–2020. https://doi.org/https://doi.org/10.1002/pbc.25621

Meites, E. (2019). Human papillomavirus vaccination for adults: updated recommendations of the Advisory Committee on Immunization Practices. *MMWR. Morbidity and mortality weekly report*, *68*

Melo, I. M. A., Viana, M. R. P., Pupin, B., Bhattacharjee, T. T., & de Azevedo Canevari, R. (2021). PCR-RFLP and FTIR-based detection of high-risk human papilloma virus for cervical cancer screening and prevention. *Biochemistry and Biophysics Reports*, *26*, 100993. https://doi.org/10.1016/j.bbrep.2021.100993

Mendoza-Nava, H., Ferro-Flores, G., Ocampo-García, B., Serment-Guerrero, J., Santos-Cuevas, C., Jiménez-Mancilla, N., Luna-Gutiérrez, M., & Camacho-López, M. A. (2012). Laser Heating of Gold Nanospheres Functionalized with Octreotide: In Vitro Effect on HeLa Cell Viability. Photomedicine and Laser Surgery, 31(1), 17–22. https://doi.org/10.1089/pho.2012.3320

Mering, C. V., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., & Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic acids research*, *31*(1), 258-261.doi:10.1093/nar/gkg034.

Michalas, S. P. (2000). The Pap test: George N. Papanicolaou (1883–1962): a screening test for the prevention of cancer of uterine cervix. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, *90*(2), 135-138. https://doi.org/10.1016/S0301-2115(00)00260-8

Miliotou, N. A., & Papadopoulou, C. L. (2018). CAR T-cell Therapy: A New Era in Cancer Immunotherapy. In Current Pharmaceutical Biotechnology (Vol. 19, Issue 1, pp. 5–18). https://doi.org/http://dx.doi.org/10.2174/1389201019666180418095526

Mitra, A., MacIntyre, D. A., Marchesi, J. R., Lee, Y. S., Bennett, P. R., & Kyrgiou, M. (2016). The vaginal microbiota, human papillomavirus infection and cervical intraepithelial neoplasia: what do we know and where are we going next? Microbiome, 4(1), 58. https://doi.org/10.1186/s40168-016-0203-0

Molano, M., Posso, H., Weiderpass, E., Van den Brule, A. J. C., Ronderos, M., Franceschi, S., ... & Munoz, N. (2002). Prevalence and determinants of HPV infection among Colombian women with normal cytology. *British journal of cancer*, *87*(3), 324-333.doi:10.1038/sj.bjc.6600442.

Molina, M. A., Carosi Diatricch, L., Castany Quintana, M., Melchers, W. J., & Andralojc, K. M. (2020). Cervical cancer risk profiling: molecular biomarkers predicting the outcome of hrHPV infection. Expert Review of Molecular Diagnostics, 20(11), 1099–1120. https://doi.org/10.1080/14737159.2020.1835472

Morris, M., Eifel, P. J., Lu, J., Grigsby, P. W., Levenback, C., Stevens, R. E., Rotman, M., Gershenson, D. M., & Mutch, D. G. (1999). Pelvic Radiation with Concurrent Chemotherapy Compared with Pelvic and Para-Aortic Radiation for High-Risk

Cervical Cancer. New England Journal of Medicine, 340(15), 1137–1143. https://doi.org/10.1056/NEJM199904153401501

Morris, M., Eifel, P. J., Lu, J., Grigsby, P. W., Levenback, C., Stevens, R. E., ... & Mutch, D. G. (1999). Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aortic radiation for high-risk cervical cancer. *New England Journal of Medicine*, *340*(15), 1137-1143.. DOI:10.1056/NEJM199904153401501

Muñoz, N., Bosch, F. X., De Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K. V., ... & Meijer, C. J. (2003). Epidemiologic classification of human papillomavirus types associated with cervical cancer. *New England journal of medicine*, *348*(6), 518-527.https://doi.org/10.1056/NEJMoa021641

Mutuku, O. M. (2020). *The utility of a Manual Liquid Based Cytology in Screening for Pre-cancerous Lesion and Cervical Ca* (Doctoral dissertation, JKUAT COHES). http://localhost/xmlui/handle/123456789/5802

Nagy, Á., Lánczky, A., Menyhárt, O., & Győrffy, B. (2018). Author Correction: Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Scientific reports*, *8*.doi:10.1038/s41598-018-29514-3.

Nam, E. J., Kim, J. W., Hong, J. W., Jang, H. S., Lee, S. Y., Jang, S. Y., Lee, D. W., Kim, S. W., Kim, J. H., Kim, Y. T., Kim, S., & Kim, J. W. (2008). Expression of the p16 and Ki-67 in relation to the grade of cervical intraepithelial neoplasia and high-risk human papillomavirus infection. Journal of Gynecologic Oncology, 19(3), 162–168. https://doi.org/10.3802/jgo.2008.19.3.162

Nandakumar, A., Ramnath, T., & Chaturvedi, M. (2009). The magnitude of cancer cervix in India. The Indian Journal of Medical Research, 130(3), 219–221.

Naucler, P., Ryd, W., Törnberg, S., Strand, A., Wadell, G., Elfgren, K., ... & Dillner, J. (2009). Efficacy of HPV DNA testing with cytology triage and/or repeat HPV DNA

testing in primary cervical cancer screening. *Journal of the National Cancer Institute*, *101*(2), 88-99. https://doi.org/10.1093/jnci/djn444

Neyaz, M. K., & Ahmad, S. (2019). Biomarkers for the Early Detection of Cervical Cancer BT - Preventive Oncology for the Gynecologist (S. Mehta & A. Singla (Eds.); pp. 117–129). Springer Singapore. https://doi.org/10.1007/978-981-13-3438-2_10

Neveling, K. (2007). *Molecular causes and consequences of genetic instability with respect to the FA/BRCA caretaker pathway* (Doctoral dissertation, Universität Würzburg).

Niikura, H., Okamoto, S., Otsuki, T., Yoshinaga, K., Utsunomiya, H., Nagase, S., Takano, T., Ito, K., Watanabe, M., & Yaegashi, N. (2012). Prospective Study of Sentinel Lymph Node Biopsy Without Further Pelvic Lymphadenectomy in Patients With Sentinel Lymph Node–Negative Cervical Cancer. International Journal of Gynecologic Cancer, 22(7), 1244 LP – 1250. https://doi.org/10.1097/IGC.0b013e318263f06a

Nobre, R. J., de Almeida, L. P., & Martins, T. C. (2008). Complete genotyping of mucosal human papillomavirus using a restriction fragment length polymorphism analysis and an original typing algorithm. *Journal of clinical virology*, *42*(1), 13-21. https://doi.org/10.1016/j.jcv.2007.11.021

Onwuka, C. O., & Ekanem, I. A. (2017). The utility of visual inspection with acetic acid in cervical cancer screening. *ECOA*, *103*, 7-14.

Onyango, C. G., Ogonda, L., Guyah, B., Shiluli, C., Ganda, G., Orang'o, O. E., & Patel, K. (2020). Novel biomarkers with promising benefits for diagnosis of cervical neoplasia: a systematic review. Infectious Agents and Cancer, 15(1), 68. https://doi.org/10.1186/s13027-020-00335-2

Oyaga-Iriarte, E., Insausti, A., Sayar, O., & Aldaz, A. (2019). Prediction of irinotecan toxicity in metastatic colorectal cancer patients based on machine learning models with pharmacokinetic parameters. Journal of Pharmacological Sciences, 140(1), 20–25. https://doi.org/10.1016/j.jphs.2019.03.004

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics*, *12*(2), 87-98. https://doi.org/10.1038/nrg2934

Papanicolaou, G. N., & Traut, H. F. (1997). The diagnostic value of vaginal smears in carcinoma of the uterus. 1941. Archives of Pathology & Laboratory Medicine, 121(3), 211–224.

Pardini, B., De Maria, D., Francavilla, A., Di Gaetano, C., Ronco, G., & Naccarati, A. (2018). MicroRNAs as markers of progression in cervical cancer: a systematic review. BMC Cancer, 18(1), 696. https://doi.org/10.1186/s12885-018-4590-4

Parikh, J. H., Barton, D. P., Ind, T. E., & Sohaib, S. A. (2008). MR imaging features of vaginal malignancies. *Radiographics*, *28*(1), 49-63. https://doi.org/10.1148/rg.281075065

Peiretti, M., Zapardiel, I., Zanagnolo, V., Landoni, F., Morrow, C. P., & Maggioni, A. (2012). Management of recurrent cervical cancer: A review of the literature. Surgical Oncology, 21(2), e59–e66. https://doi.org/https://doi.org/10.1016/j.suronc.2011.12.008

Peng, Z., Gong, Y., & Liang, X. (2021). Role of FAT1 in health and disease. *Oncology letters*, *21*(5), 1-13. https://doi.org/10.3892/ol.2021.12659

Perkins, R. B., Wentzensen, N., Guido, R. S., & Schiffman, M. (2023). Cervical cancer screening: a review. *Jama*, *330*(6), 547-558. doi:10.1001/jama.2023.13174

Phippen, N., Leath III, C., Miller, C., Lowery, W., Havrilesky, L., & Barnett, J. (2013). Is a home-based palliative care treatment strategy preferable to standard chemotherapy in recurrent cervical cancer? Gynecologic Oncology, 131(1), 277–278. https://doi.org/10.1016/j.ygyno.2013.04.027

Poondla, N., Madduru, D., Duppala, S. K., Velpula, S., Nunia, V., Kharb, S., ... & Suravajhala, P. (2021). Cervical cancer in the era of precision medicine: A perspective from developing countries. *Advances in Cancer Biology-Metastasis*, *3*, 100015.https://doi.org/10.1016/j.adcanc.2021.100015

Porras, C., Wentzensen, N., Rodríguez, A. C., Morales, J., Burk, R. D., Alfaro, M., ... & Schiffman, M. (2012). Switch from cytology-based to human papillomavirus test-based cervical screening: Implications for colposcopy. *International journal of cancer*, *130*(8), 1879-1887.https://doi.org/10.1002/ijc.26194

Prabhu, T. R. B. (2015). *A Practical Approach to Cervical Cancer Screening Techniques*. JP Medical Ltd.

Prendiville, W., & Sankaranarayanan, R. (2017). *Colposcopy and treatment of cervical precancer*. International Agency for Research on Cancer, World Health Organization.

Rachana, L. Y. (2018). *" Cytopathological Analysis and Comparison of Cervical Smears Using Revised Bethesda System 2014 and Old Bethesda System 2001"-A Case Series* (Doctoral dissertation, Rajiv Gandhi University of Health Sciences (India)).

Ramachandran, D., & Dörk, T. (2021). Genomic Risk Factors for Cervical Cancer. In Cancers (Vol. 13, Issue 20). https://doi.org/10.3390/cancers13205137

Ramzan, M., ul Ain, N., Ilyas, S., Umer, M., Bano, S., Sarwar, S., ... & Shakoori, A. R. (2015). A cornucopia of screening and diagnostic techniques for human papillomavirus associated cervical carcinomas. *Journal of virological methods*, *222*, 192-201. https://doi.org/10.1016/j.jviromet.2015.06.015

Rema, P., & Ahmed, I. (2016). Conservative Surgery for Early Cervical Cancer. Indian Journal of Surgical Oncology, 7(3), 336–340. https://doi.org/10.1007/s13193-015-0476-y

Riaz, N., Havel, J. J., Makarov, V., Desrichard, A., Urba, W. J., Sims, J. S., Hodi, F. S., Martín-Algarra, S., Mandal, R., Sharfman, W. H., Bhatia, S., Hwu, W.-J., Gajewski, T. F., Slingluff, C. L. J., Chowell, D., Kendall, S. M., Chang, H., Shah, R., Kuo, F., … Chan, T. A. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. Cell, 171(4), 934-949.e16. https://doi.org/10.1016/j.cell.2017.09.028

Rizvi, H., Sanchez-Vega, F., La, K., Chatila, W., Jonsson, P., Halpenny, D., Plodkowski, A., Long, N., Sauter, J. L., Rekhtman, N., Hollmann, T., Schalper, K. A., Gainor, J. F., Shen, R., Ni, A., Arbour, K. C., Merghoub, T., Wolchok, J., Snyder, A., … Hellmann, M. D. (2018). Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, 36(7), 633–641. https://doi.org/10.1200/JCO.2017.75.3384

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., Lee, W., Yuan, J., Wong, P., Ho, T. S., Miller, M. L., Rekhtman, N., Moreira, A. L., Ibrahim, F., Bruggeman, C., Gasmi, B., Zappasodi, R., Maeda, Y., Sander, C., … Chan, T. A. (2015). Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science (New York, N.Y.), 348(6230), 124–128. https://doi.org/10.1126/science.aaa1348

Romagosa, C., Simonetti, S., López-Vicente, L., Mazo, A., Lleonart, M. E., Castellvi, J., & Ramon y Cajal, S. (2011). p16(Ink4a) overexpression in cancer: a tumor

suppressor gene associated with senescence and high-grade tumors. Oncogene, 30(18), 2087–2097. https://doi.org/10.1038/onc.2010.614

Santesso, N., Schünemann, H., Blumenthal, P., De Vuyst, H., Gage, J., Garcia, F., Jeronimo, J., Lu, R., Luciani, S., Quek, S. C., Awad, T., & Broutet, N. (2012). World Health Organization Guidelines: Use of cryotherapy for cervical intraepithelial neoplasia. International Journal of Gynecology & Obstetrics, 118(2), 97–102. https://doi.org/https://doi.org/10.1016/j.ijgo.2012.01.029

Sarathi, S., Hemavathy, V., & Vijayalakshmi, R. (2015). Cervical Cancer Kills One Indian Women Every 7 Minutes. International Journal of Innovative Research and Development || ISSN 2278 – 0211, 4(1), 132–134. http://www.ijird.com/index.php/ijird/article/view/57709

Sarnaik, A. (2019). T Cell Therapy Shows Promise for Melanoma and Cervical Cancer. 2, 3–5.

Saslow, D., Solomon, D., Lawson, H. W., Killackey, M., Kulasingam, S. L., Cain, J., ... & Myers, E. R. (2012). American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *American journal of clinical pathology*, *137*(4), 516-542.. doi: 10.3322/caac.21139

Saunders, C. T., Wong, W. S. W., Swamy, S., Becq, J., Murray, L. J., & Cheetham, R. K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics (Oxford, England), 28(14), 1811–1817. https://doi.org/10.1093/bioinformatics/bts271

Schmidt, M. W., Battista, M. J., Schmidt, M., Garcia, M., Siepmann, T., Hasenburg, A., & Anic, K. (2022). Efficacy and safety of immunotherapy for cervical cancer—a systematic review of clinical trials. *Cancers*, *14*(2), 441.https://doi.org/10.3390/cancers14020441

Sen, P., Ghosal, S., Hazra, R., Arega, S., Mohanty, R., Kulkarni, K. K., ... & Ganguly, N. (2020). Transcriptomic analyses of gene expression by CRISPR knockout of miR-214 in cervical cancer cells. *Genomics*, *112*(2), 1490-1499.https://doi.org/10.1016/j.ygeno.2019.08.020.

Shivapurkar, N., Toyooka, S., Toyooka, K. O., Reddy, J., Miyajima, K., Suzuki, M., Shigematsu, H., Takahashi, T., Parikh, G., Pass, H. I., Chaudhary, P. M., & Gazdar, A. F. (2004). Aberrant methylation of trail decoy receptor genes is frequent in multiple tumor types. International Journal of Cancer, 109(5), 786–792. https://doi.org/https://doi.org/10.1002/ijc.20041

Siegel, R., Naishadham, D., & Jemal, A. (2013). Cancer statistics, 2013. CA: A Cancer Journal for Clinicians, 63(1), 11–30. https://doi.org/10.3322/caac.211

Sinha, V. B., Pandey, N., & Taneja, P. (2016). Biomarker genes for gynaecological cancers. *Research Journal of Pharmacy and Technology*, *9*(10), 1641-1646.

doi : 10.5958/0974-360X.2016.00329.2

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., Walsh, L. A., Postow, M. A., Wong, P., Ho, T. S., Hollmann, T. J., Bruggeman, C., Kannan, K., Li, Y., Elipenahli, C., Liu, C., Harbison, C. T., Wang, L., Ribas, A., … Chan, T. A. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. The New England Journal of Medicine, 371(23), 2189–2199. https://doi.org/10.1056/NEJMoa1406498

Snyder, A., Nathanson, T., Funt, S. A., Ahuja, A., Buros Novik, J., Hellmann, M. D., Chang, E., Aksoy, B. A., Al-Ahmadie, H., Yusko, E., Vignali, M., Benzeno, S., Boyd, M., Moran, M., Iyer, G., Robins, H. S., Mardis, E. R., Merghoub, T., Hammerbacher, J., … Bajorin, D. F. (2017). Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer:

An exploratory multi-omic analysis. PLoS Medicine, 14(5), e1002309. https://doi.org/10.1371/journal.pmed.1002309

Sobin, L. H., Gospodarowicz, M. K., & Wittekind, C. (Eds.). (2011). *TNM classification of malignant tumours*. John Wiley & Sons.

Solomon, D., Davey, D., Kurman, R., Moriarty, A., O'Connor, D., Prey, M., Raab, S., Sherman, M., Wilbur, D., Wright Thomas, J., Young, N., & Workshop, for the F. G. M. and the B. 2001. (2002). The 2001 Bethesda SystemTerminology for Reporting Results of Cervical Cytology. JAMA, 287(16), 2114–2119. https://doi.org/10.1001/jama.287.16.2114

Soto, D., Song, C., & McLaughlin-Drubin, M. E. (2017). Epigenetic Alterations in Human Papillomavirus-Associated Cancers. Viruses, 9(9). https://doi.org/10.3390/v9090248

Stanley, M. (2007). Prophylactic HPV vaccines. Journal of Clinical Pathology, 60(9), 961 LP – 965. https://doi.org/10.1136/jcp.2006.040568

Steenbergen, R. D. M., Kramer, D., Braakhuis, B. J. M., Stern, P. L., Verheijen, R. H. M., Meijer, C. J. L. M., & Snijders, P. J. F. (2004). TSLC1 Gene Silencing in Cervical Cancer Cell Lines and Cervical Neoplasia. JNCI: Journal of the National Cancer Institute, 96(4), 294–305. https://doi.org/10.1093/jnci/djh031

Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., ... & Stratton, M. R. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, *486*(7403), 400-404.doi:10.1038/NATURE11017.

Stolnicu, S., & Goldfrank, D. (2021). Anatomy, Histology, Cytology, and Colposcopy of the Cervix. *Atlas of Diagnostic Pathology of the Cervix: A Case-Based Approach*, 1-23.

Suarez, S. S., & Pacey, A. A. (2006). Sperm transport in the female reproductive tract. *Human reproduction update*, *12*(1), 23-37. https://doi.org/10.1093/humupd/dmi047

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/https://doi.org/10.3322/caac.21660

Swain, M. (2023). Update in Pathological Classification of Cervical Intraepithelial Neoplasia and Cervical Cancer. *Journal of Colposcopy and Lower Genital Tract Pathology*, *1*(2), 56-63.

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., ... & von Mering, C. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*, *49*(D1), D605-D612. doi:10.1093/NAR/GKAA1074.

Tampa, M., Mitran, C. I., Mitran, M. I., Nicolae, I., Dumitru, A., Matei, C., ... & Georgescu, S. R. (2020). The role of beta HPV types and HPV-associated inflammatory processes in cutaneous squamous cell carcinoma. *Journal of Immunology Research*, *2020*.https://doi.org/10.1155/2020/570163.

Tang, Q., Liu, L., Zhang, H., Xiao, J., & Hann, S. S. (2020). Regulations of miR-183-5p and snail-mediated shikonin-reduced epithelial-mesenchymal transition in cervical cancer cells. *Drug design, development and therapy*, 577-589. DOI: 10.2147/DDDT.S236216

Tewari, K. S., & Monk, B. J. (2014). New strategies in advanced cervical cancer: from angiogenesis blockade to immunotherapy. Clinical Cancer Research, 20(21), 5349–5358. https://doi.org/10.1158/1078-0432.CCR-14-1099

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., Powderly, J. D., Carvajal, R. D., Sosman, J. A., Atkins, M. B., Leming, P. D., Spigel, D. R., Antonia, S. J., Horn, L., Drake, C. G., Pardoll, D. M., Chen, L., Sharfman, W. H., Anders, R. A., … Sznol, M. (2012). Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. New England Journal of Medicine, 366(26), 2443–2454. https://doi.org/10.1056/NEJMoa1200690

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC Medical Informatics and Decision Making, 19(1), 281. https://doi.org/10.1186/s12911-019-1004-8

Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M. H. G., Goldinger, S. M., Utikal, J., Hassel, J. C., Weide, B., Kaehler, K. C., Loquai, C., Mohr, P., Gutzmer, R., Dummer, R., Gabriel, S., … Garraway, L. A. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science (New York, N.Y.), 350(6257), 207–211. https://doi.org/10.1126/science.aad0095

Van Doorslaer, K., Tan, Q., Xirasagar, S., Bandaru, S., Gopalan, V., Mohamoud, Y., ... & McBride, A. A. (2012). The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic acids research*, *41*(D1), D571-D578. https://doi.org/10.1093/nar/gks984

Viana, M. R. P., Melo, I. M. A., Pupin, B., Raniero, L. J., & de Azevedo Canevari, R. (2020). Molecular detection of HPV and FT-IR spectroscopy analysis in women with normal cervical cytology. *Photodiagnosis and Photodynamic Therapy*, *29*, 101592. https://doi.org/10.1016/j.pdpdt.2019.101592

Villa, L. L., Costa, R. L. R., Petta, C. A., Andrade, R. P., Ault, K. A., Giuliano, A. R., Wheeler, C. M., Koutsky, L. A., Malm, C., Lehtinen, M., Skjeldestad, F. E., Olsson, S.-E., Steinwall, M., Brown, D. R., Kurman, R. J., Ronnett, B. M., Stoler, M. H.,

Topalian, S. L., Hodi, F. S., Brahmer, J. R., Gettinger, S. N., Smith, D. C., McDermott, D. F., Powderly, J. D., Carvajal, R. D., Sosman, J. A., Atkins, M. B., Leming, P. D., Spigel, D. R., Antonia, S. J., Horn, L., Drake, C. G., Pardoll, D. M., Chen, L., Sharfman, W. H., Anders, R. A., … Sznol, M. (2012). Safety, Activity, and Immune Correlates of Anti–PD-1 Antibody in Cancer. New England Journal of Medicine, 366(26), 2443–2454. https://doi.org/10.1056/NEJMoa1200690

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC Medical Informatics and Decision Making, 19(1), 281. https://doi.org/10.1186/s12911-019-1004-8

Van Allen, E. M., Miao, D., Schilling, B., Shukla, S. A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M. H. G., Goldinger, S. M., Utikal, J., Hassel, J. C., Weide, B., Kaehler, K. C., Loquai, C., Mohr, P., Gutzmer, R., Dummer, R., Gabriel, S., … Garraway, L. A. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science (New York, N.Y.), 350(6257), 207–211. https://doi.org/10.1126/science.aad0095

Van Doorslaer, K., Tan, Q., Xirasagar, S., Bandaru, S., Gopalan, V., Mohamoud, Y., ... & McBride, A. A. (2012). The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. *Nucleic acids research*, *41*(D1), D571-D578. https://doi.org/10.1093/nar/gks984

Viana, M. R. P., Melo, I. M. A., Pupin, B., Raniero, L. J., & de Azevedo Canevari, R. (2020). Molecular detection of HPV and FT-IR spectroscopy analysis in women with normal cervical cytology. *Photodiagnosis and Photodynamic Therapy*, *29*, 101592. https://doi.org/10.1016/j.pdpdt.2019.101592

Villa, L. L., Costa, R. L. R., Petta, C. A., Andrade, R. P., Ault, K. A., Giuliano, A. R., Wheeler, C. M., Koutsky, L. A., Malm, C., Lehtinen, M., Skjeldestad, F. E., Olsson, S.-E., Steinwall, M., Brown, D. R., Kurman, R. J., Ronnett, B. M., Stoler, M. H.,

Ferenczy, A., Harper, D. M., … Barr, E. (2005). Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. The Lancet Oncology, 6(5), 271–278. https://doi.org/10.1016/S1470-2045(05)70101-7

Villalobos-Oyarce, R., & Navarro-Venegas, C. (2023). How to differentially detect potentially risky human papilloma virus strains in the human population?. *World Journal of Advanced Research and Reviews*, *18*(2), 244-247.

Virmani, A. K., Muller, C., Rathi, A., Zoechbauer-Mueller, S., Mathis, M., & Gazdar, A. F. (2001). Aberrant methylation during cervical carcinogenesis. Clinical Cancer Research : An Official Journal of the American Association for Cancer Research, 7(3), 584–589.

von Knebel Doeberitz, M., Reuschenbach, M., Schmidt, D., & Bergeron, C. (2012). Biomarkers for cervical cancer screening: the role of p16INK4a to highlight transforming HPV infections. *Expert review of proteomics*, *9*(2), 149-163. https://doi.org/10.1586/epr.12.13

Wang, Dan, et al. "Target Identification-Based Analysis of Mechanism of Betulinic Acid-Induced Cells Apoptosis of Cervical Cancer SiHa." Natural Product Communications 17.7 (2022): 1934578X221115528.https://doi.org/10.1177/1934578X221115528

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, *10*(1), 57-63..doi:10.1038/NRG2484.

Wei, Z., Liu, X., Cheng, C., Yu, W., & Yi, P. (2021). Metabolism of amino acids in cancer. *Frontiers in cell and developmental biology*, *8*, 603837.https://doi.org/10.3389/fcell.2020.603837

Widschwendter, A., Ivarsson, L., Blassnig, A., Müller, H. M., Fiegl, H., Wiedemair, A., Müller-Holzner, E., Goebel, G., Marth, C., & Widschwendter, M. (2004). CDH1 and CDH13 methylation in serum is an independent prognostic marker in cervical cancer patients. International Journal of Cancer, 109(2), 163–166. https://doi.org/https://doi.org/10.1002/ijc.11706

Widschwendter, A., Müller, H. M., Fiegl, H., Ivarsson, L., Wiedemair, A., Müller-Holzner, E., Goebel, G., Marth, C., & Widschwendter, M. (2004). DNA methylation in serum and tumors of cervical cancer patients. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 10(2), 565–571. https://doi.org/10.1158/1078-0432.ccr-0825-03

Wieners, G., Mohnike, K., Peters, N., Bischoff, J., Kleine-Tebbe, A., Seidensticker, R., ... & Ricke, J. (2011). Treatment of hepatic metastases of breast cancer with CT-guided interstitial brachytherapy–a phase II-study. *Radiotherapy and Oncology*, *100*(2), 314-319.

Woodman, C. B. J., Collins, S. I., & Young, L. S. (2007). The natural history of cervical HPV infection: unresolved issues. Nature Reviews Cancer, 7(1), 11–22. https://doi.org/10.1038/nrc2050

Wu, Y. A., Wang, X., Wu, F., Huang, R., Xue, F., Liang, G., ... & Huang, Y. (2012). Transcriptome profiling of the cancer, adjacent non-tumor and distant normal tissues from a colorectal cancer patient by deep sequencing.. https://doi.org/10.1371/journal.pone.0041001

Xia, S.-J., Gao, B.-Z., Wang, S.-H., Guttery, D. S., Li, C.-D., & Zhang, Y.-D. (2021). Modeling of diagnosis for metabolic syndrome by integrating symptoms into physiochemical indexes. Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie, 137, 111367. https://doi.org/10.1016/j.biopha.2021.111367

Xu, W., Xu, M., Wang, L., Zhou, W., Xiang, R., Shi, Y., Zhang, Y., & Piao, Y. (2019). Integrative analysis of DNA methylation and gene expression identified cervical cancer-specific diagnostic biomarkers. Signal Transduction and Targeted Therapy, 4(1), 55. https://doi.org/10.1038/s41392-019-0081-6

Xu, Z., Huang, L., Zhang, T., Liu, Y., Fang, F., Wu, X., ... & Hu, P. (2022). Shikonin inhibits the proliferation of cervical cancer cells via FAK/AKT/GSK3β signalling. *Oncology Letters*, *24*(3), 1-10.. https://doi.org/10.3892/ol.2022.13424

Yang, H. J., Liu, V. W. S., Wang, Y., Chan, K. Y. K., Tsang, P. C. K., Khoo, U. S., Cheung, A. N. Y., & Ngan, H. Y. S. (2004). Detection of hypermethylated genes in tumor and plasma of cervical cancer patients. Gynecologic Oncology, 93(2), 435–440. https://doi.org/10.1016/j.ygyno.2004.01.039

Yang, Y., Xu, L., Sun, L., Zhang, P., & Farid, S. S. (2022). Machine learning application in personalised lung cancer recurrence and survivability prediction. Computational and Structural Biotechnology Journal, 20, 1811–1820. https://doi.org/10.1016/j.csbj.2022.03.035

Yang, Z., Jia, Y., Wang, S., Zhang, Y., Fan, W., Wang, X., ... & Yang, H. (2023). Retinoblastoma-binding protein 5 regulates H3K4 methylation modification to inhibit the proliferation of melanoma cells by inactivating the Wnt/β-catenin and epithelial-mesenchymal transition pathways. *Journal of Oncology*, *2023*.https://doi.org/10.1155/2023/5093941

Yao, S., Xu, J., Zhao, K., Song, P., Yan, Q., Fan, W., ... & Lu, C. (2018). Down-regulation of HPGD by miR-146b-3p promotes cervical cancer cell proliferation, migration and anchorage-independent growth through activation of STAT3 and AKT pathways. *Cell death & disease*, *9*(11), 1055.https://doi.org/10.1038/s41419-018-1059-y.

Yeung, C. L. A., Tsang, T. Y., Yau, P. L., & Kwok, T. T. (2017). Human papillomavirus type 16 E6 suppresses microRNA-23b expression in human cervical cancer cells through DNA methylation of the host gene C9orf3. *Oncotarget*, *8*(7), 12158.https://doi.org/10.18632/oncotarget.14555.

Yu, M. Y., Tong, J. H., Chan, P. K., Lee, T. L., Chan, M. W., Chan, A. W., ... & To, K. F. (2003). Hypermethylation of the tumor suppressor gene RASSFIA and frequent concomitant loss of heterozygosity at 3p21 in cervical cancers. *International journal of cancer*, *105*(2), 204-209.https://doi.org/10.1002/ijc.11051

Yu, M., Ting, D. T., Stott, S. L., Wittner, B. S., Ozsolak, F., Paul, S., ... & Haber, D. A. (2012). RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. *Nature*, *487*(7408), 510-513.https://doi.org/10.1038/nature11217

Zambrano, P., Segura-Pacheco, B., Perez-Cardenas, E., Cetina, L., Revilla-Vazquez, A., Taja-Chayeb, L., Chavez-Blanco, A., Angeles, E., Cabrera, G., Sandoval, K., Trejo-Becerril, C., Chanona-Vilchis, J., & Duenas-González, A. (2005). A phase I study of hydralazine to demethylate and reactivate the expression of tumor suppressor genes. BMC Cancer, 5, 44. https://doi.org/10.1186/1471-2407-5-44

Zammataro, L., Lopez, S., Bellone, S., Pettinella, F., Bonazzoli, E., Perrone, E., Zhao, S., Menderes, G., Altwerger, G., Han, C., Zeybek, B., Bianchi, A., Manzano, A., Manara, P., Cocco, E., Buza, N., Hui, P., Wong, S., Ravaggi, A., … Santin, A. D. (2019). Whole-exome sequencing of cervical carcinomas identifies activating ERBB2 and PIK3CA mutations as targets for combination therapy. Proceedings of the National Academy of Sciences, 116(45), 22730–22736. https://doi.org/10.1073/pnas.1911385116

Zhang, J., Chiodini, R., Badr, A., & Zhang, G. (2011). The impact of next-generation sequencing on genomics. *Journal of genetics and genomics*, *38*(3), 95-109.https://doi.org/10.1016/j.jgg.2011.02.003

Zhang, L., Jiang, Y., Lu, X., Zhao, H., Chen, C., Wang, Y., ... & Yan, F. (2019). Genomic characterization of cervical cancer based on human papillomavirus status. *Gynecologic oncology*, *152*(3), 629-637.https://doi.org/10.1016/j.ygyno.2018.12.017

Zhang, S., Ma, L., Meng, Q. W., Zhou, D., & Moyiding, T. (2017). Comparison of laparoscopic-assisted radical vaginal hysterectomy and abdominal radical hysterectomy in patients with early stage cervical cancer: a retrospective study. *Medicine*, *96*(36), e8005. *DOI:* 10.1097/MD.0000000000008005

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, *9*, 1-8.. https://doi.org/10.1186/1471-2105-9-40

Zhao, J., Cao, H., Zhang, W., Fan, Y., Shi, S., & Wang, R. (2021). SOX14 hypermethylation as a tumour biomarker in cervical cancer. *Bmc Cancer*, *21*(1), 675.https://doi.org/10.1186/s12885-021-08406-2

Zhao, Y., Chen, X., & Yin, J. (2019). Adaptive boosting-based computational model for predicting potential miRNA-disease associations. Bioinformatics (Oxford, England), 35(22), 4730–4738. https://doi.org/10.1093/bioinformatics/btz297