Appendices

APPENDICES

Appendix I
QIAamp DNA Mini Kit

Principle

The QIAamp DNA Mini Kit is designed for the purification of high-quality genomic DNA from various sample types, including blood, tissue, and cells. The process utilizes a silica membrane technology, which allows DNA to bind under specific conditions while contaminants are washed away. The purified DNA is then eluted in a suitable buffer or water for downstream applications such as PCR, sequencing, or other molecular analyses.

Materials

- Buffer AVL: Lysis buffer containing proteinase K (if necessary).
- Buffer AW1 and AW2: Wash buffers used to remove contaminants.
- Buffer AE: Elution buffer for purified DNA.
- QIAamp Mini spin column with a silica membrane for DNA binding.
- Ethanol (absolute)
- Centrifuge for spinning the columns
- Pipettes and tips
- Sample collection tubes

Methodology

To begin the DNA purification process using the QIAamp DNA Mini Kit, the sample was first collected and prepared to ensure it was suitable for extraction. Next, Buffer AVL was added to the sample, along with Proteinase K if required, to facilitate cell lysis and protein digestion. The mixture was incubated at 56°C for 10 to 30 minutes to ensure thorough lysis. Following this incubation, ethanol was added to the lysate to promote DNA binding to the silica membrane. The resulting lysate was then loaded onto the QIAamp Mini spin column and centrifuged to allow the DNA to bind to the membrane. Subsequently, the column was washed with Buffer AW1 and AW2 to remove impurities. Finally, the purified DNA was eluted by adding Buffer AE to the column, allowing it to sit at room temperature briefly before centrifuging to collect the high-quality genomic DNA, ready for further molecular analysis.

Appendix II

NanoDropTM 1000 Spectrophotometer (Thermo Fisher Scientific)

Principle:

The NanoDrop™ 1000 Spectrophotometer operates on the principle of measuring absorbance of light by a sample at specific wavelengths. It employs a small volume of liquid (typically 1-2 µL) placed between two optical fibers, which minimizes sample consumption and allows for high-precision measurements. The device uses a photodiode array to capture the spectrum of light absorbed by the sample, providing a rapid analysis of nucleic acids, proteins, and other biomolecules.

Materials:

- NanoDrop™ 1000 Spectrophotometer
- Sample cuvettes or pipettes for loading samples
- Deionized water or buffer solutions for blanking and cleaning
- Nucleic acid or protein samples for analysis

Methodology:

To perform a measurement using the NanoDrop™ 1000 Spectrophotometer, the instrument was first calibrated and set up properly. A small drop of deionized water or the appropriate buffer solution was placed on the measurement pedestal to create a baseline for absorbance. The arm of the spectrophotometer was closed to initiate the blanking process, establishing a zero point for subsequent measurements. After blanking, the water or buffer was removed, and a small volume (1-2 µL) of the sample solution was applied directly onto the pedestal. The arm was closed again, and the appropriate measurement program was selected on the instrument's software, depending on whether nucleic acids, proteins, or other samples were being measured. The NanoDrop™ 1000 then automatically measured the absorbance across a range of wavelengths and displayed the results, including concentration and purity ratios (A260/A280, A260/A230) on its screen. After measurements, the optical surfaces of the spectrophotometer were thoroughly cleaned with a lint-free cloth and deionized water to ensure that no residues affected future analyses.

Appendix III

Agarose gel electrophoresis

Agarose Gel Electrophoresis is a fundamental technique in molecular biology for separating and analysing nucleic acids and proteins.

Principle:

Agarose gel electrophoresis separates DNA and RNA by size and charge. When an electric field is applied, the negatively charged molecules move toward the positive side. Smaller molecules move faster, so they separate. The DNA is stained with dyes like ethidium bromide or SYBR Green and viewed under UV light.

Materials:

- Agarose powder
- Buffer solution and Running buffer (e.g., TAE or TBE buffer)
- Gel casting tray & Gel comb
- Electrophoresis apparatus
- DNA ladder
- Staining dye (e.g., ethidium bromide, SYBR Green)
- UV transilluminator (for visualization)

Methodology:

To perform agarose gel electrophoresis, a 0.8-2% agarose gel was first prepared by dissolving the appropriate amount of agarose powder in a buffer solution, heating it until fully melted, and then allowing it to cool slightly before pouring it into a gel casting tray with a comb inserted to form wells. Once the gel solidified, the comb was carefully removed, and the gel was placed into the electrophoresis chamber, filling it with running buffer to cover the gel. DNA samples were then prepared by mixing them with a loading dye, which helped visualize the sample during loading and electrophoresis. The samples were loaded into the wells of the gel, alongside a DNA ladder for size reference. The electrophoresis apparatus was connected to a power supply, the voltage was set

according to the gel size and buffer system, and the gel was run for 30 minutes to several hours, depending on the desired separation. After electrophoresis was complete, the gel was carefully removed and stained with a DNA-binding dye for visualization. Using a UV transilluminator, the gel was observed to analyze the separated DNA fragments, comparing them to the DNA ladder to determine their sizes. Finally, the results were documented.

Appendix IV

Twist Library Preparation EF Kit with Amp Mix

Principle:

The Twist Library Preparation EF Kit with Amp Mix is designed for efficient preparation of DNA libraries for high-throughput sequencing. This kit utilizes enzymatic reactions for fragmenting and ligating DNA, allowing for the creation of libraries that can be effectively amplified and sequenced. The process includes the end repair of DNA fragments, A-tailing, adapter ligation, and PCR amplification, enabling the generation of sequence-ready libraries with high yield and quality.

Materials:

Twist Library Preparation EF Kit components: (End repair enzyme mix, A-tailing enzyme mix, Adapter ligation enzyme mix, PCR amplification mix (Amp Mix), Wash buffer, Elution buffer), Input DNA sample (genomic, amplicon, or other sources), Thermal cycler, Magnetic bead purification system (optional, for library cleanup), and Bioanalyzer or qPCR system (for library quality control)

Methodology:

To prepare a DNA library using the Twist Library Preparation EF Kit with Amp Mix, the input DNA was first quantified and diluted to the recommended concentration. End repair was performed by mixing the input DNA with the end repair enzyme mix in a microcentrifuge tube, followed by incubation at the specified temperature for the recommended time to ensure complete repair of DNA ends. The A-tailing enzyme mix was then added to the reaction to attach an "A" base to the 3' ends of the DNA fragments. After a brief incubation, adapter ligation was carried out by adding the adapter ligation enzyme mix and appropriate adapters to the mixture, followed by incubation to facilitate the ligation of adapters to the A-tailed fragments. After ligation, the ligated library was purified using a magnetic bead purification system to remove any unligated adapters and enzyme components. The library was then amplified using the provided PCR amplification mix (Amp Mix) in a thermal cycler, with the number of cycles adjusted based on the desired yield. Once PCR amplification was complete, a final purification step

was performed to clean the amplified library and remove excess primers and dNTPs. The quality and quantity of the prepared library were then assessed using a Bioanalyzer or qPCR system to confirm it met the criteria for high-throughput sequencing. The resulting library was ready for use in subsequent sequencing reactions to generate high-quality data for genomic analysis.

Appendix V

Illumina NovaSeq 6000

Principle:

The Illumina NovaSeq 6000 is a high-throughput sequencing platform that uses sequencing by synthesis (SBS) technology. Fluorescently labeled nucleotides are added to a growing DNA strand during sequencing. The system uses a flow cell with multiple lanes to sequence multiple samples at once, making it fast and cost-effective. It can generate millions to billions of reads in a single run and is used for applications like whole-genome sequencing, targeted sequencing, and RNA sequencing.

Materials:

- Illumina NovaSeq 6000 sequencer
- NovaSeq 6000 Reagent Kits, (Sequencing reagents (for SBS), Flow cells (SP, S1, S2, or S4))
- Library preparation kits (specific to the application)
- Analysis software (such as BaseSpace Sequence Hub or local analysis pipelines)
- Wash solutions and buffers

Methodology:

To perform sequencing on the Illumina NovaSeq 6000, the DNA or RNA library was first prepared according to the guidelines of the chosen library preparation kit, ensuring it was of high quality and concentration. Once the library was ready, its quality and quantity were assessed using a Bioanalyzer or qPCR to ensure it met the input requirements for sequencing. The prepared library was then loaded onto the designated flow cell, with the appropriate flow cell type selected based on the desired output and application. The flow cell was inserted into the NovaSeq 6000 sequencer, and the sequencing run was initiated by selecting the desired sequencing mode and read length in the control software.

The sequencing process began with the bridge amplification of the library fragments on the flow cell surface, creating clusters of identical DNA molecules. Following amplification, sequencing by synthesis was performed, where fluorescently labeled reversible terminator nucleotides were sequentially added, and the incorporated bases were detected using a high-resolution imaging system. This process was repeated for a set number of cycles, generating millions of short reads of DNA sequences.

After the sequencing run was complete, raw data in the form of image files was generated, which were then processed and converted into sequence data (FASTQ format) using the onboard analysis software. Finally, the sequencing results were analyzed for various applications, such as variant calling, gene expression analysis, or metagenomics, using appropriate bioinformatics tools. The NovaSeq 6000's ability to handle large-scale sequencing projects efficiently made it a powerful tool for researchers in genomics and personalized medicine.

Appendix VI

Burrows-Wheeler Transform (BWT) algorithm

Principle:

BWA is a software tool for mapping high-throughput sequencing data against a reference genome. It uses the Burrows-Wheeler Transform (BWT) algorithm to efficiently align sequences.

Materials:

- Reference genome (FASTA format)
- Sequencing data (FASTQ format)
- Computing resources (e.g., Linux, RAM)

Methodology

The Burrows-Wheeler Transform (BWT) methodology began with taking an input string, typically appended with a unique end-of-string marker like "$" to ensure all rotations were distinct. The core process involved generating all possible cyclic rotations of the input string, which were then stored in a matrix. This matrix was subsequently sorted lexicographically, meaning it was ordered based on dictionary order of the rows.

The BWT output was constructed by taking the last column of the sorted matrix, which contained characters that represented the transform of the original string. The beauty of BWT lay in its ability to group similar characters together, making it highly efficient for compression when followed by techniques such as Run-Length Encoding (RLE) or Move-to-Front (MTF) encoding. Importantly, BWT was fully reversible: the inverse BWT could reconstruct the original string using the last column and a lexicographically sorted version of that column, allowing full retrieval of the original data.

This methodology was particularly powerful in bioinformatics, where BWT was used to store and search large genomic datasets efficiently, as it provided a highly compressed form of the data while allowing fast pattern searches.

Appendix VII

Sequence Alignment/Map

Principle:

SAM is a text-based file format for storing sequence alignment data, allowing efficient storage and retrieval of alignments.

Materials:

Sequencing data (FASTQ), Reference genome (FASTA), Alignment software (e.g., BWA, Bowtie, STAR), SAMtools software, and Computing resources (e.g., Linux, RAM)

Methodology:

The Sequence Alignment/Map (SAM) format was a widely used method for storing and representing aligned sequence data, particularly in next-generation sequencing (NGS) workflows. The SAM format was text-based and facilitated the storage of large volumes of sequence alignment data by efficiently representing sequences that had been aligned to a reference genome. The methodology began with the alignment of short reads, generated from sequencing platforms, to a reference genome using algorithms such as Burrows-Wheeler Aligner (BWA) or Bowtie. These alignment tools generated a map of where each read aligned within the reference genome, and this information was captured in SAM format.

The SAM file consisted of a header section and an alignment section. The header contained metadata, such as the reference genome, sequence identifiers, and details about the alignment process. The alignment section recorded the essential information for each aligned read, including the read's name, the reference sequence it aligned to, its position, mapping quality, and a CIGAR string that described how the read aligned (e.g., matches, insertions, or deletions). Each line in the alignment section corresponded to one read and included details like the read sequence and quality scores.

SAM could also be converted into Binary Alignment/Map (BAM) format, which was a compressed binary version of SAM. This transformation allowed for efficient storage and faster retrieval of alignment data. The SAM/BAM format was critical in bioinformatics for downstream analysis, such as variant calling, where differences between the aligned reads and the reference genome were identified. Through its structured and efficient representation of aligned sequence data, SAM enabled large-scale genomic analyses with precision and speed.

Appendix VIII

Sorting Intolerant From Tolerant

Principle:

SIFT predicts the effects of amino acid substitutions on protein function using sequence homology and machine learning.

Materials:

- Protein sequence (FASTA format)
- SIFT software (available online or as a standalone tool)

Methodology:

The SIFT (Sorting Intolerant From Tolerant) algorithm was a computational method used to predict the potential impact of amino acid substitutions on protein function. It was primarily employed to assess whether a particular missense mutation (an amino acid change) was likely to be damaging or tolerated within the context of the protein's structure and function. The methodology behind SIFT was based on the evolutionary conservation of amino acids in protein sequences. It operated under the principle that functionally important amino acids tended to be conserved across different species, while less critical positions could tolerate more variability.

The SIFT algorithm worked by aligning the query protein sequence with homologous sequences from related species and generating a multiple sequence alignment. This alignment was used to determine how conserved each amino acid position was across evolution. SIFT assigned a score to each possible amino acid substitution based on how often that substitution occurred in the alignment. Scores ranged from 0 to 1, where substitutions with scores below a threshold (typically 0.05) were predicted to be deleterious, meaning they were likely to affect protein function, while scores above this threshold were considered tolerated.

By leveraging evolutionary information, SIFT provided insights into the functional impact of mutations identified in genetic studies, such as those related to disease pathogenesis or personalized medicine. The algorithm was commonly used in genomic research to prioritize variants for further study and to help guide decisions about which mutations may warrant experimental validation.

Appendix IX

PolyPhen-2 HDIVAR

Principle:

PolyPhen-2 HDIVAR predicts the functional impact of human genetic variants, particularly missense mutations, using structural and comparative evolutionary considerations.

Materials:

- Protein sequence (FASTA format)
- 3D protein structure (PDB format)
- Multiple sequence alignment (MSA) tools (e.g., PSI-BLAST)
- PolyPhen-2 HDIVAR software (available online or as a standalone tool)

Methodology:

First, the algorithm compared the amino acid sequence of the mutated protein with homologous sequences from different species to evaluate evolutionary conservation. Highly conserved regions were considered functionally important, and any substitution in these regions was more likely to disrupt protein function. Next, PolyPhen-2 HDIVAR assessed the structural context of the mutation by examining the three-dimensional structure of the protein (if available) or by using structural models. It looked at how the mutation affected important features like protein folding, stability, or interaction with other molecules.

PolyPhen-2 then combined this information with data on known pathogenic variants to generate a score ranging from 0 to 1. Scores closer to 1 indicated that the mutation was likely to be damaging, while scores closer to 0 suggested it was benign. The HDIVAR model specifically focused on high-diversity human variation and was trained on known disease-causing mutations from the Human Gene Mutation Database (HGMD), making it suitable for identifying clinically relevant variants.

This methodology enabled researchers and clinicians to predict the functional impact of genetic variants and prioritize them for further experimental validation or clinical interpretation, particularly in the context of hereditary diseases.

Appendix X

Likelihood Ratio Test

Principle:

The LRT (Likelihood Ratio Test) is a statistical method used to compare the fit of two competing models: one that includes additional parameters and another that is simpler. The test evaluates whether the more complex model significantly improves the fit of the data compared to the simpler model. The null hypothesis in the LRT assumes that the simpler model is sufficient, while the alternative hypothesis suggests that the additional parameters improve the model's explanatory power. The test statistic follows a chi-square distribution, and a significant p-value indicates that the more complex model is a better fit for the data.

Materials:

- dbSNP database for variant cross-referencing
- Statistical software for LRT calculations

Methodology:

The LRT (Likelihood Ratio Test Prediction) method evaluated the potential functional impact of genetic mutations by comparing two evolutionary models: the neutral evolution hypothesis (H0), which assumed no selection pressure on the codon, and the negative selection hypothesis (H1), which suggested the codon evolved under selective pressure, indicating the mutation was deleterious. The process began by aligning homologous sequences and using a statistical framework to calculate the likelihood of the observed mutation data under both models. A Likelihood Ratio (LRT) score was computed by taking the difference between the likelihoods of H0 and H1, which was then multiplied by 2. The resulting score helped determine whether the mutation was neutral (high score), deleterious (low score), or unknown (inconclusive score). Mutations with lower LRT scores were more likely to be deleterious, implying they could contribute to diseases. The tool classified mutations into three categories: Deleterious (D), Neutral (N), or Unknown (U), providing insights into the mutation's potential functional consequences. This method was particularly useful in identifying harmful mutations that could affect protein function and contribute to disease progression, especially in cancer research.

Appendix XI

Mutation Taster

Principle:

MutationTaster predicts the functional impact of genetic variants, including point mutations, insertions, and deletions, on protein function.

Materials:

- Genetic variant information (e.g., chromosome, position, reference allele, alternate allele)
- Genome assembly (e.g., hg19, hg38)
- MutationTaster software (available online or as a standalone tool)

Methodology:

First, it examined evolutionary conservation, comparing the mutated sequence with orthologous sequences from different species to determine if the affected region was functionally important. Highly conserved sequences were more likely to be crucial for normal biological function, and mutations in these regions were more likely to have deleterious effects.

Next, MutationTaster evaluated the potential effect of the mutation on protein-coding regions. It assessed whether the variant led to nonsense mutations (which could introduce premature stop codons), missense mutations (which resulted in amino acid changes), or frameshift mutations (which could disrupt the entire downstream protein sequence). It also considered how the mutation might affect splice sites, regulatory regions, or other non-coding elements that were vital for gene expression and function.

To enhance accuracy, MutationTaster incorporated known pathogenic variants from databases such as ClinVar and the Human Gene Mutation Database (HGMD). It used this information, along with in silico models and data on protein structure, to provide a score that categorized the variant as either disease-causing ("pathogenic") or benign. Additionally, it provided detailed reports explaining the

reasoning behind its predictions, making the tool valuable for clinical diagnostics and genetic research.

This comprehensive approach allowed MutationTaster to offer a robust assessment of variant pathogenicity, aiding in the identification of mutations that were likely to contribute to genetic diseases.

Appendix XII

(Functional Analysis through Hidden Markov Model)

Principle

FATHMM (Functional Analysis through Hidden Markov Models) predicts the functional impact of genetic variants, particularly non-synonymous single-nucleotide polymorphisms (nsSNPs), by leveraging evolutionary conservation and functional domain analysis.

Materials

- Protein sequences: In FASTA format for alignment and analysis.
- FATHMM software or web server: Accessible online or as standalone software.

Methodology

FATHMM was a computational tool designed to predict the functional impact of non-synonymous genetic mutations, particularly those in coding regions of the genome. It utilized a hidden Markov model (HMM) to analyze the mutation's potential effect on protein function. The methodology began by aligning the sequence of the mutated gene with multiple homologous sequences to assess evolutionary conservation. The model considered the sequence's evolutionary context, focusing on how mutations affected the protein's functional domains and structure. The FATHMM algorithm then computed a score (FATHMM_score), which ranged from 0 to 1, where lower scores indicated a higher likelihood of the mutation being deleterious (D) and higher scores suggested a tolerated mutation (T). This model was trained on a large dataset of known disease-associated mutations to refine its predictions. By analyzing mutations in the context of protein evolution and structure, FATHMM helped prioritize mutations that were more likely to contribute to disease, such as cancer or genetic disorders, for further experimental validation.

Appendix XIII

Radial Support Vector Machine

Principle:

RadialSVM (Radial Support Vector Machine) Protein Variation Effect Analyzer predicts the functional impact of missense mutations in proteins by combining evolutionary conservation, structural data, and physicochemical properties.

Materials

- Protein sequences: In FASTA format, representing the wild-type and mutant proteins.
- Protein structural data: From databases like Protein Data Bank (PDB)
- RadialSVM software: Available online or as a standalone tool for variant analysis.
- Training data: Datasets of known deleterious and neutral mutations, such as those from ClinVar or HGMD, for model validation.

Methodology

The RadialSVM Protein Variation Effect Analyzer utilized sequence homology and machine learning techniques to predict the functional impact of protein-coding variants. The analysis began with preparing input data, including wild-type and mutant protein sequences in FASTA format. Homologous sequences were collected from publicly available databases and aligned using multiple sequence alignment tools to assess evolutionary conservation. Variants in conserved regions were flagged as potentially impactful due to their significance in maintaining protein functionality.

Key features, such as physicochemical properties, structural characteristics, and conservation scores, were extracted from the aligned sequences. These features were then used as input for a support vector machine (SVM) model with a radial basis function (RBF) kernel. The trained SVM model classified variants as

deleterious (D) or neutral (N) based on a dataset of known mutations. Each variant was assigned a RadialSVM score, with higher scores indicating a greater likelihood of being deleterious. The analysis provided insights into the potential functional impacts of variants, aiding in identifying mutations that could significantly disrupt protein function or stability.

Appendix XIV

Logistic Regression

Principle:

Logistic Regression (LR) is a statistical method used to predict binary or categorical outcomes based on one or more independent variables. It models the relationship between the independent variables and the probability of a particular outcome using a logistic function (sigmoid curve). The algorithm calculates the odds ratio for the presence of the outcome and transforms it into a probability, enabling predictions of classes such as "deleterious" versus "neutral" mutations in biological studies.

Materials:

- Logistic regression library or algorithm (e.g., sklearn in Python)
- Preprocessed and normalized input data
- Cross-validation tools for model evaluation

Methodology:

The logistic regression analysis began with data preparation, ensuring the dataset contained well-labeled features as independent variables and a binary outcome as the dependent variable. The data was split into training and testing subsets for model development and validation. Features were standardized or normalized to ensure equal contributions from all variables to the model.

The logistic regression model was constructed using the training dataset, with coefficients for each feature calculated to maximize the likelihood of correctly classifying the outcomes. The logistic function transformed the linear combination of features into probabilities. A threshold (e.g., 0.5) was applied to these probabilities to classify the outcomes as either "deleterious" or "neutral."

The model was then tested on the validation dataset to evaluate its predictive accuracy, sensitivity, and specificity. Performance metrics, such as the area under the receiver operating characteristic (ROC) curve, were used to assess

the model's discriminative power. Hyperparameters were tuned to optimize the model, and cross-validation was employed to ensure robustness against overfitting. Once validated, the model was applied to new data to predict the functional impact of genetic variants, providing a probabilistic framework for mutation classification.

Appendix XV

In silico PCR

Principle:

In silico PCR, as implemented in the UCSC Genome Browser, is a computational technique used to simulate the polymerase chain reaction (PCR) process. It allows researchers to predict the amplification of DNA sequences based on input primer sequences and a reference genome. The tool identifies regions in the genome where the primers anneal and calculates the expected product size. This approach is particularly valuable for validating primer specificity and designing PCR experiments without requiring laboratory trials.

Materials:

- Primer sequences: Forward and reverse primers in 5' to 3' orientation.
- Reference genome: UCSC Genome Browser reference genome assembly (e.g., hg38 for humans).
- Access to UCSC In Silico PCR tool: Available online through the UCSC Genome Browser platform.
- File formats: Optional input/output files in text format for primer sequences and results.

Methodology:

The process began by accessing the UCSC *In Silico* PCR tool on the UCSC Genome Browser platform. The forward and reverse primer sequences were entered into the designated fields, ensuring their accuracy in the 5' to 3' orientation. Additional parameters, such as the maximum product size and allowed mismatches, were configured to customize the search. Upon initiation, the tool compared the primers to the selected reference genome using sequence alignment algorithms. It identified exact or near-exact matches for the primers and determined whether they were correctly oriented to amplify a specific genomic region.

The tool outputted the genomic coordinates of the amplified region, the size of the PCR product, and the sequence of the amplified fragment. Researchers analyzed the output to verify primer specificity, ensuring the primers amplified only the target region without nonspecific binding elsewhere in the genome. If the predicted amplicon matched the desired region and the product size was within the expected range, the primers were deemed suitable for experimental use. The results could also be visualized in the UCSC Genome Browser to assess the genomic context. This computational approach streamlined primer design and validation, saving time and resources in experimental PCR setup.

Appendix XVI

Touchdown PCR

Principle:

Touchdown PCR is a modified PCR technique that increases specificity by gradually decreasing the annealing temperature during cycling, allowing for more specific binding of primers to the target sequence.

Materials:

- DNA template
- Primers (forward and reverse)
- PCR buffer
- dNTPs
- Mg2+
- Taq DNA polymerase
- Thermal cycler

Methodology:

Touchdown PCR was employed as a variation of the standard polymerase chain reaction (PCR) to enhance specificity and efficiency by gradually lowering the annealing temperature during the initial cycles. The methodology began with a high annealing temperature, typically a few degrees higher than the calculated melting temperature (Tm) of the primers. This elevated temperature facilitated the binding of primers exclusively to the most perfectly matched target sequences, reducing the likelihood of non-specific binding.

As the PCR progressed, the annealing temperature was progressively decreased in small increments, usually by 0.5–1°C per cycle, until it reached a lower temperature closer to the optimal Tm for primer binding. This gradual reduction enabled the selective amplification of specific products as the reaction proceeded, effectively promoting the amplification of the correct target sequence while minimizing non-specific sequences or primer-dimers.

Touchdown PCR proved particularly useful when working with complex or problematic templates, where specificity was critical, or when primers exhibited a tendency to bind to non-target sequences. By initiating the reaction with stringent conditions and relaxing them over time, the method maximized the yield of the desired product while minimizing the formation of unwanted by-products. This made it a powerful technique for applications such as cloning, sequencing, and genotyping.

Appendix XVII

Sanger sequencing

Principle:

Sanger sequencing, also known as dideoxy chain termination, is a DNA sequencing method that uses dideoxynucleotides (ddNTPs) to terminate DNA synthesis at specific points, creating fragments of varying lengths. These fragments are then separated by size using electrophoresis, allowing for the determination of the DNA sequence.

Materials:

- DNA template
- Primers (forward and reverse)
- dNTPs (dATP, dCTP, dGTP, dTTP)
- ddNTPs (ddATP, ddCTP, ddGTP, ddTTP)
- DNA polymerase (e.g., Taq or Sequenase)
- Buffer
- Electrophoresis equipment (e.g., ABI 3730)
- Sequencing software

Methodology:

Sanger sequencing, also referred to as the chain-termination method, was widely used to determine the nucleotide sequence of DNA. The process began with the amplification of the target DNA through standard PCR, followed by a sequencing reaction. In this reaction, the amplified DNA was combined with a primer, DNA polymerase, regular deoxynucleotides (dNTPs), and fluorescently labeled dideoxynucleotides (ddNTPs). The ddNTPs were crucial for chain termination, as they lacked the 3’-hydroxyl group required to form a phosphodiester bond with the subsequent nucleotide.

During DNA synthesis, DNA polymerase incorporated nucleotides into the growing DNA strand. When a ddNTP was integrated, it halted DNA elongation, resulting in fragments of varying lengths. Each fragment ended with a ddNTP

labeled with a fluorescent dye specific to one of the four nucleotides (A, T, C, or G). These fragments were then separated by size using capillary electrophoresis. As the fragments passed through the capillary, a laser excited the fluorescent tags, and a sensor detected the emitted signals, identifying the sequence based on the fluorescent colors.

The detected data were translated into a chromatogram, visually representing the sequence of nucleotides. Sanger sequencing provided highly accurate sequences of up to 900 base pairs, establishing itself as the gold standard for validating genetic mutations, confirming PCR products, and conducting small-scale sequencing projects. Although newer high-throughput sequencing methods have emerged, Sanger sequencing remained a reliable and cost-effective approach for targeted sequencing and the verification of genetic variants.

Supplementary Files

SUPPLEMENTARY TABLES

All the Supplementary Tables are provided as soft copy in the enclosed CD-ROM.

Supplementary File 1: Detailed Mutational Data of Squamous Cell Cervical Cancer Samples (CC1-CC5).

Supplementary File 2: Data for Filtering Process Steps and High-Confidence Mutation Identification.

Supplementary File 3: Novel Variants Identified in Squamous Cell Cervical Cancer (Unreported in dbSNP and ClinVar).

Annexures

ANNEXURES

Annexure I

Cervical Cancer Awareness and Socioeconomic Questionnaire

கர்ப்பப்பை வாய் புற்றுநோய் விழிப்புணர்வு மற்றும் சமூக பொருளாதார பின்னணி

Patient name : _____

நோயாளியின் பெயர் : _____

Section 1: Demographic Information / பிரிவு 1: அடிப்படைதகவல்

1. **Age /** வயது

   ☐ ≤20 years / ≤20 ஆண்டுகள்

   ☐ 21–30 years / 21-30 ஆண்டுகள்

   ☐ 31–40 years / 31-40 ஆண்டுகள்

   ☐ 41–50 years/ 41-50 ஆண்டுகள்

   ☐ ≥51 years / ≥51 ஆண்டுகள்

2. **Place of Residence /** வசிக்கும் இடம்

   ☐ Rural / கிராமப்புறம்

   ☐ Urban / நகர்ப்புறம்

Section 2: Education and Socioeconomic Status / பிரிவு 2: கல்வி மற்றும் சமூகப் பொருளாதார நிலை

3. **What is your highest level of education completed? /** உங்களின் மிக உயர்ந்த கல்வி நிலை என்ன?

   ☐ No formal education / முறையான கல்வி இல்லை

   ☐ Primary education / ஆரம்பக் கல்வி

   ☐ Secondary education / இடைநிலைக் கல்வி

   ☐ College or higher / கல்லூரி அல்லது அதற்கு மேல்

4. **What is your current marital status? /** உங்கள் தற்போதைய திருமண நிலை என்ன?

   ☐ Married / திருமணமானவர்

   ☐ Unmarried / திருமணமாகாதவர்

5. **What is your family's annual income? /** உங்கள் குடும்பத்தின் ஆண்டு வருமானம் என்ன?

   ☐ Below ₹60,000 / ₹60,000க்கு கீழே

   ☐ ₹60,000 to ₹1,20,000 / ₹60,000 முதல் ₹1,20,000 வரை

   ☐ ₹1,20,000 to ₹5,00,000 / ₹1,20,000 முதல் ₹5,00,000 வரை

   ☐ Above ₹5,00,000 / ₹5,00,000க்கு மேல்

6. **What is your current profession? /** உங்கள் தற்போதைய தொழில் என்ன?

   ☐ Student / மாணவர்

   ☐ Housewife / இல்லத்தரசி

   ☐ Employed / வேலை செய்பவர்

Section 3: Awareness of Cervical Cancer and HPV

பிரிவு 3: கர்ப்பப்பை வாய்ப் புற்றுநோய் மற்றும் HPV பற்றிய விழிப்புணர்வு

7. **Have you heard about cervical cancer? /** கர்ப்பப்பை வாய்ப் புற்றுநோய் பற்றி கேள்விப்பட்டிருக்கிறீர்களா?

   ☐ Yes / ஆம்

   ☐ No / இல்லை

8. **Do you know about the Human Papillomavirus (HPV)? /** மனித பாப்பிலோமா வைரஸ் (HPV) பற்றி உங்களுக்குத் தெரியுமா?

   ☐ Yes / ஆம்

   ☐ No / இல்லை

9. **What is your primary source of information about cervical cancer? / கர்ப்பப்பை வாய்ப் புற்றுநோயைப் பற்றிய உங்கள் முதன்மையான தகவல் என்ன?**

   ☐ Television / தொலைக்காட்சி

   ☐ Newspaper/Magazines / செய்தித்தாள்/பத்திரிக்கைகள்

   ☐ Medical practitioner / மருத்துவ பயிற்சியாளர்

   ☐ Friends/Neighbors/Relatives / நண்பர்கள்/உறவினர்கள்

   ☐ Internet / இணையம்

   ☐ Education (School/College/University) / கல்வி

   ☐ I don't know / எனக்குத் தெரியாது

Section 4: Knowledge of Cervical Cancer Screening and Vaccination

பிரிவு 4: கர்ப்பப்பை வாய்ப் புற்றுநோய் பரிசோதனை மற்றும் தடுப்பூசி பற்றிய புரிதல்

10. **பேப் டெஸ்ட் (கர்ப்பப்பை வாய்ப் புற்றுநோய்க்கான ஸ்கிரீனிங் செயல்முறை) பற்றி உங்களுக்குத் தெரியுமா?**

    ☐ Yes / ஆம்

    ☐ No / இல்லை

11. **Do you know about the HPV vaccine? / HPV தடுப்பூசி பற்றி உங்களுக்குத் தெரியுமா?**

    ☐ Yes / ஆம்

    ☐ No / இல்லை

12. **Have you received the HPV vaccine?**

    ☐ Yes / ஆம்

    ☐ No / இல்லை

13. Do you think it is important to undergo a cytological examination (e.g., Pap test) for cervical cancer screening?

☐ Yes / ஆம்

☐ No / இல்லை

Section 5: Knowledge of Cervical Cancer Symptoms

பிரிவு 5: கர்ப்பப்பை வாய்ப் புற்றுநோய் அறிகுறிகள் பற்றிய புரிதல்

14. Which of the following symptoms do you associate with cervical cancer? / பின்வரும் அறிகுறிகளில் எது கர்ப்பப்பை வாய்ப் புற்றுநோயுடன் தொடர்புடையது?

☐ Itching in the genital area / பிறப்புறுப்பு பகுதியில் அரிப்பு

☐ Pain in the pelvis / இடுப்பு பகுதியில் வலி

☐ Irregular and painful menstrual periods / ஒழுங்கற்ற மற்றும் வலிமிகுந்த மாதவிடாய்

☐ Pain or bleeding during or after sexual activity / பாலியல் செயல்பாட்டின் போது அல்லது அதற்குப் பிறகு வலி அல்லது இரத்தப்போக்கு

☐ Discomfort during urination / சிறுநீர் கழிக்கும் போது அசௌகரியம்

☐ Unusual or smelly vaginal discharge / வெள்ளை வெளியேற்றம்

☐ I don't know / எனக்குத் தெரியாது

Section 6: Knowledge of Risk Factors for Cervical Cancer

பிரிவு 6: கர்ப்பப்பை வாய்ப் புற்றுநோய்க்கான ஆபத்து காரணிகள் பற்றிய அறிவு

15. Which of the following do you think are risk factors for cervical cancer?

பின்வருவனவற்றில் எது கர்ப்பப்பை வாய்ப் புற்றுநோய்க்கான ஆபத்து காரணிகள் என்று நீங்கள் நினைக்கிறீர்கள்?

☐ Being overweight / அதிக எடையுடன் இருப்பது

☐ Early pregnancy / மிக இளம் வயதில் கர்ப்பம்

☐ Early start of sexual activity / சிறு வயதிலேயே பாலியல் செயல்பாடு

☐ Genetic factors / மரபணு காரணிகள்

☐ Having a weakened immune system / பலவீனமான நோயெதிர்ப்பு

☐ HPV infection / HPV தொற்று

☐ Long-term use of oral contraceptives / வாய்வழி கருத்தடைகளின் நீண்டகால பயன்பாடு

☐ Marriage at an early age / சிறு வயதிலேயே திருமணம்

☐ Multiple sexual partners / பல பாலியல் நண்பர்கள்

☐ Repeated pregnancy

☐ Smoking

☐ I don't know

Consent Form

We are conducting a study on **Cervical Cancer Awareness and Socioeconomic Factors** to better understand the levels of knowledge, awareness, and socioeconomic influences related to cervical cancer. Your participation will help us identify gaps and areas for improvement in public health interventions. Participation in this study is completely voluntary, and you may choose to withdraw at any time. All the information you provide will remain confidential and will be used solely for research purposes. There are no risks associated with participation, and no identifiable personal information will be shared. By proceeding with the questionnaire, you acknowledge that you have understood the above and agree to participate in this study.

கர்ப்பப்பை வாய்ப் புற்றுநோய் தொடர்பான அறிவு, விழிப்புணர்வு மற்றும் சமூகப் பொருளாதார தாக்கங்களின் அளவுகளை நன்கு புரிந்துகொள்ள, கர்ப்பப்பை வாய்ப் புற்றுநோய் விழிப்புணர்வு மற்றும் சமூகப் பொருளாதாரக் காரணிகள் குறித்த ஆய்வை நடத்தி வருகிறோம். பொது சுகாதாரத் தலையீடுகளில் முன்னேற்றத்திற்கான இடைவெளிகளையும் பகுதிகளையும் கண்டறிய உங்கள் பங்கேற்பு எங்களுக்கு உதவும். இந்த ஆய்வில் பங்கேற்பது முற்றிலும் தன்னார்வமானது, எந்த நேரத்திலும் நீங்கள் திரும்பப் பெறலாம். நீங்கள் வழங்கும் அனைத்து தகவல்களும் ரகசியமாக இருக்கும் மற்றும் ஆராய்ச்சி நோக்கங்களுக்காக மட்டுமே பயன்படுத்தப்படும். பங்கேற்புடன் தொடர்புடைய ஆபத்துகள் எதுவும் இல்லை, மேலும் அடையாளம் காணக்கூடிய தனிப்பட்ட தகவல்கள் எதுவும் பகிரப்படாது. கேள்வித்தாளைத் தொடர்வதன் மூலம், மேற்கூறியவற்றை நீங்கள் புரிந்துகொண்டுள்ளீர்கள் என்பதை ஒப்புக்கொள்கிறீர்கள் மற்றும் இந்த ஆய்வில் பங்கேற்க ஒப்புக்கொள்கிறீர்கள்.

If you consent to participate, please check the box below:

☐ I agree to participate in this study / இந்த ஆய்வில் பங்கேற்க ஒப்புக்கொள்கிறேன்.

ANNEXURE II

INSTITUTIONAL HUMAN ETHICS COMMITTEE

Avinashilingam
Institute for Home Science and Higher Education for Women
(Deemed to be university under Category 'A' by MHRD, Estd. u/s 3 of UGC Act 1956) Re-accredited with 'A^{++}' Grade by NAAC.
Recognised by UGC Under Section 12 B
Coimbatore- 641043, Tamil Nadu, India

05.01.2023

To
Ms. B. Sudha
Department of Biotechnology
Avinashilingam Institute for Home Science and
Higher Education for Women
Coimbatore- 641043

Dear Sudha,

Ref: Your proposal No. IHEC/22-23/BT-01 entitled "Identification and Validation of Biomarkers Associated with Squamous Cervical Carcinoma'' submitted for approval of IHEC on 19.11.2022.

The Institutional Human Ethics Committee of our University hereby grants approval to your research proposal No.IHEC/22-23/BT-01 entitled "Identification and Validation of Biomarkers Associated with Squamous Cervical Carcinoma" submitted by you. The Approval number for the same is AUW/IHEC/BT-22-23/XPD-01.

We wish you all the best in your research endeavours.

Regards

5.1.2023

Dr. A Thirumani Devi
Member Secretary

Annexure III

Cervical Cancer: Socio-Demographic, Clinical Profile, and Treatment Experiences

கர்ப்பப்பை வாய்ப் புற்றுநோய்: சமூக-மக்கள்தொகை, மருத்துவ விவரம் மற்றும் சிகிச்சை அனுபவங்கள்

Section 1: Socio-Demographic Information/பிரிவு 1: சமூக-மக்கள்தொகை தகவல்

1. **What is your age group? / 1. உங்கள் வயது என்ன?**

   ☐ Below 30 years / 30 ஆண்டுகளுக்கு கீழ்

   ☐ 30–39 years / 30-39 ஆண்டுகள்

   ☐ 40–49 years / 40-49 ஆண்டுகள்

   ☐ 50–59 years / 50-59 ஆண்டுகள்

   ☐ 60–69 years / 60-69 ஆண்டுகள்

   ☐ 70–79 years / 70-79 ஆண்டுகள்

   ☐ 80 years or older / 80 வயது அல்லது அதற்கு மேற்பட்டவர்கள்

2. **What is your current residence? / 2. உங்கள் தற்போதைய குடியிருப்பு என்ன?**

   ☐ Urban / நகர்ப்புறம்

   ☐ Rural / கிராமப்புறம்

3. **What is your occupation? / 3. உங்கள் தொழில் என்ன?**

   ☐ Daily wage laborer / தினசரி கூலி தொழிலாளி

   ☐ Involved in agriculture / விவசாயம்

   ☐ Working in textiles / ஜவுளி

   ☐ Healthcare worker / சுகாதார பணியாளர்

   ☐ Homemaker / இல்லத்தரசி

   ☐ Teacher / ஆசிரியர்

   ☐ Other (please specify): _____ / மற்றவை: _____

Section 2: Medical and Health History

பிரிவு 2: மருத்துவம் மற்றும் சுகாதார விவரம்

4. **Do you have any family members who have been diagnosed with cancer?/ புற்றுநோயால் பாதிக்கப்பட்டுள்ள குடும்ப உறுப்பினர்கள் யாராவது இருக்கிறார்களா?**

   ☐ Yes / ஆம்

   ☐ No / இல்லை

5. **Have you undergone any surgeries in the past? (Select all that apply) / 5. நீங்கள் கடந்த காலத்தில் ஏதேனும் அறுவை சிகிச்சை செய்திருக்கிறீர்களா? (பொருந்தும் அனைத்தையும் தேர்ந்தெடுக்கவும்)**

   ☐ No previous surgeries/ முந்தைய அறுவை சிகிச்சைகள் இல்லை

   ☐ Cataract surgery / கண்புரை அறுவை சிகிச்சை

   ☐ Hysterectomy/ கருப்பை நீக்கம்

   ☐ Other (please specify): _____ / மற்றவை

   ☐ Not applicable / பொருந்தாது

6. **Have you had any exposure to radiation in your previous treatments? /** உங்கள் முந்தைய சிகிச்சைகளில் கதிர்வீச்சுக்கு ஏதேனும் வெளிப்பாடு இருந்ததா?

   ☐ Yes / ஆம்

   ☐ No / இல்லை

7. **Do you have any other medical conditions (comorbidities)? உங்களுக்கு வேறு ஏதேனும் மருத்துவ நிலைமைகள் (கூட்டு நோய்கள்) உள்ளதா?**

   ☐ Yes / ஆம், if yes , what it is ?/ ஆம் எனில், அது என்ன _____

   ☐ No / இல்லை

8. **What is your current menstrual status?/ உங்கள் தற்போதைய மாதவிடாய் நிலை என்ன?**

   ☐ Post-menopausal/ மாதவிடாய் நின்ற பின்

   ☐ Pre-menopausal/ மாதவிடாய் நிற்கும் முன்

Section 3: Symptoms and Presenting Complaints

பிரிவு 3: அறிகுறிகள்

9. **Which of the following symptoms have you experienced? (Select all that apply) / பின்வரும் எந்த அறிகுறிகளை நீங்கள் அனுபவித்தீர்கள்? (பொருந்தும் அனைத்தையும் தேர்ந்தெடுக்கவும்)**

   ☐ Abdominal pain / வயிற்று வலி

   ☐ Abnormal vaginal discharge (leucorrhea) / அசாதாரண யோனி வெளியேற்றம் (லுகோரியா)

   ☐ Unintentional weight loss / எதிர்பாராத எடை இழப்பு

   ☐ Vaginal bleeding after menopause / மாதவிடாய் நின்ற பிறகு பிறப்புறுப்பு இரத்தப்போக்கு

   ☐ Other abnormal vaginal bleeding / பிற அசாதாரண யோனி இரத்தப்போக்கு

☐ Loss of appetite / பசியின்மை

☐ Back pain / முதுகு வலி

☐ Abdominal swelling / வயிற்று வீக்கம்

☐ Painful urination (dysuria) / வலிமிகுந்த சிறுநீர் கழித்தல்

☐ Changes in bowel habits / குடல் பழக்கவழக்கங்களில் ஏற்படும் மாற்றங்கள்

☐ Other (please specify): _____ / மற்றவை

Section 4: Diagnosis Information (To be filled by Medical Practitioners or Nurses)

10. **What type of cervical cancer have you been diagnosed with? (Select one)**

☐ Squamous cell carcinoma

☐ Adenocarcinoma

☐ Non-keratinizing carcinoma

☐ Invasive carcinoma

☐ Large cell carcinoma

☐ Other (please specify): _____

☐ Not available

Section 5: Treatment Experience (To be filled by Medical Practitioners or Nurses)

11. **Have you received any treatment for your cervical cancer?**

(To be filled by Medical Practitioners or Nurses)

☐ Yes

☐ No (If no, please proceed to the conclusion)

12. **What type of treatment have you undergone? (Select one)**

☐ Monotherapy (single treatment)

☐ Combination therapy (multiple treatments)

---

13. **If you selected Monotherapy, please specify the type of treatment:**

☐ Chemotherapy

☐ Radiotherapy

---

14. **If you selected Combination Therapy, please indicate the treatments you have received:**

☐ Cisplatin with conventional therapy

☐ Carboplatin with intensity-modulated radiotherapy (IMRT)

☐ Paclitaxel with carboplatin and electron beam radiotherapy (ECRT)

☐ Other (please specify): _____

Consent Form

We are conducting a study titled **"Cervical Cancer: Socio-Demographic, Clinical Profile, and Treatment Experiences"** to understand the socio-demographic, clinical, and treatment-related factors associated with cervical cancer. This research aims to contribute to improving awareness, early detection strategies, and healthcare outcomes.

Your participation in this study is completely voluntary. You may refuse to answer any question or withdraw at any time without any consequences. The information you provide will remain strictly confidential and will be used solely for research purposes. No identifying details will be shared publicly or with third parties. The questionnaire will take approximately **10–15 minutes** to complete, and there are no risks or discomforts involved in participating.

By proceeding with this questionnaire, you confirm that you have read and understood the purpose of this study and the information provided. You agree to participate voluntarily and allow your responses to be used for research purposes.

கர்ப்பப்பை வாய்ப் புற்றுநோயுடன் தொடர்புடைய சமூக-மக்கள்தொகை, மருத்துவ மற்றும் சிகிச்சை தொடர்பான காரணிகளைப் புரிந்து கொள்ள, "கர்ப்பப்பை வாய்ப் புற்றுநோய்: சமூக-மக்கள்தொகை, மருத்துவ விவரம் மற்றும் சிகிச்சை அனுபவங்கள்" என்ற தலைப்பில் ஒரு ஆய்வை நடத்துகிறோம். இந்த ஆராய்ச்சி விழிப்புணர்வை மேம்படுத்துதல், முன்கூட்டியே கண்டறிதல் உத்திகள் மற்றும் சுகாதார விளைவுகளை மேம்படுத்துவதை நோக்கமாகக் கொண்டுள்ளது.

இந்த ஆய்வில் உங்கள் பங்கேற்பு முற்றிலும் தன்னார்வமானது. நீங்கள் எந்த கேள்விக்கும் பதிலளிக்க மறுக்கலாம் அல்லது எந்த நேரத்திலும் எந்த விளைவுகளும் இல்லாமல் திரும்பப் பெறலாம். நீங்கள் வழங்கும் தகவல் கண்டிப்பாக ரகசியமாக இருக்கும் மற்றும் ஆராய்ச்சி நோக்கங்களுக்காக மட்டுமே பயன்படுத்தப்படும். அடையாளம் காணும் விவரங்கள் எதுவும் பொதுவில் அல்லது மூன்றாம் தரப்பினருடன் பகிரப்படாது. கேள்வித்தாள் முடிக்க தோராயமாக 10-15 நிமிடங்கள் எடுக்கும், மேலும் பங்கேற்பதில் எந்த இடர்களும் அல்லது அசௌகரியங்களும் இல்லை.

இந்தக் கேள்வித்தாளைத் தொடர்வதன் மூலம், இந்த ஆய்வின் நோக்கம் மற்றும் வழங்கப்பட்ட தகவலைப் படித்துப் புரிந்துகொண்டீர்கள் என்பதை உறுதிப்படுத்துகிறீர்கள். நீங்கள் தானாக முன்வந்து பங்கேற்க ஒப்புக்கொள்கிறீர்கள் மற்றும் உங்கள் பதில்களை ஆராய்ச்சி நோக்கங்களுக்காகப் பயன்படுத்த அனுமதிக்கிறீர்கள்.

Participant Name / பங்கேற்பாளர் பெயர் :_____

Signature/ கையொப்பம் :_____

Date / தேதி: _____

Annexure IV

Sri Ramakrishna Hospital

Medical Service : M/s. S.N.R. SONS CHARITABLE TRUST

SRI RAMAKRISHNA HOSPITAL ETHICAL COMMITTEE

395, SAROJINI NAIDU ROAD, SIDHAPUDUR, COIMBATORE - 641 044.

Phone : 0422 - 4500000, E-mail : ec@sriramakrishnahospital.co.in website : sriramakrishnahospital.com

Ethics Committee Registration No. ECR/690/Inst/TN/2014/RR-18

Ethics Committee Chairman
Dr. Murali. P. M. M.Sc.,Ph.D.,D.Sc.,

Ethics Committee Vice Chairman
Dr. Vimal Veereshwarayya, Ph.D.,RAC.,

Ethics Committee Member Secretary
Dr. Isaac Christian Moses., MD.,FICP.,FACP.,

Ethics Committee Basic Scientist
Dr. Paramasivam. N, MD(Pharm),DA.,

Ethics Committee Clinical Scientist
Dr. Booma. V, MD(Paediatrics)
Dr. Karthikesh. K, MS.,FRCS.,DNB.,M.Ch.,
Dr. Loganathan. N, MBBS.,MD(GM)., DM,
Dr. S. Lokeshwaran, MBBS.,MD., DNB,EDIC, PDCC.,

Ethics Committee Social Scientist
Dr. Nagalingam. M, MSW, Ph.D.,

Ethics Committee Legal Expert
Mr. Sivakumar. V, B.Sc., B.L.,

Ethics Committee Layperson
Mr. Subramanian. V, B.A.,

EC/2020/ 2402/CR/65 11.03.2020

ETHICAL CLEARANCE CERTIFICATE

Project title: ""SOCIO-DEMOGRAPHIC PROFILING AND COMPARATIVE STUDY OF PREVALENCE OF HPV GENOTYPES CAUSING CERVICAL CANCER AMONG RURAL POPULATION IN TAMIL NADU "

Researcher : **MS. B.SUDHA, M.Sc, Research Scholar-Cancer biology, Avinashilingam Institute for Home Science**

The following members of the Ethics Committee were present at the meeting held on 24.02.2020 at 2.30 pm in Auditorium, Sri Ramakrishna Hospital Campus, Coimbatore.

S. No	Members Name	Qualification	Designation	Address	Affiliation to the Institution (Yes/ No)
1.	Dr. Murali.P.M	M.Sc., Ph.D., D.Sc.,	Chairperson	Jananom Private Limited, 26-1 Natesar Nagar, Kovaipudur, Coimbatore	No
2.	Dr. Karthikesh	MS.,FRCS.,DNB.,M.Ch	Active Member Secretary/ Clinical Scientist	Consultant - Oncologist, Sri Ramakrishna Hospital, No. 395, Sarojini Naidu Road, Siddhapudur,CBE	Yes
3.	Dr. Nagalingam.M	MSW, Ph.D.,	Social Scientist	Assistant Professor, Department of Social Work, IGNTU, Amarkantak, Lalpur, Anuppur, MP–484 886	No
4.	Mr. Sivakumar.V	B.Sc., B.L.,	Legal Expert	9 Ground Floor, Parsn Trade Plaza, 156, Dr.Nanjappa Road, CBE	No
5.	Mr. Subramanian.V	BA	Layperson	Supreme mills, ERA Mohan Nagar, Kalapatti Road, Coimbatore	No
6.	Dr. Paramasivan.N	MBBS., MD (Pharma).,	Basic Scientist	Sri Ramakrishna Dental College, S N.R. College Road, Coimbatore	Yes
7.	Dr. Subramanian.K	M.Sc., M.Phil., PhD.,	Statistician	Associate Professor & Head- Department of Statistics PSG College of Arts & Science, Coimbatore	No

Sri Ramakrishna Hospital

Medical Service : M/s. S.N.R. SONS CHARITABLE TRUST

SRI RAMAKRISHNA HOSPITAL ETHICAL COMMITTEE

395, SAROJINI NAIDU ROAD, SIDHAPUDUR, COIMBATORE - 641 044.
Phone : 0422 - 4500000, E-mail : ec@sriramakrishnahospital.co.in website : sriramakrishnahospital.com

Ethics Committee Registration No. ECR/690/Inst/TN/2014/RR-18

8.	Dr. P. Sukumaran	MBBS.MS.MCh	Subject expert	Medical Director Ramakrishna Hospital, 395, Sarojini Naidu road, Sidhapudur, Coimbatore	Yes
9.	Dr. T.K.Ravi,	M.Pharm, Ph.D,FAGE	Subject expert	Principal Sri Ramakrishna College of Pharmacy, 395, Sarojini Naidu road, Sidhapudur, Coimbatore	Yes
10.	Dr. Vasanth Raj	Mpharm,Phd	Subject expert	Research Officer Sri Ramakrishna Hospital, 395, Sarojini Naidu road, Sidhapudur, Coimbatore	Yes
11.	Dr. Prabha Udayakumar	MBBS., MD	Subject Expert	HOD, Anesthesiology Sri Ramakrishna Hospital, 395, Sarojini Naidu road, Sidhapudur, Coimbatore	Yes
12.	Dr. Banumathy	DGO.,DNB.,FICOG	Subject Expert	Consultant-Gynecology, Sri Ramakrishna Hospital, 395, Sarojini Naidu road, Sidhapudur, Coimbatore	Yes
13.	Dr. S. Sriram	Mpharm, Phd	Subject expert	HOD- Pharmacy practice COP, SRIPMS	Yes

This is to certify that the research work entitled "**SOCIO-DEMOGRAPHIC PROFILING AND COMPARATIVE STUDY OF PREVALENCE OF HPV GENOTYPES CAUSING CERVICAL CANCER AMONG RURAL POPULATION IN TAMIL NADU"** placed before the Institutional Ethical Committee and has been approved on **10.03.2020** as there is no objection to hold this research work.

The Ethics committee expects to be informed about the progress of the study, any changes in the protocol and patient information / informed consent and asks to be provided a copy of the final report.

The Ethics committee wishes for the research.

Yours truly,

Member Secretary, Institutional Ethics Committee.

MEMBER SECRETARY
SRI RAMAKRISHNA HOSPITAL ETHICAL COMMITTEE
No. 395, SAROJINI NAIDU ROAD,
SIDDHAPUDUR, COIMBATORE - 641 044

Sri Ramakrishna Hospital

Medical Service : M/s. S.N.R. SONS CHARITABLE TRUST

395, SAROJINI NAIDU ROAD, SIDHAPUDUR, COIMBATORE - 641 044.
Phone : 0422- 4500000,Website :www.sriramakrishnahospital.com
E-mail ID : sriramakrishnahospital@snrsonstrust.org

SRI RAMAKRISHNA HOSPITAL- SCIENTIFIC COMMITTEE

22nd February 2020

MS. B.SUDHA
Research Scholar – Biotechnology
Avinashilingam University,

Dear **MS. B.SUDHA**

The Institutional Scientific Committee of Sri Ramakrishna Hospital reviewed and discussed your application to conduct the study proposal entitled "**SOCIO-DEMOGRAPHIC PROFILING AND COMPARATIVE STUDY OF PREVALENCE OF HPV GENOTYPES CAUSING CERVICAL CANCER AMONG RURAL POPULATION IN TAMIL NADU.**"

The following documents were reviewed:

- Study Protocol
- Study procedure
- Informed consent document in Tamil & English
- Investigator study undertaking

The following members of the Scientific Committee were present at the meeting held on 18.02.2020 & 19.02.2020 (3.00pm to 4.30pm) in Auditorium , Sri Ramakrishna Hospital campus, Coimbatore.

Dr. Sukumaran.P.MS,M.Ch,FIACS	:	Medical Director
Dr. Karthikesh ., MS.,FRCS.,DNB., M.Ch	:	Surgical Oncologist
Dr.Chandramohan.S,DMRD.,DNB	:	Consultant Radiologist
Dr. Prabha Udayakumar , MD,(Anaes)	:	Consultant Anesthesiology
Dr. T.K.Ravi , M.Pharm,Ph.D, FAGE	:	Principal , College of Pharmacy
Dr. Subramanian.K	:	Statistician
Dr. P Vasanthraj	:	Research Officer
Dr. Banumathy , DGO,DNB,FICOG	:	Consultant Gynecologist
Dr. Sarveswaran, MS	:	Consultant Surgeon
Dr. Jambulingam, MD	:	Consultant Physician

The Scientific Committee has no objection for the study to be conducted at Sri Ramakrishna Hospital. The study has been forwarded to Ethical Committee for approval.

Yours truly,

Dr. Sukumaran.P, MS,M.Ch,FIACS., (Medical Director)

Publications

4

Avinashilingam Institute for HomeScience and Higher Education for Women
(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category 'A' by MHRD Re-accredited with A++ Grade by NAAC. CGPA 3.65/4, Category I by UGC, Coimbatore - 641 043, Tamil Nadu, India)

Appendix L2

(Item No 5 of Checklist)

Details of Research Publications

S.No	Article	Journal	Other Details Vol/No/Page No/ Year	Published in UGC CARE / Scopus Indexed/ Web of Science
1 ✓	Absence of Knowledge and Awareness About Cervical Cancer Among Educated Women: A Need for Education About Cervical Cancer	Indian Journal of Gynecologic Oncology	20, 1-8, 2022	Scopus
2 ✓	Identification of hub genes and role of CDKN2A as a biomarker in cervical cancer: An in-silico approach	Human Gene	33, 1-12, 2022	Scopus
3 ✓	Identification of Key Candidate Genes in the Progression of Cervical Cancer: An in Silico Analysis	Indian Journal of Gynecologic Oncology	20, 1-8, 2022	Scopus
4 ✓	Translation of Gene Expression Data Into Personalized Treatment in Cervical Cancer: Machine Learning Approach	Indian Journal of Gynecologic Oncology	22, 1-7, 2024	Scopus

*Proof of list of Journals from Internet to be attached along with copies of reprints.

Scholar : B. Sudha.

Supervisor :

Checked By:

HoD/Dean of Respective School 4/9/2024

The scholar Mrs. Sudha B (19PHBTF001) has published four research articles in the following journals:
1. Indian Journal of Gynecologic oncology - indexed in Scopus and
2. Human Gene - indexed in Scopus.
This may be considered.

04.09.24
Asst Librarian

Indian Journal of Gynecologic Oncology (2022)20:11
https://doi.org/10.1007/s40944-022-00611-6

ORIGINAL ARTICLE

Absence of Knowledge and Awareness About Cervical Cancer Among Educated Women: A Need for Education About Cervical Cancer

Balraj Sudha[1] · Nachimuthu Senthil Kumar[2] · Sundaravadivelu Sumathi[1]

Received: 28 October 2021 / Revised: 10 January 2022 / Accepted: 19 January 2022

Abstract

Purpose Women lack knowledge about HPV infection, vaccines, screening modalities, symptoms, and risk factors which is a major contributor in delayed diagnosis of cervical cancer and short life expectancy. This study aimed to examine the knowledge level about HPV infection, cervical cancer, and preventive measures such as HPV vaccination among women in Tamil Nadu, India.

Methods This was a web-based cross-sectional survey using predesigned questionnaire. Totally, 2100 women registered. Their responses were analyzed with a simple Fisher exact test.

Results More than 50% of people were not aware of HPV infection, cervical cancer, and HPV vaccines, even women who have a better educational background.

Conclusion This study highlights the necessity to spread awareness about cervical cancer and the significant risk factors, symptoms, and preventive methods of cervical cancer.

Keywords Cervical cancer · HPV infection · Human papilloma virus · Sexually transmitted diseases · Awareness

Introduction

Globally, cervical carcinoma is a significant problem in women's health. In 2018, nearly 5.7 lakh cases, and 3.11 lakh people died due to cervical cancer. In Eastern, Western, Middle, and Southern Africa, cervical cancer was the leading cause of cancer-related death in women. Eswatini has the highest rate of cervical cancer, with approximately 65% of women diagnosed with the disease before 75 years. In 146 (79%) of 185 countries, cervical cancer was one of the top three cancers that affected the population under 45 years [1]. Cancer is the second leading cause of death in India, after cardiovascular diseases (CVD). India's five most common cancers among women are breast, cervical, oropharyngeal, lung, and colorectal. India has the highest number of fatalities from cervical cancer. Even though various measures for cervical cancer screening are available, only a few women are aware of this disease. Health education is needed to educate the public about the ill effects of cervical cancer in India [2]. Women in India experience a 1.6% cumulative risk of acquiring cervical cancer and a 1.0% cumulative death incidence. India accounts for approximately one-third of all cervical cancer fatalities worldwide [3]. Cervical cancer is the second most common cancer affecting women in Tamil Nadu, with 80% of women in the low socioeconomic classes, especially in rural areas [4].

HPV infections are also most prevalent among young adults, yet progressing into cervical cancer commonly takes a long time for high-risk HPV infections. HPV is the causative agent of cervical cancer and is a sexually transmitted disease [5]. The most collective reason for the higher proportion of death due to cervical cancer in India was the shortage of responsiveness to screening programs, warning signs, risk factors, and protective methods. The communal clinical characteristic spectrum of cervical carcinoma was an abdominal ache, untiring back pain,

✉ Sundaravadivelu Sumathi
sumathi_bc@avinuty.ac.in

[1] Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, TamilNadu 641043, India

[2] Department of Biotechnology, Mizoram University, Aizawl, Mizoram 796 004, India

Published online: 06 February 2022

bleeding before and after intercourse, urinary haste, white vaginal deport, malodorous vagina, etc. [6]. Many studies clearly explain the connection between HPV infection and cervical carcinoma [7]. All over the world, 70% of the cases of cervical carcinoma were related to HPV infection. Further, the risk factors for cancer in cervix cells were numerous sexual partners, sexual contact at an early age, sustained use of oral contraceptive pills, first pregnancy at an early age, repeated pregnancy, manifold abortions, and smoldering [8]. Although the HPV infection was addressed as the leading risk factor for the progress of cervical cancer, immunization with the HPV vaccine may be one of the modalities to overcome the risk of acquainting cancer of the cervix.

Early detection and precision treatment may decrease the deaths due to cervical cancer. Many reports have revealed that if cervical cancer is detected and cured at reasonable periods, the survival rate of an individual is assured [9]. This decline in the incidence of cervical cancer is due to socio-demographic shifts, such as better education, increased marital age of women, delayed first puberty, and fewer children. As there has been deficient coverage, especially in rural India, the contribution of screening is likely to be limited [10]. On the other hand, in developing countries like India, maximum cases were detected at the final stages, which shortened the survival percentage of individuals. The diagnosis was prevalent at a later stage mainly due to the lack of awareness about HPV infection, vaccination, and access to screening modalities [11]. Unfortunately, in India, the execution of screening series for cervical cancer is quite not conceivable due to lack of awareness, false impression about gynecological syndromes, and lack of national screening plans for cervical cancer [12]. The present work focused on assessing the awareness about HPV infection, vaccine, cervical cancer risk factors, symptoms, and screening strategies for women from rural and urban areas of Tamil Nadu, South India.

Materials and Methods

Study Design and Study Population

This survey was conducted from May 2020 to November 2020, and the majority of the participants were college students. The study group was females in 17 to 54 years age group, and free of cervical cancer. The subjects were enrolled online through popular social media sites such as mail, WhatsApp, and targeted email lists. Twenty-two hundred fifteen people completed the eligibility screening; 70 were ruled ineligible, and 2145 agreed to participate. Finally, 2100 (response rate = 94.8%) participants were enrolled in the study.

Compared to offline surveys, online surveys were paperless, provided participants with privacy, and made it easy to compile the results. Participants completed an online questionnaire using Google Forms, which was divided into sections as socio-demographic profile, knowledge of HPV infection, knowledge of HPV vaccines, understanding of the clinical manifestation of cervical cancer, and knowledge about risk factors. Consent was obtained from the study participants, and the data were collected. We requested and encouraged other women in their family also to participate in the survey. The participants were assured that all information collected during the survey were kept confidential and utilized only for the study. To enhance comprehensibility among the survey respondents, each element in the method was introduced in English accompanied with local language (Tamil) description.

Data Collection

The survey was created with the help of a Google Form. A Google Limited Liability Company (LLC) tool allows users to provide information via a personalized survey or quiz. The collected data was automatically integrated into a dynamic Google excel sheet linked to the survey form. We ensured that all registered participants completed the questionnaire.

Statistical Analysis

Responses about the survey from the particular question were noted. The documented data were evaluated using Microsoft Excel and GraphPad Prism 8 software. The consolidated information was expressed in percentage and means. Fisher's exact test was used to compare the numerical analysis of the results between rural and urban regions. A P value less than 0.05 was considered as statistically significant.

Results

Socio-demographic Characteristics of Enrolled Individuals (n = 2100)

For this study, the participated individuals were categorized into two groups based on those living in community, i.e., rural (800) and urban (1300). Table 1 displays the socio-demographic characteristic spectrum of study people. The mean age of rural participants was 30 and urban participants was 35 years. Among the total of 2100 participants, 800 were from rural and 1300 were from the urban regions. Out of which, 584 (73%) and 829 (63.76%)

Table 1 Socio-demographic characteristic spectrum of study people (n = 2100)

Variable	Rural (%) n = 800	Urban (%) n = 1300
Age		
≤ 20	73	63.769
21–30	21.6	27.92
31–40	3.375	4.5
41–50	1.5	2.6
≥ 51	0.5	1.2
Education		
No formal education	0.125	0.07
Primary	0.5	0.6
Secondary	1.25	0.7
College	98.1	98.5
Marital status		
Married	10.37	11.3
Unmarried	89.6	88.69
Family annual income (in Rupees)		
Below 60,000	39.3	25.3
60,000 to 1,20,000	45	50.5
1,20,000 to 5,00,000	12.25	17.9
Above 5,00,000	3.37	6.2
Profession		
Student	86	85.2
Housewife	3.87	3.1
Working	10.12	11.6

participants were within the age group below 20 years from rural and urban regions, respectively. Majority of the participants from both rural and urban had a college-level education (98.1% and 98.5%, respectively) and were unmarried (89.6% and 88.69%, respectively). Nearly half of the women in both rural and urban group had their annual income between 60,000 and 1,20,000 (45% and 50.5% respectively) (Table 1).

Knowledge About HPV Infection, HPV Vaccines, and Cervical Cancer

The registered participants from both urban and rural areas had possessed partial knowledge about cervical cancer (63.1% and 71.3%, respectively). Familiarity about human papilloma virus among female in rural and urban community was quite good (74.5% and 70.38%, respectively). Awareness about Pap test was not satisfactory among individuals who resided both in rural and urban areas (18.6% and 27.23%, respectively). Majority of people were not getting the proper information about Pap test (81.37% and 72.76%). Attaining information about HPV vaccines and their benefits was very low in both community group peoples (26.75% and 19.76%, respectively). HPV vaccination rate was very low in women who lived in rural and urban regions (11% and 14.7% respectively). 35.75% people were not familiar with the cervical cancer in rural area, while 18.8% people not familiar in urban area.

Participants in urban region (26.6%) were better informed about the HPV infection through educational source which was comparatively higher than respondents from rural background (17.5%). About 16.87% of women in rural and 22.6% of women in urban received the information related to cervical cancer through media. Sharing and acquiring knowledge about cancer in cervix through social media; medical practitioners; and friends and relatives was very low in both rural and urban regions (Table 2).

Regarding attitude of women, only 33.23% of women belonging to urban areas were ready to undergo cytological screening. In rural region, 37.8% of individuals were willing for cytological screening. Nearly above half of the women belonging to rural and urban area were not ready for cytological screening (62.1% and 66.76%, respectively).

Knowledge About Clinical Spectrum of Cervical Cancer

The frequently experienced symptoms in rural females were painful and irregular periods (16.87%); pain/bleeding during or after sexual activity (7%); unusual/smelly discharge from vagina (4.87%); and pain in pelvis (4.6%). The commonly admitted signs and indications of cervical cancer in urban women were painful and irregular menstruation (20.5%); pain/bleeding during or after sexual activity (9.4); and unusual/smelly discharge from vagina (6.1%). Women who lacked awareness about cervical cancer are unable to identify its symptoms at the initial stage. Among the women of both communities, more than half of them were not familiar with the symptoms associated with the progression of cervical cancer as shown in Fig. 1.

Knowledge About Risk Factors

Proficiency about the risk factors associated with the cervical cancer was assessed in women from both communities. In rural community, the most common risk factors were HPV infection (22.87%) and multiple sexual partners (7.25), whereas in urban area, it was HPV infection (24.38%); multiple sexual partners; having weak immune system (4.69%); smoking (4.46%); and long-term use of oral contraceptives (3.3%). Being overweight, early pregnancy, marriage at early age, and repeated pregnancy were addressed as a least known possibility factors of cervical cancer in rural and urban women. About 48.87% women

Table 2 Knowledge about cervical cancer, HPV, and HPV vaccines (n = 2100)

Variable	Rural %	Urban%	P value
Knowledge about cervical cancer			0.0001*
I know	63.1	71.3	
I don't know	36.87	28.69	
Knowledge about HPV			0.0454*
I know	74.5	70.38	
I don't know	25.5	29.6	
Knowledge about Pap test			0.00001*
I know	18.6	27.23	
I don't know	81.37	72.76	
Knowledge about HPV vaccine			0.0002*
I know	26.75	19.76	
I don't know	73.25	80.23	
Receiving HPV vaccine			0.0144*
Yes	11	14.7	
No	89	85.23	
Source of information about cervical cancer			
Television	11	9.9	
Newspaper/magazines	3.75	5.1	
Medical practitioner	5.37	7.3	
Friends/neighbors/relatives	9.75	9.5	
Internet	16.87	22.6	
Education	17.5	26.6	
I don't know	35.75	18.8	
Do you think that you should undergo cytological examination			0.034*
Yes	37.8	33.23	
No	62.1	66.76	

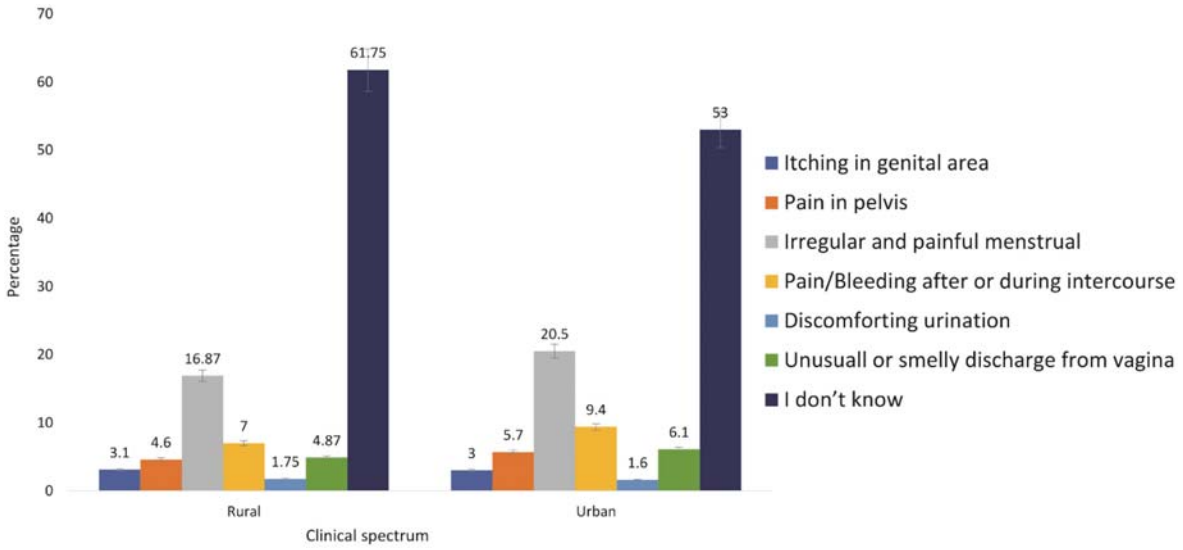


Fig. 1 Assessment of symptoms of cervical cancer in rural and urban women (n = 2100)

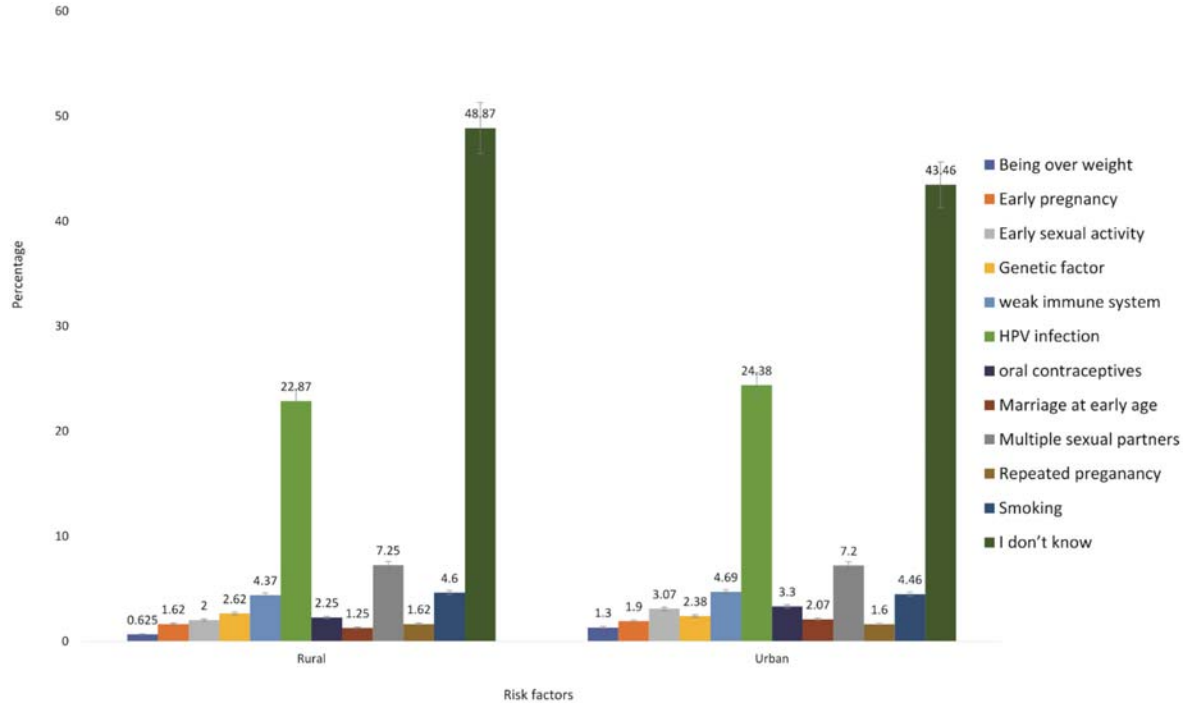


Fig. 2 Knowledge about risk factors associated with cervical cancer in rural and urban female (n = 2100)

from rural and 43.46% from urban women were not aware about the risk factors (Fig. 2).

Discussion

Worldwide, cervical carcinoma is one of the most common cancers that affect women. More than 67,000 cervical cancer deaths occur every year in India due to the absence of knowledge about cervical cancer progression, screening plans, and preventive measures [13]. Embedding awareness and knowledge programs with intervention strategies for cervical cancer or HPV screening and HPV vaccination would also help Indian women reduce HPV infection and control cervical cancer [14].

The earlier reported study focuses on rural women with primary education at the school level and lacks awareness about cervical carcinoma [4, 15]. In rural and urban areas of India, several studies have been conducted to assess young women's awareness of cervical cancer and HPV [16–20]. There is a lack of awareness and knowledge about cervical cancer among young undergraduates and post-graduate students from rural and urban areas [14, 21–23]. Some of the reasons for India's low HPV vaccination rate could be attributed to various factors such as income, sociocultural, and spiritual prejudices. And our study also clearly suggests that a fundamental lack of awareness even among educated women is the leading cause of this type of cancer. In the present survey, we have collected information from around 2100 women both from rural and urban areas, which revealed the latest responsiveness level about cervical carcinoma, its screening methods, clinical spectrum, and possible factors in the rural and urban region of Tamil Nadu, India. In our study population, maximum participants (86% and 85.2%, respectively) were students with college-level education. Despite having a college-level education, only half of the participants knew about cervical carcinoma (63.1% and 71.3%). These results were found to be significant in both communities ($p < 0.05$). Attentiveness about cervical cancer in both communities was not satisfactory. They were hesitant to come forward for primary screening, diagnosis, and cancer treatment in cervix cells. As per the report of a prior study in rural areas of India, about 38% of individuals were familiar with cervical cancer [24]. A survey about cervical cancer awareness among nursing staff in central India stated that 86% of them knew cervical carcinoma. A study conducted in the Pondicherry region revealed that the awareness rate was 45% and 3% for cervical cancer and HPV vaccines [19, 25]. 48% of graduated students from different fields were not even familiar with cervical carcinoma and HPV. 90% of graduated students did not have basic information

about cervical cancer and the major risk factors. A similar trend was observed in developed countries like the USA and UK [26–28]. This highlights the fact that even people with primary education have less knowledge and awareness about cervical carcinoma and immunization protocols.

As per the suggestion of direct detection, the prevalence and death rate would be reduced with cytological screening such as Pap smear test. The decrease in cervical cancer occurrence and successive reduction in death rate can be achieved through screening programs with Pap smear test, which is considered secondary protective measures for early diagnosis in the USA and Canada. Various studies conducted in Greece, Taiwan, and Bhutan declared that the awareness level of cervical carcinoma was 94%, 80%, and 53%, respectively [29–31]. In our present study, most of the individuals were not willing for cytological screening for early detection, which might be due to less knowledge about the significance of the Pap test. Another observation was that the familiarity with the warning signs and indications associated with cervical cancer. But people were not aware that these clinical spectrums were connected with cervical cancer or some other cancer. The maximum documented symptoms in this study population were irregular and painful menstrual period, pain/bleeding during or after sexual activity, unusual/smelly discharge from the vagina, and pain in the pelvis. Many women consider signs such as pelvic pain, vaginal discharge, and malodorous vagina which was normal. So, there was no necessity to worry about the symptoms, as mentioned earlier.

From the results of this present study, HPV infection, multiple sexual partners, smoking, a weak immune system, and genetic factors were documented as possible factors among the rural community for cervical cancer. On the other hand, HPV infection, multiple sexual partners, a weak immune system, smoking, long-term use of oral contraceptives, early start of sexual activity, and early pregnancy were considered predominant cancer in urban areas. A study done in Thailand pointed out that first sexual intercourse at a young age (81.8%) and having multiple sexual partners (85.6%) were possible factors for cervical cancer [32]. Koshy et al., (2017) reported that 3% of the females were aware of a maximum of the possible factors involved in cervical cancer, while half of the participants did not know about even one possibility factor of cervical cancer. Kadian et al. (2020) found that most women in North India were not aware of HPV infection that was a significant risk factor for cancer progression. About 60% of women have not heard of cervical carcinoma in Tamil Nadu [4]. In another study conducted in 2020, 85% of women had poor knowledge of cervical cancer, and fewer than 25% were aware of cervical cancer signs, risk factors, or preventive strategies [15]. Knowledge level regarding cervical cancer was the same Worldwide [5, 19, 26–28, 33–35].

In contrast to other countries, no national HPV immunization initiatives have been introduced in India, where the majority of the population is positive for high-risk oncogenic HPV types, specifically HPV type 16 [36–38]. PATH, in collaboration with the Indian Council of Medical Research, conducted a small HPV vaccine immunogenicity experiment in the Indian states of Andhra Pradesh and Gujarat, but these were halted due to entirely irrelevant mortalities, clinical, safety, and ethical challenges. As a result, HPV vaccination in India is limited to medical centers, and vaccine normalization is primarily limited among the educated population. Studies have found that the majority of participants or their parents are initially reluctant to have their daughters immunized, leading to a shortage of knowledge about the HPV vaccine, its effectiveness and safety, as well as sociocultural, financial, and spiritual concerns [17, 39, 40].

The outcome of this study showed that the knowledge on HPV vaccination was very insignificant among both rural and urban women. Proper medical supervision and awareness programs are almost non-existent in our study population. In advance, HPV vaccine introduction in the National Immunization Program, there is a necessity to reduce the price and time of vaccine use [41]. As a result, an initial assessment is required to increase awareness and understanding of HPV-related cancers and their prevention through immunization in adolescent and teenage girls, destigmatize HPV infection, and ensure the successful execution of screening and vaccination programs to minimize cervical cancer death rates and burden. It may also help prevent high-risk HPV-related diseases like head and neck cancer, esophageal cancer, and other common cancers in India. Accordingly, a good vaccination program series constructed depending upon the local epidemiological HPV profile will minimize the cervical cancer problem in women of Tamil Nadu, India. Cervical cancer is a preventable malignancy with little attention and care taken by women.

Conclusion

Cervical cancer is easily preventable with suitable vaccination and continuous screening. So, every woman should be aware and educate themselves and their family members with nurses and medical care officers; conferences and talks about HPV-related diseases in educational institution should be organized; social workers like Anganwadi and ASHA workers; NGOs and women representatives should create awareness. The present study is the first large-scale survey report of urban and rural educated women which

highlights that there is a requisite to organize repetitive screening test and awareness program for HPV vaccination for reducing burden of cervical cancer. It could be a preventive measure compared to other strategies, and endeavor should be arranged to practice it properly and reduce the death rate gradually.

Acknowledgements We thanks to all the participants for their support of our study.

Authors Contributions All authors conceived and designed the study. BS, SS, and NSK conducted research, provided research materials, and collected data. BS organized data. BS and SS wrote initial draft. BS, NSK, and SS analyzed and interpreted data, and all authors for final draft of article, and provided logistic support. All authors have critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to this study.

References

1. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, Bray F. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob Health. 2020;8(2):e191–203. https://doi.org/10.1016/S2214-109X(19)30482-6.
2. Manikandan S, Behera S, Naidu NM, Angamuthu V, Mohammed OFB, Debata A. Knowledge and awareness toward cervical cancer screening and prevention among the professional college female students. J Pharm Bioallied Sci. 2019;11(2):S314.
3. Mishra R. An epidemiological study of cervical and breast screening in India: district-level analysis. BMC Womens Health. 2020;20(1):1–15. https://doi.org/10.1186/s12905-020-01083-6.
4. Tamilarasi R, Maheshwari L, Siddharth R, Sanjeev. A study of awareness on cervical cancer and prevalence of pathological leucorrhoea among women in rural Chennai. Int J Commun Med Public Health. 2018;5:2118–22.
5. Ebu NI, Abotsi-Foli GE, Gakpo DF. Nurses' and midwives' knowledge, attitudes, and acceptance regarding human papillomavirus vaccination in Ghana: a cross-sectional study. BMC Nurs. 2021;20(1):1–10. https://doi.org/10.1186/s12912-020-00530-x.
6. WHO. Comprehensive cervical cancer control. A guide to essential practice. WHO, Geneva. 2006; http://www.who.int/reproductivehealth/publications/cervical_cancer_gep/index.html. Accessed June 2021.
7. WHO. Human Papillomavirus (HPV). 2020. http://www.who.int/immunization/diseases/hpv/en/. Accessed June 2021.
8. Biswas LN, Manna B, Maiti PK, Sengupta S. Sexual risk factors for cervical cancer among rural Indian women: a case-control study. Int J Epidemiol. 1997;26(3):491–5. https://doi.org/10.1093/ije/26.3.491.
9. Tsikouras P, Zervoudis S, Manav B, Tomara E, Iatrakis G, Romanidis C, Bothou A, Galazios G. Cervical cancer: screening, diagnosis and staging. J Buon. 2016;21(2):320–5.
10. Van Dyne E A. Establishing baseline cervical cancer screening coverage—India, 2015–2016. In: MMWR morbidity and mortality weekly report. 2019; 68: 14–19. https://doi.org/10.15585/mmwr.mm6801a4externalicon.
11. Mishra GA, Pimple SA, Shastri SS. An overview of prevention and early detection of cervical cancers. Indian J Med Paediatr Oncol. 2011;32(3):125. https://doi.org/10.4103/0971-5851.92808.
12. Narayana G, Suchitra MJ, Sunanda G, Ramaiah JD, Kumar BP, Veerabhadrappa KV. Knowledge, attitude, and practice toward cervical cancer among women attending Obstetrics and Gynecology Department: a cross-sectional, hospital-based survey in South India. Indian J Cancer. 2017;54(2):481. https://doi.org/10.1186/s12905-020-01083-6.
13. Kaarthigeyan K. Cervical cancer in India and HPV vaccination. Indian J Med Paediatr Oncol. 2012;33(1):7. https://doi.org/10.4103/0971-5851.96961.
14. Rashid S, Labani S, Das BC. Knowledge, awareness and attitude on HPV, HPV vaccine and cervical cancer among the college students in India. PLoS ONE. 2016;11(11):e0166713. https://doi.org/10.1371/journal.pone.0166713.
15. Reichheld A, Mukherjee PK, Rahman SM, David KV, Pricilla RA. Prevalence of cervical cancer screening and awareness among women in an urban community in South India—a cross sectional study. Ann Glob Health. 2020;86(1):30. https://doi.org/10.5334/aogh.2735.
16. Asthana S, Satyanarayana L. Cervical cancer screening from the perspective of a rural population of northern India. Curr Med Res Pract. 2013;3(3):128.
17. Hussain S, Bharadwaj M, Nasare V, Kumari M, Sharma S, Hedau S, Das BC. Human papillomavirus infection among young adolescents in India: impact of vaccination. J Med Virol. 2012;84(2):298–305. https://doi.org/10.1002/jmv.22261.
18. Mehta S, Rajaram S, Goel G, Goel N. Awareness about human papilloma virus and its vaccine among medical students. Indian J Commun Med. 2013;38(2):92. https://doi.org/10.4103/0970-0218.112438.
19. Saha A, Chaudhury AN, Bhowmik P, Chatterjee R. Awareness of cervical cancer among female students of premier colleges in Kolkata. India Asian Pac J Cancer Prev. 2010;11(4):1085–90.
20. Sharma K, Kathait A, Jain A, Kujur K, Raghuwanshi S, Bharti AC, Saklani AC, Das BC. Higher prevalence of human papillomavirus infection in adolescent and young adult girls belonging to different Indian tribes with varied socio-sexual lifestyle. PLoS ONE. 2015;10(5):e0125693. https://doi.org/10.1371/journal.pone.0125693.
21. Das BC, Hussain S, Nasare V, Bharadwaj M. Prospects and prejudices of human papillomavirus vaccines in India. Vaccine. 2008;26(22):2669–79. https://doi.org/10.1016/j.vaccine.2008.03.056.
22. Hussain S, Nasare V, Kumari M, Sharma S, Khan MA, Das BC, Bharadwaj M. Perception of human papillomavirus infection, cervical cancer and HPV vaccination in North Indian population. PLoS ONE. 2014;9(11):e112861. https://doi.org/10.1371/journal.pone.0112861.
23. Prusty BK, Kumar A, Arora R, Batra S, Das BC. Human papillomavirus (HPV) DNA detection in self-collected urine. Int J Gynecol Obstet. 2005;90(3):223–7. https://doi.org/10.1016/j.ijgo.2005.06.004.
24. Arunadevi V, Prasad G. Knowledge and awareness of cervical cancer among women in rural India. Int J Curr Res Rev. 2015;7(21):29. https://doi.org/10.31782/2231-2196.
25. Jain SM, Bagde MN, Bagde ND. Awareness of cervical cancer and Pap smear among nursing staff at a rural tertiary care hospital in Central India. Indian J Cancer. 2016;53(1):63. https://doi.org/10.4103/0019-509X.180823.
26. Ingledue K, Cottrell R, Bernard A. College women's knowledge, perceptions, and preventive behaviors regarding human

papillomavirus infection and cervical cancer. Am J Health Stud. 2004;19(1):28.
27. Philips Z, Johnson S, Avis M, Whynes DK. Human papillomavirus and the value of screening: young women's knowledge of cervical cancer. Health Educ Res. 2003;18(3):318–28. https://doi.org/10.1093/her/cyf022.
28. Yacobi E, Tennant C, Ferrante J, Pal N, Roetzheim R. University students' knowledge and awareness of HPV. Prev Med. 1999;28(6):535–41. https://doi.org/10.1006/pmed.1999.0486.
29. Bakogianni GD, Goutsou SC, Liti MV, Rizopoulou SI, Nikolakopoulos KM, Nikolakopoulou NM. Knowledge, attitude, and practice of cervical cancer screening among Greek students: a short report. Int J Adolesc Med Health. 2012;24(4):329–30. https://doi.org/10.1515/ijamh.2012.047.
30. Dhendup T, Tshering P. Cervical cancer knowledge and screening behaviors among female university graduates of year 2012 attending national graduate orientation program, Bhutan. BMC Women's Health. 2014;14(1):1–7. https://doi.org/10.1186/1472-6874-14-44.
31. Hsu YY, Cheng YM, Hsu KF, Fetzer SJ, Chou CY. Knowledge and beliefs about cervical cancer and human papillomavirus among Taiwanese undergraduate women. Oncol Nurs Forum. 2011;38(4):E297–304.
32. Nganwai P, Truadpon P, Inpa C, Sangpetngam B, Mekjarasnapa M, Apirakarn M, Chumworathayi B. Knowledge, attitudes and practices vis-a-vis cervical cancer among registered nurses at the Faculty of Medicine, Khon Kaen University, Thailand. Asian Pac J Cancer Prev. 2008;9(1):15–8.
33. Al-Darwish AA, Al-Naim AF, Al-Mulhim KS, Al-Otaibi NK, Morsi MS, Aleem AM. Knowledge about cervical cancer early warning signs and symptoms, risk factors and vaccination among students at a medical school in Al-Ahsa, Kingdom of Saudi Arabia. Asian Pac J Cancer Prev. 2014;15(6):2529–32. https://doi.org/10.7314/APJCP.2014.15.6.2529.
34. Aswathy S, Quereshi MA, Kurian B, Leelamoni K. Cervical cancer screening: current knowledge & practice among women in a rural population of Kerala, India. Ind J Med Res. 2012;136(2):205–10.
35. Siddharthar J, Rajkumar B, Deivasigamani K. Knowledge, awareness and prevention of cervical cancer among women attending a tertiary care hospital in Puducherry, India. J Clin Diagn Res. 2014;8(6):01–3. https://doi.org/10.7860/JCDR/2014/8115.4422.
36. Patel KR, Vajaria BN, Begum R, Desai A, Patel JB, Shah FD, Shukla SN, Patel PS. Prevalence of high-risk human papillomavirus type 16 and 18 in oral and cervical cancers in population from Gujarat, West India. J Oral Pathol Med. 2014;43(4):293–7. https://doi.org/10.1111/jop.12147.
37. Sureshkumar BT, Shanmughapriya S, Das BC, Natarajaseenivasan K. A population-based study of the prevalence of HPV in three districts of Tamil Nadu. India Int J Gynecol Obstet. 2015;129(1):58–61. https://doi.org/10.1016/j.ijgo.2014.10.025.
38. Vinodhini K, Shanmughapriya S, Sanmugham S, Senthikumar G, Das BC, Natarajaseenivasan K. Prevalence of high-risk HPV and associated risk factors in cases of cervical carcinoma in Tamil Nadu, India. Int J Gynecol Obstet. 2012;119:253–6. https://doi.org/10.1016/j.ijgo.2012.06.019.
39. Lee PW, Kwan TT, Tam KF, Chan KK, Young PM, Lo SS, Cheung AN, Ngan HY. Beliefs about cervical cancer and human papillomavirus (HPV) and acceptability of HPV vaccination among Chinese women in Hong Kong. Prev Med. 2007;45(2–3):130–4. https://doi.org/10.1016/j.ypmed.2007.07.013.
40. Woodhall SC, Lehtinen M, Verho T, Huhtala H, Hokkanen M, Kosunen E. Anticipated acceptance of HPV vaccination at the baseline of implementation: a survey of parental and adolescent knowledge and attitudes in Finland. J Adolesc Health. 2007;40:466–9. https://doi.org/10.1016/j.jadohealth.2007.01.005.
41. Chatterjee S, Chattopadhyay A, Samanta L. HPV and cervical cancer epidemiology-current status of HPV vaccination in India. Asian Pac J Cancer Prev. 2016;17(8):3663–73. https://doi.org/10.14456/apjcp.2016.156/APJCP.2016.17.8.3663.

Current Problems in Cancer xxx (xxxx) 101157

Contents lists available at ScienceDirect

Current Problems in Cancer

journal homepage: www.elsevier.com

Journal logo

Women's ignorance and misperception of cervical cancer: Evidence-based analysis from low- and middle-income countries

Balraj Sudha [a], Nachimuthu Senthil Kumar [b], Sundaravadivelu Sumathi [a, *]

[a] Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Institute for Home science and Higher Education for Women, Coimbatore, 641 043, Tamil Nadu, India
[b] Department of Biotechnology, Mizoram University, Aizawl, 796 004, Mizoram, India

ARTICLE INFO

Keywords:
Cervical cancer
Pap smear
HPV vaccine

ABSTRACT

Cervical cancer is a major cause of morbidity and mortality particularly in low- and mid-income countries. This review synthesizes existing knowledge on cervical cancer, HPV infection, and HPV vaccination accumulated over the past decade (2015-2024), highlighting disparities in awareness and prevention strategies globally. Education level correlates with HPV vaccine awareness, yet mere familiarity with cervical cancer doesn't ensure understanding of its severity. Notably, prevention measures, including screening and HPV testing, varied significantly across countries during this period. To enhance HPV vaccine uptake moving forward, targeted efforts are necessary to educate women, particularly in low- and mid-income countries, about HPV risks as a sexually transmitted disease and the availability of affordable vaccines in government clinics.

Introduction

Cancer is the fourth-largest non-communicable disease among current epidemics and contributes significantly to overall mortality. Cervical cancer is the fourth most significant cause of cancer death, with an expected 604,000 new cases and 342,000 deaths in 2020.[1] Early marriage, childbirth before the age of 20, poor personal hygiene, multiple sexual partners, low socioeconomic status, smoking, the presence of the Human Papilloma Virus and Herpes Simplex Virus type II, use of the oral contraceptive pill, and other risk factors are known to increase the incidence of cervical cancer.[2]

Preventing HPV infection through sex abstinence is a primary method of preventing cervical cancer. In addition to following safer sex practices like routinely using latex condoms, early vaccination of female teenagers with a vaccine that is effective against primarily HPV types 16 and 18 strains may also be very helpful.[3] A sexually active woman may be infected with HPV and may, therefore, likely not benefit from vaccination. People who begin sexual activity earlier in life and have several partners are more likely to contract HPV and develop cervical cancer 5 to 20 years later. Epidemiological evidence identifies cigarette smoking as an independent risk factor for cervical cancer, distinct from the well-established etiologic role of human papillomavirus (HPV) infection.[4]

Regular Pap smear exams for sexually active females are another effective secondary prevention strategy for cervical cancer. Exfoliates of cervical cells are collected using a cotton swab and examined under a microscope. Early stages of cancer can be identified and treated before they develop into invasive cancer, such as in-situ malignancy. To sum up, cervical cancer prevention requires HPV vaccination before sexual activity begins, consistent use of latex condoms after that, a low number of sex partners, annual Pap smear

* Corresponding author.
E-mail address: sumathi_bc@avinuty.ac.in (S. Sumathi).

https://doi.org/10.1016/j.currproblcancer.2024.101157
Received 21 August 2024; Received in revised form 16 October 2024; Accepted 5 November 2024
0147-0272/

screening, and all these actions cannot be substituted for one another.[5] Women's ignorance of HPV infection, immunizations, screening methods, symptoms, and risk factors play a significant role in cervical cancer detection delayed and shorter life expectancy.[6]

As part of the global strategy to eliminate cervical cancer, countries are mandated to attain and sustain a predefined incidence threshold, as outlined by the World Health Organization's (WHO) cervical cancer elimination initiative. Policies that address the aims must be devised to achieve them. The target population must have accurate information and knowledge to make decisions that would avoid the disease and consider the methods in place to treat and manage cervical disease. The impact of cervical cancer would eventually decline due to these actions. Therefore, nations must create policies and educational initiatives to inform and educate the populace.[7] Unfortunately, these preventative techniques have not yet been consistently applied in low- and middle-income countries, and more research on the viability of scaling up these prevention activities is required.[8]

Effective treatment and improved outcomes for many cancer cases can be achieved through simple, affordable primary, secondary, and tertiary prevention strategies. Giving early detection and screening programs, as well as public awareness campaigns, more momentum could help to promote preventive measures. Numerous studies conducted in developed nations demonstrate a high correlation between early reporting for screening and community treatment. As a result, research should be done to assess the knowledge levels of various populations to create an efficient program. The current review summarises recent works that evaluate cervical cancer risk factors and raise public awareness of the disease globally.

Materials and methods

Search strategy and screening of paper

As illustrated in Fig. 1, this systematic review utilized the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to comprehensively review the literature on cervical cancer knowledge, attitudes, and prevention strategies.[9]

A systematic search of PubMed and Google Scholar databases was performed to identify published studies examining knowledge, attitudes, and preventive practices related to cervical cancer. Further, the articles were selected following inclusion criteria. The chosen articles contained information about sociodemographic detail and their understanding of the symptoms, risk factors and prevention strategies. Keywords used for the search were “cervical cancer,” “Cervical cancer prevention,” “HPV infection,” and “cervical cancer screening.” To ensure reliability and consistency, this study utilized a rigorous search protocol. The search was conducted multiple times to verify the stability of results, validate search terms, and confirm the accuracy of findings. This iterative process allowed for refinement of keywords and verification of result consistency, ultimately establishing the dependability of the results. By repeating the search process, we confirmed that our findings were reliable and consistent.

Study inclusion and exclusion criteria

In our review, we incorporated a cross-sectional study from different global populations that examined knowledge, attitudes, and screening practices for cervical cancer. The inclusion was limited to publications published from 2015 to 2024 in English across a nine-year timeframe. Studies with unclear outcome measurements, case reports, study populations other than females, women infected with HIV, duplicates, and review papers were also excluded.

Results and discussion

Study characteristics

A comprehensive review of 25 studies conducted across various countries assessed cervical cancer awareness, knowledge, and attitudes among different populations. The studies, conducted between 2011 and 2022, involved sample sizes ranging from 150 to 2,100 participants, covering women aged 15-75 years with varying literacy levels, marital status, and socioeconomic backgrounds.

Knowledge about cervical cancer, HPV, symptoms, and risk factors

Knowing cervical cancer is very much important among women. Even educated women are not much aware of the severity of cervical cancer, clinical signs and symptoms, risk factors and prevention modalities. Table 1 summarises the sociodemographic profile of different study population. The following content elaborates on the connections between the variables.

Knowledge about HPV and cervical cancer

Human Papillomavirus (HPV) is the most common sexually transmitted infection, yet many women lack awareness about its connection to cervical cancer. Recent studies conducted across various countries reveal varying levels of knowledge regarding cervical cancer and HPV.

In some countries, awareness levels are encouragingly high. For instance, in Kuwait, 89.7 % of women have heard of cervical cancer. Similarly, in Bangladesh, studies have consistently shown high awareness, with 81 % to 87 % of women aware of cervical cancer. India also reports significant awareness, with 71.3 % of urban women and 63.1 % of rural women knowledgeable about cervical can-

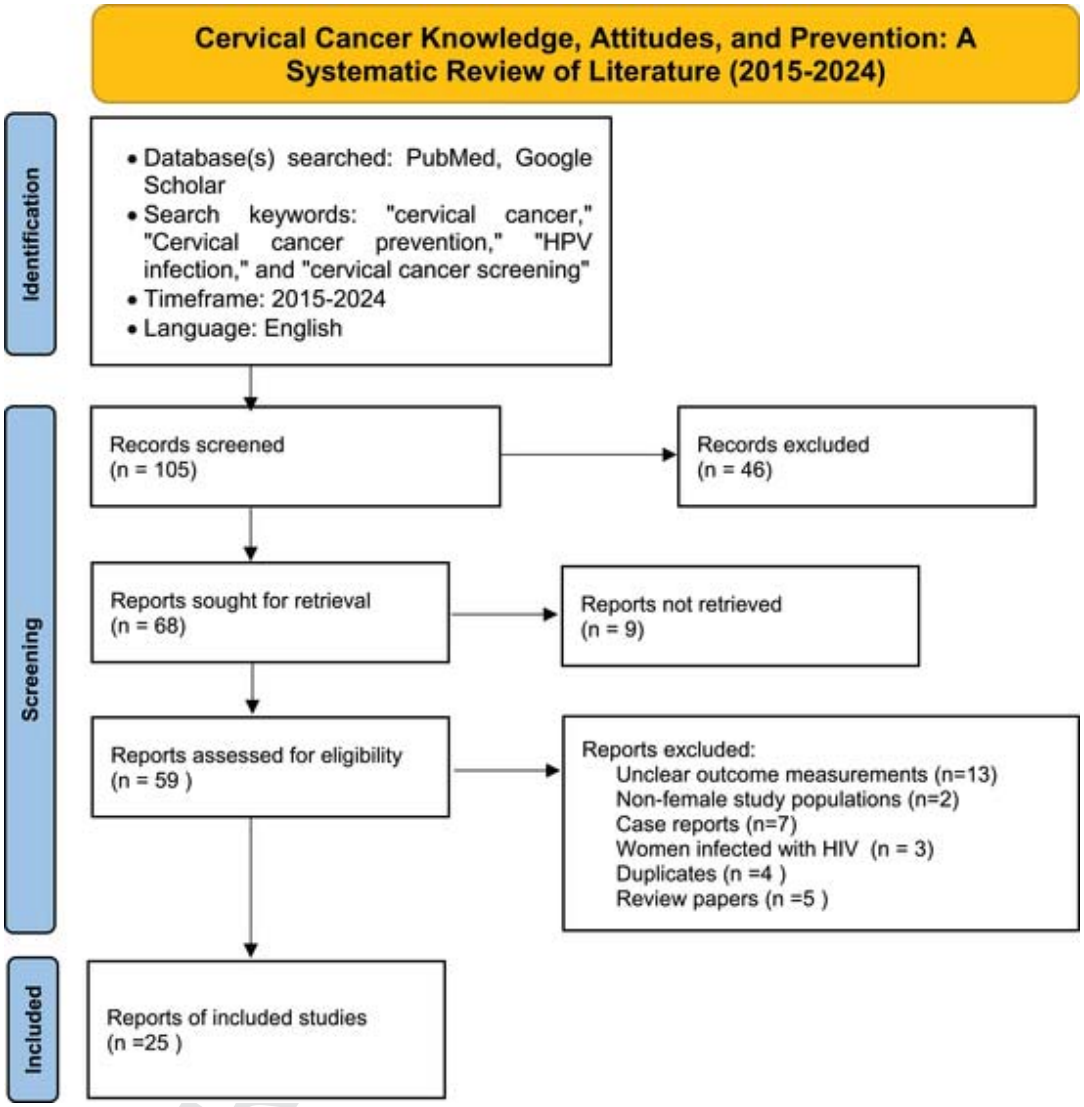


Fig. 1. PRISMA Flow diagram: Study selection process for cervical cancer knowledge, attitudes, and prevention studies (2015-2024).

cer. Central Ethiopia and Cameroon also demonstrate moderate to high awareness, with 68.8 % and 58 % of women aware of cervical cancer, respectively.

However, other regions report concerning gaps in knowledge. In South Africa, a staggering 73.8 % of women are unaware of what HPV is, and only 45.6 % recognize its connection to cervical cancer. Ghanaian women also exhibit limited understanding, with 69.7 % having insufficient knowledge and only 9.7 % demonstrating high knowledge. A study in Kolkata, India, found that merely 15 % of women were aware of cervical cancer, while 36 % knew about HPV.

Literacy and place of residence significantly influence cervical cancer knowledge. Educated women in India and Ethiopia show partial knowledge, indicating room for improvement. The varying awareness levels across countries and regions underscore the need for targeted education and awareness campaigns.

Overall, these findings highlight the importance of addressing knowledge gaps and promoting awareness about cervical cancer and HPV to facilitate effective prevention and control strategies.

Table 1
The sociodemographic profile of different study population.

S.No	Place	Year of study	Sample size	Population characterstics	Findings	Reference
1	Bangladesh.	2022	1090	Age - 18 - 65 years, Literacy status - 28.9 %, Marital status - 77.8 % married, Socioeconomic status - 56.8 % belongs to middle income	**1.** Knowledge of cervical cancer - 45.2 %; **2.** Knowledge of signs and symptoms - Bleeding in between periods, Menstrual period that are longer than as usual; **3.** Knowledge of risk factors - Long term use of contraceptive pill 44.6 %, Starting to have sex at a young age (before age 17) 57.1 %	[10]
2	India	2016 to 2018	1500	Age - 15 - 75 years, Literacy status - 79.8 % college level, Marital status - 52.4 % married	**1.** Knowledge of cervical cancer - 26.5 %; **2.** Knowledge about pap test - 20.6; **3.** Knowledge about HPV- 10 %; Knowledge about HPV vaccine - 2.6 %; **4.** Knowledge of signs and symptoms - Painful/irregular menstruation, Pain/bleeding during or after sexual activity; **5.** Knowledge of risk factors - Long term use of contraceptive pills, Early start of sexual activity, Multiple sexual partners	[11]
3	Ethiopia	2014	667	Age 15 - 30 years, Literacy status - 100 %	**1.** Heard about cervical cancer - 60.6 %; **2.** Knowledge of signs and symptoms - Vaginal bleeding - 22.0 %, Vaginal foul smelling 31.6 %; **3.** Knowledge about risk factors - Multiple sex partners - 52.2 %; **4.** Awareness of pap smear screening - 23.8 %; **5.** Attitude toward screening - 68.9 %	[12]
4	India	2012 – 2013	354	Age - 21 – 40 years, Marital status - 45.2 % married	**1.** Heard about cervical cancer - 90.3 %; **2.** Knowledge about risk factors - Multiple sex partners - 52.2 %; **3.** Knowledge of signs and symptoms - Offensive foul-smelling discharge - 63.8 %, Irregular bleeding - 50.6 %; **4.** Awareness of pap smear screening - 79.1 %; **5.** 16.6 % had undergone a Pap test	[13]
5	India	2012	373	Age - 30 – 60 years, Literacy status - 65 % illitrate	**1.** Aware of cervical cancer - 53.8 %; **2.** Knowledge of signs and symptoms - Discomfort during intercourse - 45.7 %, Postmenopausal bleeding 44.7 %; **3.** Konwledge of risk factors – Non maintenance of menstrual hygiene - 53.2 %; **4.** Knowledge of Pap smear - 7 %	[14]
6	Bangladesh	2011	1590	Age - 30–59 years, Literacy status - 12 % higher level, Marital status - 88.9 % married	**1.** Heard of cervical cancer - 81.3 %; **2.** Heard of cervical cancer screening - 59.8 %	[15]
7	India	2011	224	Age - 20 – 59 years, Marital status - 92.4 % married, Socioeconomic status - 39.2 % belongs to low income	**1.** Heard about cervical cancer - 72.7 %; **2.** Knowledge about risk factors - Genital infection – HPV, HIV, Chlamydia -39.20 %, Early age of first pregnancy - 33.9 %	[16]
8	Nigeria	2013	317	Age - <30 – 89 years, Marital status - 62.8 % married, Literacy status - 90.5 % literate,	**1.** Heard of cervical cancer - 37.2 %; **2.** Heard of HPV vaccine - 6 %; Amongst those aware of cervical cancer, doing Pap smear - 8.5 %; **3.** Knowledge of risk factor - HPV 5.1 %;	[17]
9	central Ethiopia	2020	414	Age - 18 – 49 years; Marital status - 61.4 % married; Literacy status - 33.3 % high school;	**1.** Heard of cervical cancer - 68.8 %; **2.** Knowledge of symptoms- Vaginal bleeding 20.5 %, Vaginal foul-smelling discharges - 32.1 %; **3.** Knowledge of risk factor - Early sexual intercourse - 20 %, Having multiple sexual partners - 17.6 %; **4.** Knowledge of HPV vaccine - 7.2 %; **5.** Knowledge of pap smear - 6.8 %	[18]

(*continued on next page*)

1. Table 1 (*continued*)

S.No	Place	Year of study	Sample size	Population characterstics	Findings	Reference
10	Ethiopia	2018	410	Age - 15 to 49 years; Marital status - 70.2 % married; Literacy status - 20.6 % college and above;	**1.** Knowledge of cervical cancer - 31 %; **2.** Positive attitude toward screening - 57.8 %	[19]
11	Nigeria	NA	305	Age - 15 to 49 years; Marital status - 73.1 % married; Literacy status - secondary education - 54.1 %, tertiary education - 8.5 %;	**1.** Heard of cervical cancer - 12.8 %; **2.** Knowledge of risk factor - Early age at first sex - 3.6 %; **3.** Knowledge of symptoms - Foul smelling vaginal discharge - 5.6 %; **4.** Heard of cervical cancer screening - 7.9 %; **5.** Positive attitude toward screening - 88.9 %	[20]
12	Bangladesh	NA	956	Age - 30 years above; Marital status - 88 % married; Literacy status - Primary education - 73.1 %, secondary education - 3 %;	**1.** Knowledge of cervical cancer - 87 %; **2.** Knowledge of risk factor - Early marriage - 55 %; **3.** Knowledge of symptoms - Heaviness in lower abdomen - 30 %; **4.** Knowledge of treatment of Ca cervix - 65 %; **5.** Knowledge of Cervical cancer vaccine - 37 %;	[21]
13	North-East Nigeria	2014	978	Age - 15-49 years; Marital status - 57.9 % married; Literacy status - Tertiary education - 28.1 %	**1.** Knowledge of cervical cancer - 73.2 %; **2.** Knowledge of risk factor - 88.1 %; **3.** Knowledge of symptoms - 76.5 %; **4.** Knowledge Cervical cancer prevention - 90.3 %; **5.** Knowledge of Cervical cancer vaccine - 37 %;	[22]
14	Northern Tanzania	2017 – 2019	2192	Age - <25-55+ years; Literacy status - Secondary and above - 28.7 %	**1.** Knowledge of cervical cancer - 22.1 %; **2.** Aware of HPV infection - 22.8 %; **3.** Aware of HPV vaccines - 46.4 %; **4.** Knowledge of HPV - 29.4 %; **5.** Ever screened for CC - 34.0 %	[23]
15	Ghana	NA	200	Age - 15 – above 40 years, Marital status - 46 % married, Literacy status - 29.5 % illitrate,	**1.** Heard of Cervical cancer - 55.5 %; **2.** Knowledge about signs and symptoms - Bleeding between periods - 25 %, Severe lower abdominal pain - 21.5 %, Unusual vaginal discharge - 20 %; **3.** Knowledge about risk factors: Having many sexual partners - 52 %, Having a weakened immune system - 30 %, Infection with HPV - 28 %	[24]
16	India	2015	1020	Age - 20 – 49 years, Marital status - 50.2 % married, Literacy status - 48.9 % higher,	**1.** Good knowledge about cervical cancer sign and symptoms - 45.5 %, **2.** risk factors - 40.1 %	[25]
17	India	2014 – 2017	1140	Age - 20 – 65 years, Marital status - 89.3 % married, Literacy status - 34 % illitrates, Socio economic status - low 61.1 %	**1.** Heard of cervical cancer - 82.9 %; **2.** Knowledge regarding cervical cancer - 25 %; **3.** Positive attitude towards screening - 99.9 %	[26]
18	western Ethiopia	2017	830	Age - 17 – 26 years, Marital status - 4.5 % married, Literacy status - engineering students - 42.7 %	**1.** Heard about cervical cancer and its risk factors - 54.4 %; **2.** Knowledge about pap smear - 61.1 %, visual inspection with acetic acid - 38.8 %; **3.** Positive attitude towards screening - 44.1 %	[27]
19	Bangladesh	2019 – 2020	600	Age - 15 and 29 years, ≥ 30 years; Marital status - 62.8 % married; Literacy status - Higher education - 35.3 %; Socioeconomic status - About 70 % belongs to low income	**1.** Heard about cervical cancer - 71.8 %, **2.** Among them, 2.3 % had undergone screening; **3.** Vaccinated women - 5.3 %; **4.** Positive attitude about HPV vaccine - 76.6 %; **5.** Knowledge of risk factors - multiple sexual partners - 63.6 %; **6.** Knowledge about sign and symptoms - Increased vaginal discharge - 64 %	[28]
20	Kuwait	2020 - 2021	250	Age - 17 - 26 years; Marital status - 18.5 % married; Literacy status - 100 %;	**1.** Heard about cervical cancer - 89.7 %, **2.** Heard about HPV vaccine - 23.3 %, **3.** Vaccinated participants - 5.8 %; **4.** Heard about cytological examination - 78.8 %;	[29]
21	Ghana	2015	288	Age - 19 to 64 years; Marital status - 48.3 % married; Literacy status - completed university - 60 %;	**1.** Heard about cervical cancer - 56.9 %, **2.** Heard about HPV vaccine - 56.9 %, **3.** Heard about cytological examination - 37.5 %;	[30]

(*continued on next page*)

1. Table 1 (*continued*)

S.No	Place	Year of study	Sample size	Population characterstics	Findings	Reference
22	Cameroon	NA	433	Age - 18 to 68 years; Marital status - 39.3 % married; Literacy status - Tertiary education - 24 %;	**1.** Heard of cervical cancer - 58 %; **2.** Knowledge of risk factors – 58.99 %; **3.** screened for cervical cancer - 28.4 %	[31]
23	Saudi Arabia	2019	1489	Age - 18 to 25 years; Marital status - 53.3 % married; Literacy status - higher level education - 67.4 %;	**1.** Knowledge of symptoms - Vaginal bleeding 79.8 %; **2.** Knowledge of risk factors - Family history 53.5 %; **3.** Heard of the pap smear test - 51.9 %; **4.** Heard of the HPV vaccine - 12.6 %	[32]
24	northern Tanzania	2020	297	Age - 18 to 55 years; Marital status - 29 % married; Literacy status - Secondary education or above - 22.9 %;	**1.** Knowledge of signs and symptoms - adequate 27.9 %; **2.** Knowledge of risk factors - 38.4 %; **3.** Knowledge of prevention strategies - 52.5 %; **4.** Positive attitude towards screening - 66.7 %	[33]
25	India	2020	2100	Age - 17 to 54 years; Marital status - 98.1 % from rural and 98.5 % from urban were married; Literacy status - above 98 % had college level education	**1.** Knowledge about cervical cancer - 63.1 % from rural and 71.3 % from urban; **2.** Knowledge about HPV - 74.5 % from rural and 70.38 % from urban; **3.** Knowledge about Pap test - 18.6 % from rural and 27.23 % from urban; **4.** Knowledge about HPV vaccine - 26.75 % from rural and 19.76 % from urban; **5.** Positive attitude towards cytological screening - 37.8 % from rural and 33.23 % from urban;	[6]

NA = Not applicable

Knowledge about symptoms

Studies on cervical cancer awareness revealed alarming gaps in knowledge regarding symptoms. In South Africa, a staggering 58.8 % of participants believed vaginal discharge and/or bleeding did not warrant cervical cancer screening, while 13.7 % were uncertain. Similarly, in Ethiopia, only 22 % and 31.6 % of students recognized vaginal bleeding and foul-smelling discharge as symptoms, respectively, while a concerning 45.2 % of Central Ethiopian participants were completely unaware of any symptoms.

In contrast, a study in Saudi Arabia showed more promising results, with 79.8 % of participants correctly identifying vaginal bleeding as a symptom, and 43.7 % and 19.3 % recognizing dyspareunia and leg pain, respectively. A study in Bangladesh identified a range of symptoms, including abdominal heaviness, foul-smelling discharge, weight loss, irregular bleeding, and postmenopausal bleeding.

These findings highlight the urgent need for improved education on cervical cancer symptoms to facilitate early detection and reduce disease burden. The lack of awareness about symptoms may delay diagnosis, and women unaware of symptoms are less likely to seek medical attention. Enhancing knowledge of cervical cancer signs and causes is essential for effective prevention and control strategies.

Overall, these studies underscore the importance of addressing knowledge gaps and promoting awareness about cervical cancer symptoms to save lives and improve health outcomes.

Knowledge about risk factors

Cervical cancer is primarily caused by Human Papillomavirus (HPV), with additional risk factors including long-term oral contraceptives, immune-suppressive illnesses, and multiple pregnancies. However, studies reveal alarming gaps in awareness regarding these risk factors. In Bangladesh, urban women were more aware of risk factors than rural women, identifying early sexual activity, multiple partners, and long-term oral contraceptive use. Conversely, HPV and lack of Pap screening tests were the least recognized risk factors.

In Central Ethiopia, 49 % of participants were unaware of cervical cancer risk factors. Other studies in Bangladesh and Saudi Arabia reported similarly low awareness, with only 13 % and 25.72 % recognizing HPV as a risk factor, respectively. A South African study found merely 18.8 % of participants correctly identified risk factors, with 8.6 % recognizing STIs, 4.6 % recognizing HPV, and 4 % recognizing multiple partners.

These findings underscore the significance of education on cervical cancer risk factors, particularly HPV, to facilitate prevention and early detection. The lack of awareness about risk factors contributes to rising cervical cancer incidence, emphasizing the need for targeted interventions to address knowledge gaps.

Knowledge about prevention strategy

The broad implementation of cytological screening programs makes cervical cancer easily avoidable. Compared to women who had never been examined before the study, those who had previously used Pap screening services were more likely to be knowledgeable about HPV. The reason these women knew more may have been revealed during the screening process. There was a link between not getting screened and not being aware of the cervical cancer risk factors .[34]

Research highlights significant gaps in cervical cancer screening awareness and practices. Women who underwent Pap smear screening were more likely to be aware of HPV, yet only 47.1 % of participants had been screened, mostly within the past five years. Alarmingly, low screening rates persist among healthcare professionals, with merely 10 % of doctors and less than 1 % of nurses having undergone screening.

In Bangladesh and India, limited knowledge about Pap smear tests prevails, with only 8.7 % and 7 % of women recognizing it as a cervical cancer screening method. Ethiopian studies similarly reported widespread unawareness of Pap smear screening. Furthermore, healthcare professionals demonstrated poor attitudes toward screening, with 90 % never referring patients.

Despite these findings, most participants recognized prevention methods, such as stopping smoking, avoiding multiple partners, and early sexual activity. Notably, early detection and HPV vaccination can prevent 67 % of cervical cancer cases. However, unfavourable attitudes toward screening persist, with 55.5 % of respondents exhibiting negative perceptions.

Multiple barriers contribute to low screening rates, including low education and awareness, low perceived risk, delayed symptoms, social stigma, cancer fear, cost, familial commitments, and humiliation.

To address these challenges, targeted education and awareness campaigns are crucial to improve cervical cancer screening rates and prevent unnecessary deaths.

Conclusion

The majority of studies included in this review indicate that a significant gap exists between women's awareness of cervical cancer and their actual screening practices, despite favourable attitudes. To effectively prevent and control cervical cancer, enhancing women's knowledge, fostering positive attitudes, and improving screening practices are crucial. This study highlighted diverse findings from various countries, underscoring the need for tailored strategies to bridge the knowledge-practice gap and optimize cervical cancer prevention, ultimately saving lives through education, awareness, and targeted interventions.

Strength and limitations

It will be useful as a guide for policymakers in designing educational programmes on cervical cancer screening and prevention to raise awareness in women and enhance screening uptake, resulting in a decrease in the cervical cancer burden.

Declaration

Consent to participate

Not applicable.

Consent for publication

Not applicable

Data availability statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Funding

Not applicable

Ethical approval and consent to participate

There was no need for ethics clearance because the review was based on previously published studies.

CRediT authorship contribution statement

Balraj Sudha: Conceptualization, Data curation, Formal analysis, Writing – original draft. **Nachimuthu Senthil Kumar:** Formal analysis, Validation. **Sundaravadivelu Sumathi:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

The authors thank the department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249. MayEpub 2021 Feb 4. PMID: 33538338. https://doi.org/10.3322/caac.21660.
2. Zhang S., Xu H., Zhang L., et al. Cervical cancer: epidemiology, risk factors and screening. *Chinese Journal of Cancer Research*. 2020;32(6):720.
3. Onuki M., Matsumoto K., Satoh T., et al. Human papillomavirus infections among Japanese women: age-related prevalence and type-specific risk for cervical cancer. *Cancer Sci*. 2009;100(7):1312–1316.
4. Konno R., Shin H.R., Kim Y.T., et al. Human papillomavirus infection and cervical cancer prevention in Japan and Korea. *Vaccine*. 2008;26:M30–M42.
5. Harper DM, Plegue M, Harmes KM, et al. Three large scale surveys highlight the complexity of cervical cancer under-screening among women 45–65 years of age in the United States. *Prev Med*. 2020;130:105880. May 2019. https://doi.org/10.1016/j.ypmed.2019.105880.
6. Sudha B., Kumar N.S., Sumathi S. Absence of knowledge and awareness about cervical cancer among educated women: a need for education about cervical cancer. *Indian J Gynecol Oncol*. 2022;20(1):11.
7. Tiiti TA, Bogers J, Lebelo RL. Knowledge of human papillomavirus and cervical cancer among women attending gynecology clinics in Pretoria, South Africa. *Int J Environ Res Public Health*. 2022;19(7):4210. Apr 1.
8. Montgomery MP, Dune T, Shetty PK, et al. Knowledge and acceptability of human papillomavirus vaccination and cervical cancer screening among women in Karnataka, India. *Journal of Cancer Education*. 2015;30:130–137. Mar.
9.
10. Alam NE, Islam M.S, Rayyan F, et al. Lack of knowledge is the leading key for the growing cervical cancer incidents in Bangladesh: A population based, cross-sectional study. *PLOSGlob Public Health*. 2022;2(1):e0000149. https://doi.org/10.1371/journal.pgph.0000149.
11.
12. Tadesse A., Tafa Segni M., Demissie H.F. Knowledge, attitude, and practice (KAP) toward cervical cancer screening among adama science and technology university female students. *Ethiopia. International Journal of Breast Cancer*. 2022;2490327.
13. Rahman H, Kar S. Knowledge, attitudes and practice toward cervical cancer screening among Sikkimese nursing staff in India. *Indian J Med Paediatr Oncol*. 2015;36:105–110.
14. Patra S, Upadhyay M, Chhabra P. Awareness of cervical cancer and willingness to participate in screening program: public health policy implications. *J Cancer Res Ther*. 2017;13:318–323. PMID: 28643754. https://doi.org/10.4103/0973-1482.187279.
15. Islam RM, Bell RJ, Billah B, et al. Lack of understanding of cervical cancer and screening is the leading barrier to screening uptake in women at midlife in Bangladesh: population based cross-sectional survey. *Oncologist*. 2015;20:1386–1392. PMID: 26590177. https://doi.org/10.1634/theoncologist.2015-0235.
16. Thippeveeranna C, Mohan SS, Singh LR, et al. Knowledge, attitude and practice of the pap smear as a screening procedure among nurses in a tertiary hospital in north eastern India. *Asian Pac J Cancer Prev*. 2013;14:849–852.
17. Wright KO, Aiyedehin O, Akinyinka MR, et al. Cervical cancer: community perception and preventive practices in an urban neighborhood of Lagos (Nigeria). *Int Sch Res Notices*. 2014;950534.
18. Gebisa T, Bala ET, Deriba BS. Knowledge, attitude, and practice toward cervical cancer screening among women attending health facilities in Central Ethiopia. *Cancer Control*. 2022;29:10732748221076680. Feb 17.
19. Chali K, Oljira D, Sileshi T, et al. Knowledge on cervical cancer, attitude toward its screening, and associated factors among reproductive age women in Metu Town, Ilu Aba Bor, South West Ethiopia, 2018: community-based cross-sectional study. *Cancer Rep*. 2021;4(5):e1382. Oct.
20.
21. Qayum MO, Billah MM, Akhter R, et al. Women's knowledge, attitude and practice on cervical cancer and its screening in Dhaka. *Bangladesh. Asian Pacific Journal of Cancer Prevention: APJCP* 2021;22(10):3327. Oct.
22. Rimande-Joel R, Ekenedo GO. Knowledge, belief and practice of cervical cancer screening and prevention among women of Taraba, North-East Nigeria. Asian Pacific journal of cancer prevention. *APJCP*. 2019;20(11):3291.
23.
24. Gyamfua AA, Nkrumah I, Ibitoye BM, et al. The level of knowledge and associated socio-demographic factors on cervical cancer among women: a cross-sectional study at Kenyase Bosore community, Ghana. *Pan African Med J*. 2019;34(1). Sep 24.
25. Garg PR, Srivastava S, Shumayla S, et al. Women's knowledge on cervical cancer risk factors and symptoms: a cross sectional study from Urban India. *Asian Pacific J Cancer Prevention*. 2022;23(3):1083–1090. Mar 1.
26. Ghosh S, Mallya SD, Shetty RS, et al. Knowledge, attitude and practices towards cervical cancer and its screening among women from tribal population: A community-based study from southern India. *J Racial and Ethnic Health Disparities* 2021;8:88–93. Feb.
27. Tilahun T, Tulu T, Dechasa W. Knowledge, attitude and practice of cervical cancer screening and associated factors amongst female students at Wollega University, western Ethiopia. *BMC Res Notes*. 2019;12(1):1–5. Dec.
28. Banik Rajon., Naher S., Rahman M., et al. Investigating Bangladeshi rural women's awareness and knowledge of cervical cancer and attitude towards HPV vaccination: a community-based cross-sectional analysis. *J Cancer Educat*. 2020;1–12.
29. Badawy NA, Al Shammari MA, Hosseini MS, et al. Knowledge and attitude related to cervical cancer prevention and screening among female students in the College of Nursing. *Kuwait. Saudi J Nurs Health Care* 2022;5(3):47–57.
30. Williams MS, Kenu E, Adanu A, et al. Awareness and beliefs about cervical cancer, the HPV vaccine, and cervical cancer screening among Ghanaian women with diverse education levels. *J Cancer Educat*. 2019;34:897–903. Oct.
31. Nkfusai NC, Cumber SN, Anchang-Kimbi JK, et al. Assessment of the current state of knowledge and risk factors of cervical cancer among women in the Buea Health District, Cameroon. *Pan Afr Med J*. 1-9. 2019;33.
32. Zahid HM, Qarah AB, Alharbi AM, et al. Awareness and practices related to cervical cancer among females in Saudi Arabia. *Int J Environ Res Public Health*. 2022;19(3):1455. Jan 27.
33. Mwantake MR, Kajoka HD, Kimondo FC, et al. Factors associated with cervical cancer screening among women living with HIV in the Kilimanjaro region, northern Tanzania: A cross-sectional study. *Prev Med Rep*. 2022;30:101985. Dec 1.
34. Tiiti TA, Bogers J, Lebelo RL. Knowledge of human papillomavirus and cervical cancer among women attending gynecology clinics in Pretoria, South Africa. *Int J Environ Res Public Health*. 2022;19(7):4210. Apr 1.

33 (2022) 201048

Contents lists available at ScienceDirect

Human Gene

journal homepage: www.journals.elsevier.com/human-gene

Identification of hub genes and role of *CDKN2A* as a biomarker in cervical cancer: An *in-silico* approach

Balraj Sudha [a], Arumugam Poornima [a], Kanagaraj Suganya [a], Kandasamy Swathi [a], Nachimuthu Senthil Kumar [b], Sundaravadivelu Sumathi [a,*], Paulchamy Chellapandi [c]

[a] *Department of Biochemistry, Biotechnology, and Bioinformatics, Avinashilingam Institute for Home Science and Higher Education for women, Coimbatore 641 043, Tamil Nadu, India*
[b] *Department of Biotechnology, Mizoram University, Aizawl 796 004, Mizoram, India*
[c] *Molecular Systems Engineering Lab, Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli 620024, Tamil Nadu, India*

ARTICLE INFO

Keywords:
Cervical Cancer
CDKN2A
Gene expression
Bioinformatics
differentially expressed genes

ABSTRACT

Cervical cancer is one of the most common gynecological cancers in women and its molecular pathogenesis and disease progression are yet fully understood. Finding new biomarkers is important to detect early diagnosis of cervical cancer to reduce its incidence and mortality rates among women. We have selected 3 microarray gene expression datasets (GSE67522, GSE138080, and GSE75132) and then analysed up-regulated and down-regulated genes in cervical cancer. The hub genes related to this disease were identified from constructed protein-protein interaction network. The statistical significance of key genes was validated with experimental data obtained from TCGA cervical cancer patients. We identified 18 differentially expressed genes from 3 microarray datasets. These genes were highly associated with DNA replication and cell proliferation pathways. Network analysis revealed that *CDKN2A* as a biomarker for cervical cancer prognosis. The expression and interactions of this gene were analysed with bioinformatic tools. Results of this study showed that *CDKN2A* has significant interactions with transcription factors, signalling molecules, and miRNAs. *In-silico* analysis of microarray data can pave the way to predict *CDKN2A* as a gene target for the diagnosis of cervical cancer. Early diagnosis of this disease would decrease morbidity and mortality among females worldwide.

1. Introduction

Cervical carcinoma is the fourth most frequent malignancy in females and the primary cause of cancer death. In 2020, more than 600,000 new cases and 342,000 deaths have been reported. Cervical cancer is the most common cancer in 23 countries and the foremost reason for cancer death in 36 countries (Stanca and Căpîlna, 2021). According to histology, cervical cancer is classified as squamous cell carcinoma, adenocarcinoma, or adeno-squamous carcinoma (Gurram et al., 2020). Occurrence and death rates have been dropped in most parts of the world over the last few decades. Decreases in the death rate are related to causes associated either with a rise in the overall socio-economic status or a reduction in the risk of chronic infection with high-risk human papillomavirus (HPV) (Sung et al., 2021). Although the disease's prognosis has been improved because of a combination of screening and surgery, it is still the fourth-highest rate of morbidity and mortality among females on the globe (Bray et al., 2018).

Traditional treatment for cervical cancer includes chemotherapy, radiotherapy, surgery, and adjuvant therapies (Cohen et al., 2019). Since surgery is only used to treat cervical carcinoma patients in the premature stages (IA-IIA), radiotherapy or concomitant chemo-radiotherapy is the mainstay of treatment for nearby progressed (IIB-IVA) or recurrent patients. Drastic hysterectomy and radical radiation, as well as concomitant cisplatin chemotherapy, are the most common treatments for cervical cancer patients. Some patients experience a relapse following surgery or radiotherapy. Relapse patients have minimal therapeutic options and a dismal prognosis (Dizon et al., 2014). The treatment effectiveness of cervical cancer remains inadequate to date. About 30% of newly diagnosed patients developed relapse and metastasis within two years, and the 5-year survival rate was less than 10% (Li et al., 2018). According to the results of the Gynaecologic Oncology Group 240 trial, cisplatin and paclitaxel-based chemotherapy, as well as tumor vascular-targeting drugs like bevacizumab (targeting vascular endothelial growth factor, VEGF), have been accepted by the U.S. Food

* Corresponding author.
E-mail address: sumathi_bc@avinuty.ac.in (S. Sumathi).

https://doi.org/10.1016/j.humgen.2022.201048
Received 12 January 2022; Received in revised form 28 February 2022; Accepted 27 April 2022
Available online 10 May 2022

and Drug Administration as the first-line therapy for pioneering treatment with recurrence or metastasis (Tewari and Monk, 2014). However, cisplatin-based chemotherapy often leads to drug resistance, and newer treatments such as VEGF targeted therapy and immunotherapy have had mixed results. Therefore, new approaches are dealt to increase the effectiveness of cervical cancer treatment are urgently needed.

Oncologists first determine the prognosis of cervical cancer patients to increase treatment effectiveness and establish precise treatment regimens. HPV is a significant predictor intended for cervical carcinoma but not a sufficient carcinogenic state (Gheit, 2019). In addition, HPV-16 and HPV-18 are the primary causes of cervical epithelial cell cancer (Shibata et al., 2019). HPV-16 is connected to approximately 80% of squamous cell cervical carcinoma whereas HPV-18 is linked to 20% of adeno-cervical carcinoma (Anderson et al., 2001). Molecular and functional investigations demonstrated a relationship between cervical cancer development and chronic high-risk human papillomavirus (HR-HPV) infection. However, the precise molecular network pathways from HPV infection to carcinogenesis remain unknown. Therefore, the investigation of the possible pathways causing tumorigenesis is important to extending survival outcomes (Wu et al., 2018).

Researchers can classify cancerous and non-cancerous tissues based on gene expression profiles from the microarray. Differentially expressed genes (DEGs) in cervical carcinoma have been identified using gene expression data. Using GEO2R, STRING database and Cytoscape software, Xue et al. discovered and validated four hub genes *CDC45*, *GINS2*, *MCM2*, and *PCNA* that were associated with the prognosis of cervical cancer patients (Xue et al., 2021). *TOP2A* may be used to predict poor prognosis and as a therapeutic target for cervical cancer treatment (Zhao et al., 2020). Genes including *ATG3*, *AIM2*, *ABCB7*, *ARHGAP6*, *AP2B*, *CNBP*, *HIST1H3C*, *IGF1*, *PKN3, THOC*, *TRIM66,* and *TP63* were identified as DEGs analysed by WebGestalt, STRING, Genemania, NetworkAnalyst software and UALCAN software in cervical cancer (Rajput et al., 2020). The public microarray datasets for cervical cancer are yet to be completely examined, and future research of these datasets may help us find biomarkers linked with the disease.

The *CDKN2A* gene covers p16 protein and plays a central role in cellular differentiation, senescence, and cell death. It also regulates proliferation and the cell cycle in tumorigenesis (Jiao et al., 2018). Overexpression of *CDKN2A* stimulates the synthesis of p16-INK4a and causes cell cycle arrest in the G1 phase (Ohtani et al., 2001). *CDKN2A* is the main cytokine signal transduction regulator that develops and predicts cancer (Pal et al., 2016; Hatzistergos et al., 2019). Basic and clinical results have shown that *CDKN2A* expression was low in ovarian cancer patients (Bowe et al., 2019). Silence of *CDKN2A*'s promotes gynecological cancer (Hosseini et al., 2017). In addition, *CDKN2A* overexpression increases the apoptosis rate and ultimately hinders the proliferation of breast cancer cells (Aftab et al., 2019). Luan et al. 2021 reported that in cervical cancer cell lines, *CDKN2A* expression was low. It also reported that *CDKN2A* expressed differentially in cervical carcinoma patients than in healthy individuals (Zhao et al., 2018) but its molecular mechanism in cervical cancer is poorly understood (Luan et al., 2021). *CDKN2A* can be used as a biomarker for cervical cancer risk. Studies have reported the role of *CDKN2A* in cervical cancer tumor samples (Luan et al., 2021; Mizuarai et al., 2011), comparison of its expression from microarray datasets and connections with other biological molecules were performed not many studies have been reported.

In this study, to obtain DEGs (Differentially expressed genes) between cervical cancer and healthy cervical tissues, we analysed GSE67522, GSE138080, and GSE75132 mRNA microarray datasets which have not been studied together in detail before. From that, 18 differentially expressed gene signatures in cervical cancer were identified using a network modelling approach. The cyclin dependant kinase inhibitor 2A (*CDKN2A*) gene was predicted as an independent predictive indicator of disease-free survival in cervical cancer. Hence, we attempted to study the role of *CDKN2A* in cervical carcinoma to investigate it as a valuable biomarker for the early diagnosis of cervical cancer patients.

2. Materials and methods

2.1. Microarray data

Totally one hundred and sixty-seven gene expression profiles related to cervical cancer were available in the GEO dataset under the expression profiling array category from 2003 till 2022. Among the various datasets available, the gene expression files of GSE67522, GSE138080, and GSE75132 datasets were chosen from the GEO database which were not taken together for comparison in any study reported so far (Supplementary Table 1). The GSE67522 dataset included twenty-two samples of healthy individuals and twenty cervical carcinoma samples, and the platform employed was the GPL10558 Illumina HumanHT-12 V4.0 expression bead chip (Saha et al., 2017). The GSE138080 dataset utilized GPL4133 Agilent-014850 Whole Human Genome Microarray 4 × 44 K G4112F platforms and comprised ten healthy cervical samples and ten cervical cancer samples (Babion et al., 2020). The GSE75132 dataset comprised eleven healthy cervix samples and one cervical carcinoma sample, and the platform employed was the GPL570 Human Genome U133 Plus 2.0 Array (Manawapat-Klopfer et al., 2016).

2.2. Analysis of differentially expressed genes

The differentially expressed genes in cervical cancer and healthy cervical samples were separated using GEO2R (ncbi.nlm.nih.gov/geo/geo2r). LIMMA is used to normalize the value and provides several *P*-value adjustment options. These adjustments are also called multiple testing corrections which attempt to correct for the occurrence of false-positive results. The adjusted *P*-values (adj.P) and Benjamini-Hochberg false discovery rates were utilized to achieve a compromise between discovering statistically important genes and reducing false positives. Only DEGs with a log fold-change of more than 2 and an adj.P-value of less than 0.05 was considered statistically significant. DAVID (http://david.abcc.ncifcrf.gov/) is a biological data and analytical database that may be used to find GO terms, convert gene IDs, and perform gene function enrichment analysis etc. Gene Ontology (GO) enrichment analysis for selected DEGs was performed using the DAVID bioinformatics software. It defines the assignment of genes and gene products in organisms, including molecular function (MF), biological process (BP), and cellular components (CC). For GO analysis, the background species were selected as *Homo sapiens* in DAVID. Background genes were considered against DEGs. Fisher's exact test was used in DAVID to measure gene enrichment in annotation terms. The significance level was established at $P < 0.05$ (Jiao et al., 2012).

2.3. Identification of hub genes

The Protein-protein interaction (PPI) networks of discovered DEGs were constructed using a network analyst integrating the IMEx interactome. InnateDB, a member of the International Molecular Exchange (IMEx) project, provides Network Analyst with a huge, high-quality PPI database. The database was created by manually merging experimental data from several PPI databases, together with BIND, BioGRID, DIP, IntAct, and MINT with data from the literature. Human interactions represented in 14,755 proteins and 145,955 were analytically authenticated interactions in the database (Xia et al., 2014).

2.4. In silico validation of DEGs with cancer data

The UALCAN is one of the considerable online sources for cancer data analysis and mining (http://ualcan.path.uab.edu/analysis.html). It was used to select the key genes involved in cancer pathways (Chandrashekar et al., 2017). We measured DEGs expression level in CESC from TCGA samples based on the individual cancer stages. Statistical significance was defined as a *P*-value of less than 0.05. The clinical consequences of identified common DEGs were validated in a large

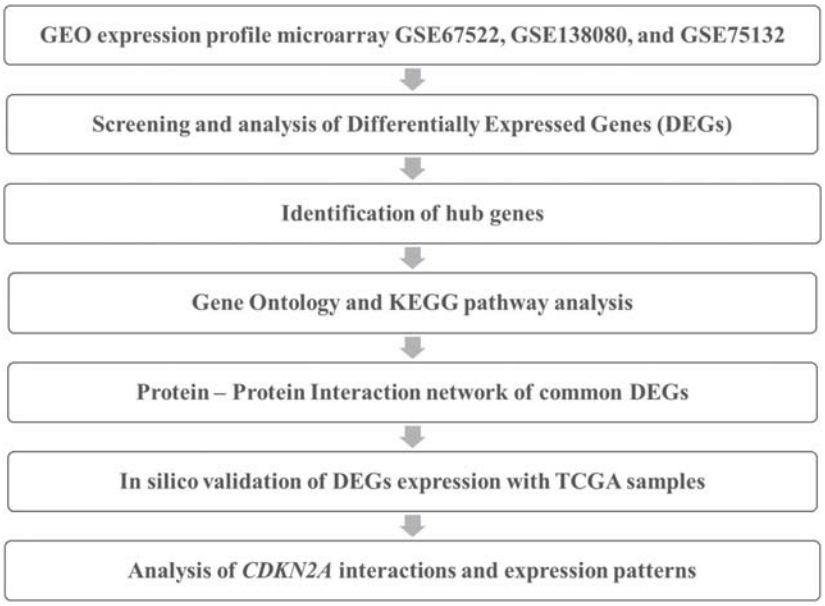


Fig. 1. Workflow of the bioinformatic analysis and *in-silico* validation.

cohort of cervical cancer samples using cBio-Portal from TCGA (Gao et al., 2013). Using the GEPIA2 tool, the expression levels of interesting genes in all type of cancers was measured with differential method (LIMMA) (Tang et al., 2019). Expression of selected genes was analysed which are specific to cervical cancer using GEPIA2. The set Log2FC value was 2.

2.5. Analysis of CDKN2A interactions and expression patterns

The transcriptional factor, signalling molecule, and miRNA network connected with *CDKN2A* were analysed using the Network Analyst database to explore the regulatory mechanism at the transcriptional level (Zhou et al., 2019). *CDKN2A* and co-expression genes were analysed for correlation using cBioPortal (Gao et al., 2013). In GEPIA2, *CDKN2A* expressions in tumour samples from TCGA were compared to combined expression data of normal adjacent sample and normal cervix sample in Genotype-Tissue Expression (GTEx) (Tang et al., 2019).

2.6. Statistical analysis

The mandatory statistical analysis was done for all the outputs obtained. Using Benjamin and Hochberg's false discovery rate, P-value was corrected for multiple comparisons. The significance of Spearman's Correlation coefficient between the expressions of mRNA z-Scores was determined using a log-rank test. A Spearman's Correlation coefficient of more than 0.5 and a P-value of less than 0.05 were considered statistically significant.

3. Results

3.1. Study characteristics

The workflow of our bioinformatics analysis and in-silico validation was shown in Fig. 1. In this investigation, 74 patient tissue samples were taken from all three datasets with full-text articles. The sufferers were

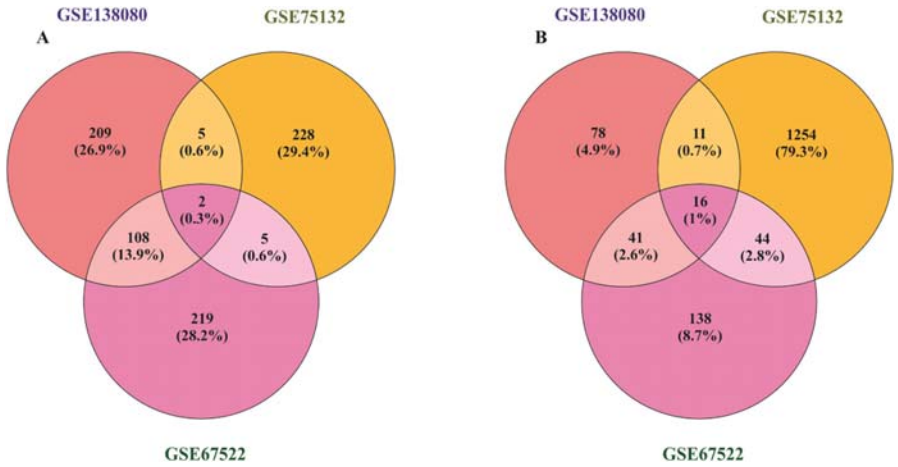


Fig. 2. Venn diagram depicting the DEGs in the selected microarray datasets using GEO2R tool. Up-regulated genes(A), Down-regulated genes (B).

Table 1
Functional enrichment of identified DEGs.

ID	Category Term	P-value	Genes
KEGG pathway			
hsa03030	DNA replication	0.041128	RFC4, MCM6
hsa03460	Fanconia anemia pathway	0.05	FANCI, RMI2
Molecular Function			
GO:0003677	DNA binding	0.004453	FANCI, RMI2, RFC4, CDKN2A, HLTF, MCM6, E2F7
GO:0005515	Protein binding	0.009976	FANCI, RFC4, CDKN2A, UHRF1, CRCT1, CKS1B, CENPF, KIAA0101, HLTF, STMN1, KIF2C, TK1, MCM6, ECT2, E2F7
GO:0042802	Identical protein binding	0.037175	UHRF1, TK1, MCM6, E2F7
GO:0005524	ATP binding	0.057366	RFC4, HLTF, KIF2C, TK1, MCM6
Biological Process			
GO:0006260	DNA replication	0.000477	RMI2, RFC4, KIAA0101, MCM6
GO:0008283	Cell proliferation	0.005565	CENPF, UHRF1, KIF2C, CKS1B
GO:0010389	Regulation of G2/M transition of mitotic cell cycle	0.009077	CENPF, CDKN2A
GO:0007019	Microtubule depolymerization	0.010081	STMN1, KIF2C
GO:0051310	Metaphase plate congression	0.012085	CENPF, KIF2C
Cellular Component			
GO:0005654	Nucleoplasm	0.000337	FANCI, CENPF, RMI2, RFC4, KIAA0101, CDKN2A, HLTF, MCM6, E2F7, CKS1B
GO:0030496	Midbody	0.006307	ASPM, CENPF, ECT2
GO:0005634	Nucleus	0.011781	ASPM, CENPF, RMI2, KIAA0101, CDKN2A, UHRF1, HLTF, KIF2C, MCM6, ECT2, E2F7
GO:0005737	Cytoplasm	0.030178	FANCI, ASPM, CENPF, RMI2, KIAA0101, CDKN2A, CRCT1, HLTF, STMN1, ECT2
GO:0000775	Chromosome, centromeric region	0.051884	CENPF, KIF2C

not given any immunological or pharmacological treatments before biopsy in any of the studies, and samples were obtained before any medication was given.

3.2. Identification of DEGs

In the GSE67522 dataset, 334 were up-regulated genes and 239 genes were down-regulated. In the GSE138080 dataset, 324 genes were up-regulated, and 146 genes were down-regulated. In the GSE75132 dataset, 240 genes were up-regulated, and 1325 genes were down-regulated. Among the eighteen genes identified as differentially expressed genes (DEGs), two and sixteen overlapped in the up-regulated (Fig. 2A) and down-regulated (Fig. 2B) genes, respectively in all the three data sets. Volcano plots were created to visualise the distribution of differentially expressed genes between the samples (Normal and CC) within each variety and between varieties for all the three sets (Supplementary Fig. 1).

3.3. Functional enrichment of DEGs

As shown in Table 1, the top-most significantly supplemented DEGs in the molecular functions were DNA binding ($P = 0.00445309$), protein binding ($P = 0.009976334$), identical protein binding ($P = 0.037175191$), and ATP binding ($P = 0.05736563$); the biological processes are DNA replication ($P = 4.77E-04$), cell proliferation ($P = 0.005564771$), G2/M transition of mitotic cell cycle regulation ($P = 0.009076825$), depolymerization of microtubule ($P = 0.010080561$) and metaphase plate congression ($P = 0.012085162$); the cellular components are nucleoplasm ($P = 3.37E-04$), midbody ($P = 0.006307423$), nucleus ($P = 0.011780609$), cytoplasm ($P = 0.030178209$) and chromosome, centromeric region ($P = 0.051883978$).

3.4. Identification of hub genes

We constructed first-order PPI network possessed 656 nodes (proteins) and 777 (interactions) with eighteen DEGs (Fig. 3). *TK1* (degree centrality = 159; betweenness centrality = 86,427.18), *CDKN2A* (degree centrality = 148; betweenness centrality = 73,699.53), *ECT2* (degree centrality = 145; betweenness centrality = 78,223.25), MCM6 (degree centrality = 76; betweenness centrality = 38,805.66) and *RFC4* (degree centrality = 57; betweenness centrality = 27,004.37) were the top-ranking nodes in the combined datasets derived from network topology measures. These proteins are associated with cervical cancer pathways including cell cycle, ubiquitin mediated proteolysis, cellular senescence, p53 signalling pathway and DNA replication.

3.5. In silico validation of DEGs in cervical cancer

Using UALCAN, we generated an interactive heatmap of eighteen differentially expressed genes in normal and cancer (Fig. 4A). Based on TCGA data and clinical patient data, UALCAN is utilised to investigate gene expression levels. It is used to compare clinical pathological parameters of patients based on pathological staging, tumour grade, and other clinical pathology features, as well as primary tumours and healthy tissue samples. The sample size of the groups namely normal, stage 1, stage 2, stage 3, and stage 4 are 3, 161, 69, 46, and 22, respectively. Here, we evaluated the expression of selected gene signatures in normal, stage 1, stage 2, stage 3, and stage 4 (Fig. 4B). Significant expression was noted for *ASPM, CI60RF75, CDKN2A, CENPF, CKS1B, E2F7, ECT2, FANCI, KIAA0101, KIF2C, MCM6, RFC4, STMN1, TK1 and UHRF1* which was greater in cervical carcinoma based on individual cancer stages than normal samples ($P < 0.05$). No significant expression was noted for *GPAT3, CRCT1* and *HLTF*.

As shown in Fig. 4C, the same kind of results were obtained while using GEPIA2. *ASPM, CI60RF75, CDKN2A, CENPF, CKS1B, E2F7, ECT2, FANCI, KIAA0101, KIF2C, MCM6, RFC4, STMN1, TK1 and UHRF* genes were highly expressed in cervical cancer compared to normal individuals. Here again, no significant expression was observed for *GPAT3, CRCT1* and *HLTF*.

3.6. CDKN2A interactions and expressions

Amongst, *CDKN2A* is identified to be involved in the p53 pathway and cell cycle prognosis. This gene was selected for further analysis (Fig. 5A). *CDKN2A* is a centre point for signalling molecule regulators namely, IKBKB, POMC, PRC2, TBX5, UV stress, NDN, BMI1, ASXL1, MYC, CTBP1, DNMT3A, and SWI/SNF Complex. In that, *CDKN2A* down-regulates the following signalling regulators: CDK2, CDK6/CCND1, CDK6, and CDK4, while UV stress and NDN up-regulates the *CDKN2A* (Fig. 5B). Transcription factors bind to certain DNA sequences and control how DNA is translated into mRNA. There are 11 transcription factors involved in the expression of *CDKN2A*. We have predicted and analysed the transcription factors targeting *CDKN2A* which are ZNF76,

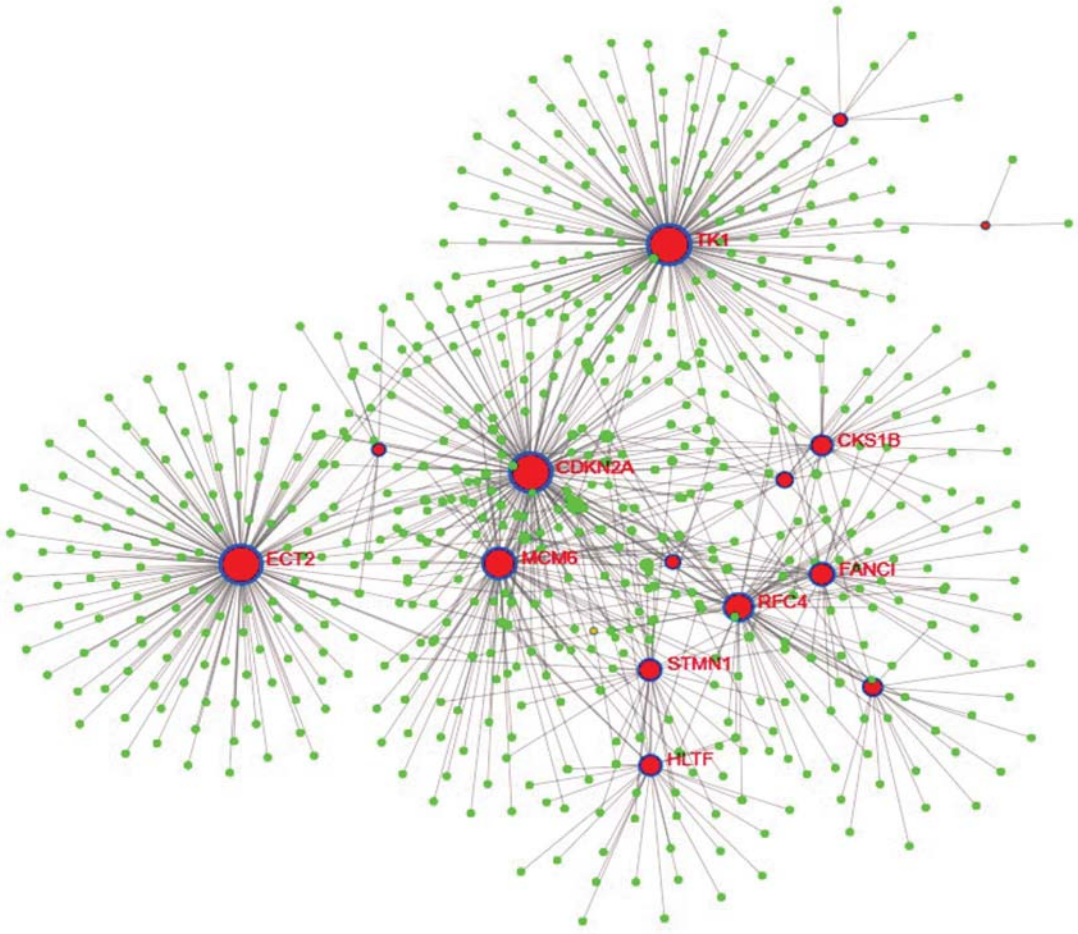


Fig. 3. Construction of protein-protein interaction network for identification of hub genes.

L3MBTL2, PRDM1, CTBP2, ZNF71, SUZ12, ZNF2, ZFP37, CTCF, EZH2, and ZNF423. An endeavour to discover the related miRNA with *CDKN2A* shows that 16 miRNAs are linked with the *CDKN2A* expression.

Correlation analysis of *CDKN2A* gene expression shows the selected genes, kinesin family member 2C (*KIF2C*, $P < 0.0001$, rs = 0.25), ubiquitin like with phd and ring finger domains 1 (*UHRF1*, $P < 0.0001$, rs = 0.09), *CDC28* protein kinase regulatory subunit 1b (*CKS1B*, ($P < 0.0001$, rs = 0.40), Fanconi anaemia, complementation group I (*FANCI*, $P < 0.0001$, rs = 0.09), minichromosome maintenance complex component 6 (*MCM6*, $P < 0.0001$, rs = 0.13), stathmin 1(*STMN1*, $P < 0.0001$, rs = 0.14), thymidine kinase 1(*TK1*, $P < 0.0001$, rs = 0.39), replication factor C subunit 4 (*RFC4*, $P < 0.0001$, rs = 0.27), cysteine rich C- Terminal 1 (*CRCT1*, $P < 0.0001$, rs = 0.29), E2F transcription factor 7 (*E2F7*, $P < 0.0001$, rs = 0.00) are strongly correlated (Fig. 5C).

4. Discussion

Cervical cancer is the deadliest group of diseases that directly affects the cell lining of the cervix surface that experiences sequential changes. Often, these precancerous cells can become cancerous so that early detection of cervical cell changes and suitable treatment reduces the possibility of cervical cancer progression (Luan et al., 2021). Hence, efforts to identify mechanisms underlying the pathogenesis of cervical cancer progression and its associated genes become important to discover new biomarkers for diagnosis.

A total of three datasets (GSE67522, GSE138080, and GSE75132) were selected for this study. Among these, GSE67522 and GSE138080 have been analysed together (Yuan et al., 2021). Other studies have also mined the GSE67522 (Basic et al., 2021), and GSE138080 datasets (Xue et al., 2021; Yuan et al., 2021; Wu and Xi, 2021; Qu et al., 2021). To the best of our knowledge, no such study was conducted with these three (GSE67522, GSE138080, and GSE75132) dataset together.

In the present study, we identified 18 differentially expressed genes (*E2F7*, ASPM, *CDKN2A*, *RMI2*, *UHRF1*, *KIF2C*, *CKS1B*, *KIAA0101*, *FANCI*, *MCM6*, *STMN1*, *CENPF, TK1*, *HLTF*, *ECT2*, *RFC4*, *CRCT1,* and *GPAT3*). The DEGs significantly enriched in the following molecular functions: cell proliferation, DNA replication, G2/M transition of mitotic cell cycle regulation, depolymerization of a microtubule, and metaphase plate congression. Eighteen differentially expressed genes were connected with proteins involved in different biological functions and also the role of *CDKN2A* in cell cycle progression was confirmed. Cell proliferation and invasion were inhibited by *CDKN2A* overexpression, and the cell cycle was arrested in the G1 phase (Ohtani et al., 2001). Therefore, the DEGs identified have diverse roles in molecular processes related to cervical cancer progression.

So far, no study has been reported about the small biomolecular interactions with *CDKN2A* gene. Among the differentially expressed genes, reported by us in the present study, no reports are available

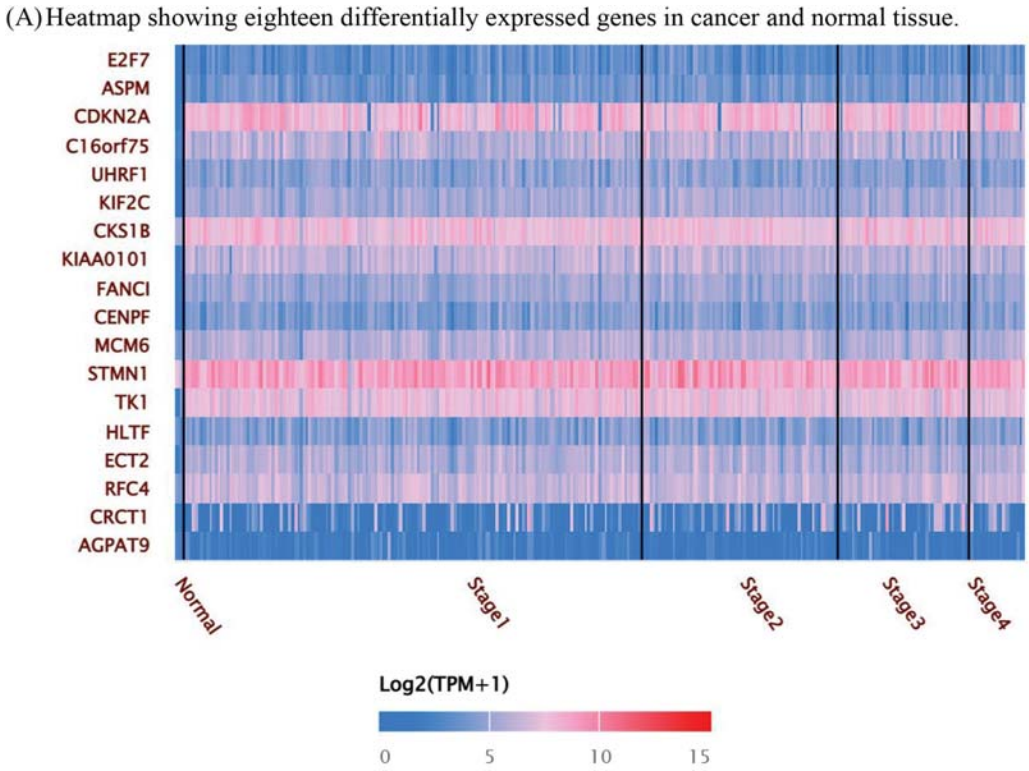
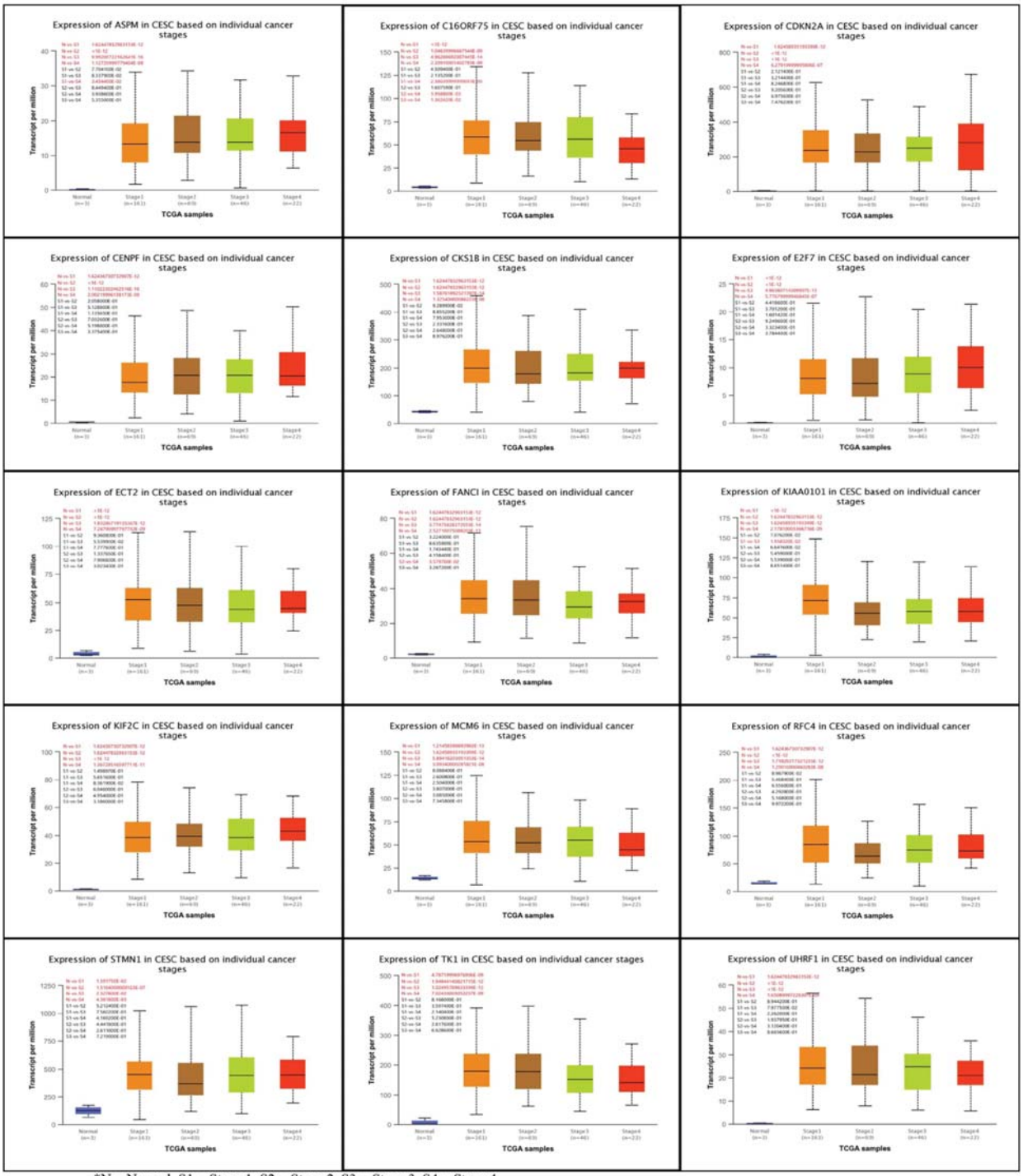


Fig. 4. *In silico* validation of DEGs in the different stages of cervical cancer samples from TCGA dataset.
(A) Heatmap showing eighteen differentially expressed genes in cancer and normal tissue.
The expression pattern is represented in the log2 transform of total transcript per million (TPM) [log2(TPM + 1)]. Each column in the heat-map represents a biological sample of normal, stage 1, stage 2, stage 3 and stage 4, and each row represents a query gene.
(B) Box-whisker plots showing the relative expression of DEGs in different stages of cervical cancer samples.
*N = Normal, S1 = Stage 1, S2 = Stage 2, S3 = Stage 3, S4 = Stage 4.
(C) Relative expression levels of eighteen differential genes in mRNA expression using GEPIA.
Black and red boxes represent the relative expression levels of genes in the normal and tumor samples, respectively. The y-axis represents the relative expression levels of genes in terms of log2 (TPM + 1) (tumor samples, 306; normal samples, 13 from GEPIA; $p \leq 0.01$). TPM, transcripts per million. The status of the expression of fifteen genes were significant. (*GPAT3, CRCT1, HLTF* was not statistically significant). **P*-values<0.05. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

regarding *CDKN2A* correlation analysis and their role in stage wise progression of cervical cancer. Few researchers have recently reported on the clinical importance of *CDKN2A* overexpression in cervical cancer. *CDKN2A* encodes 2 important proteins namely, p16INK4 and p14ARF. The two proteins are effectively involved in the regulatory cell cycle pathways namely RB1 and p53. The role of p16 was to slow down the progression of the cell cycle from the G1 phase to the S phase (Robertson and Jones, 1999). Herein, we examined the signalling molecular regulators, miRNA, and transcription factors associated with *CDKN2A*. Zhao et al., 2018 reported that the *CDKN2A* is an important gene that may be controlled by miR-424-5p or miR-9-5p. Here, using Network analyst platform, we have found that few miRNA regulators were linked with *CDKNDA* namely, has-mir-320a, has-mir-155-5p, has-mir-615-3p, has-mir-492, has-mir-423-5p, has-mir-34a-5p, has-mir-192-5p, has-mir-10b-5p, has-mir-16-5p, has-mir-24-3p, has-mir-215-5p, has-mir-455-3p, has-let-7 g-5p, has-mir-296-3p, has-mir-125b-5p and has-mir-124-3p. The results of our study suggest that *CDKN2A*-associated signalling molecules, transcription factors, and miRNAs play an important role in the molecular mechanism of cervical cancer progression.

The gene expression correlation analysis for *CDKN2A* showed a strong positive correlation with the *UHRF1*, KIF2*C*, *CKS1B*, *FANCI*, *MCM6*, *STMN1*, *TK1*, *RFC4*, *CRCT1,* and *E2F7*. *UHRF1* was an epigenetic regulator involved in the ubiquitination, DNA methylation, and protein methylation process. It was overexpressed in cervical cancer which usually accompanies HPV infection (Zhang et al., 2018). *KIF2C* plays a significant role in cell proliferation and cancer development (Shimo et al., 2007; Abdel-Fatah et al., 2011; Wang et al., 2014; Zhao et al., 2014). *KIF2C* is a direct target of the Wnt/B- catenin pathway and mediates *TORC1* signalling and can be a potential target for hepatocellular carcinoma (Wei et al., 2020). *KIF2C* was observed in most cervical cancers but not found in the normal cervix (van Dam et al., 2018). The miR-181c target gene *CKS1B* was found to be part of a highly enriched pathway in cervical cancer (Mandal et al., 2019). Also, *CKS1B* was shown to be differentially overexpressed in episomal cervical cancer instances, and its expression was found to be strongly and inversely linked with miR-181c expression in the cervical cancer subtype. *CKS1B* participates in cell cycle regulation. High expression of *CKS1B* led to the tumor initiation, maintenance, and progression which could be a potent target in many treating cancers. The molecular mechanism behind *CKS1B* and its interaction was not clear (Shi et al., 2020). *FANCI*, a key component of the FANCD2-FANCI complex, is associated with several steps in the DNA damage process. In cervical cancer, *FANCI* may regulate DNA replication, repair, and cell cycle progression via interacting kinases. According to transcriptome analysis of more than 3400 clinical

(B) Box-whisker plots showing the relative expression of DEGs in different stages of cervical cancer samples.



*N = Normal, S1 = Stage 1, S2 = Stage 2, S3 = Stage 3, S4 = Stage 4

Fig. 4. (*continued*).

(C) Relative expression levels of eighteen differential genes in mRNA expression using GEPIA.

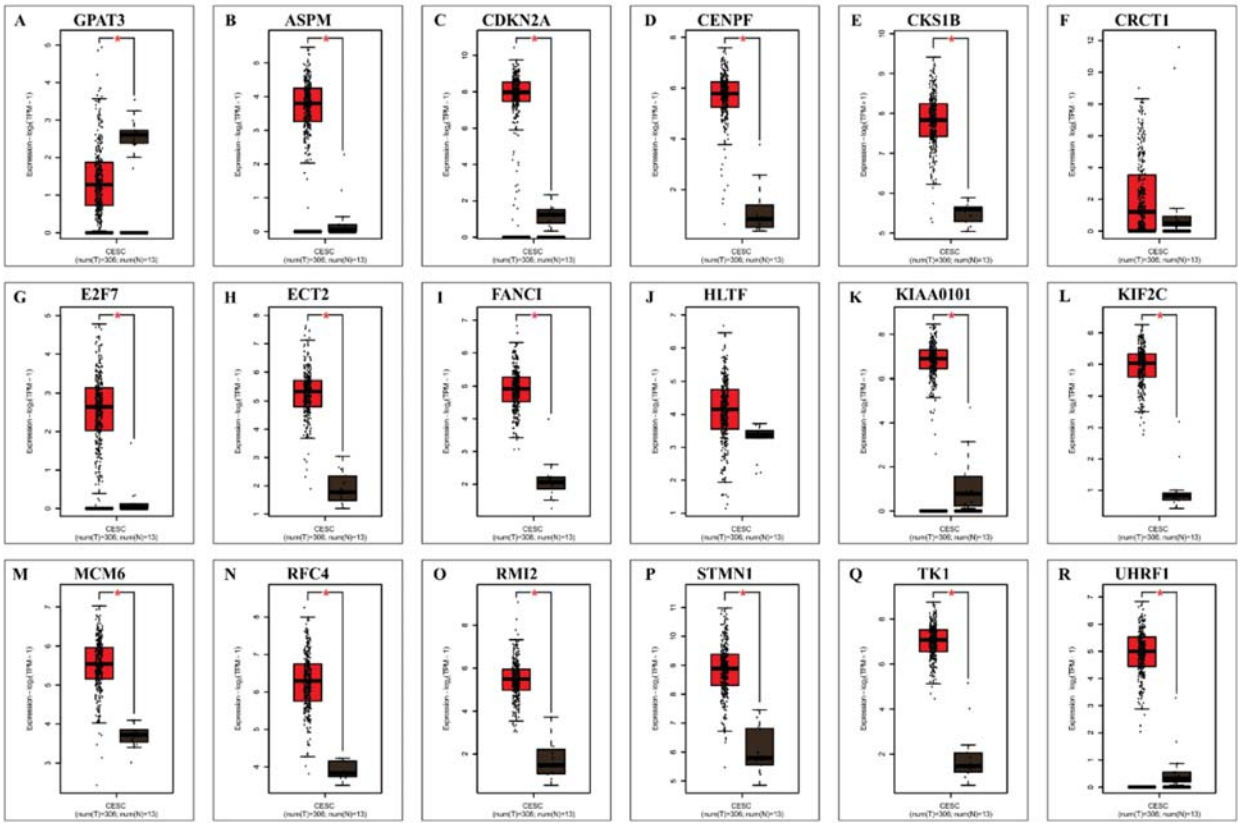


Fig. 4. (*continued*).

samples from six geographic regions and ethnic cervical cancer studies, *FANCI* mRNA levels are considerably greater in cervical cancer than in normal tissue. Furthermore, increased *FANCI* expression was found to enhance short survival in patients with cervical cancer (Liu et al., 2021; Abbas et al., 2019). Minichromosome Maintenance (MCM) proteins are involved in cell cycle progression by mediating DNA replication initiation and elongation. In cervical cancer, *MCM6* was shown to be overexpressed. *MCM6* exhibits an increased frequency of overexpression as tumor stages progress (Das et al., 2013). *STMN1* is also a biomarker in certain types of neoplasm. It serves important functions in cell cycle progression, mitosis, signal transduction, and cell migration. *TK1* has a key function in DNA synthesis and repair (Jagarlamudi and Shaw, 2018). There are not many studies reporting directly on the effect of *STMN1*, *CRCT1*, and *TK1* in cervical cancer. Downregulated *STMN1* was significantly associated with poor overall survival in patients with cervical cancer (Wang and Chen, 2018). Increased expression of *TK1* is found in solid malignancies, and it has been associated with a poor outcome (Qiu et al., 2020). The *RFC4* gene has been linked to cervical cancer (Narayan et al., 2007). Niu et al. demonstrated that *RFC4* dysregulation may contribute to cervical cancer progression and may be potential diagnostic markers (Niu et al., 2017). *E2F7* was found to be upregulated in cervical cancer samples and to be involved in DNA replication. The upregulation of *E2F7* may play a role in the development of cervical cancer by modulating DNA replication function. *E2F7* was discovered in the TF–target regulatory network, and it regulates the *CDC6* gene. The expression of the *CDC6* gene is linked to cervical cancer metastasis and invasion (Wang et al., 2009). Consequently, *E2F7* was probably linked to the development of cervical cancer by modulating the expression of *CDC6*. It is evident from our analysis that the expression of *CDKN2A* and its positively correlated genes were involved in cervical cancer.

The present study had few limitations due to different sample sizes and the correspondence data updated in the online databases because of the variation in the size and information available in the datasets. *CDKN2A* mRNA expression levels are discovered to be a biomarker for cervical cancer prognosis. However, larger sample sizes are needed to corroborate these findings and uncover possible targets for cervical cancer diagnosis and treatment. Due to a paucity of clinical sample information, multivariate Cox regression analysis could not be done to further highlight the importance of *CDKN2A*.

5. Conclusions and future work

In conclusion, Targeting the expression of *CDKN2A* may serve as a gene target as it has significant interactions with other biological factors that contributed to cervical cancer. It could be used as a biomarker and therapeutic target in the diagnosis and treatment of cervical cancer. Additional research is needed to investigate and demonstrate the exploitation of DEGs for diagnosis, prognosis, and therapy of cervical cancer.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humgen.2022.201048.

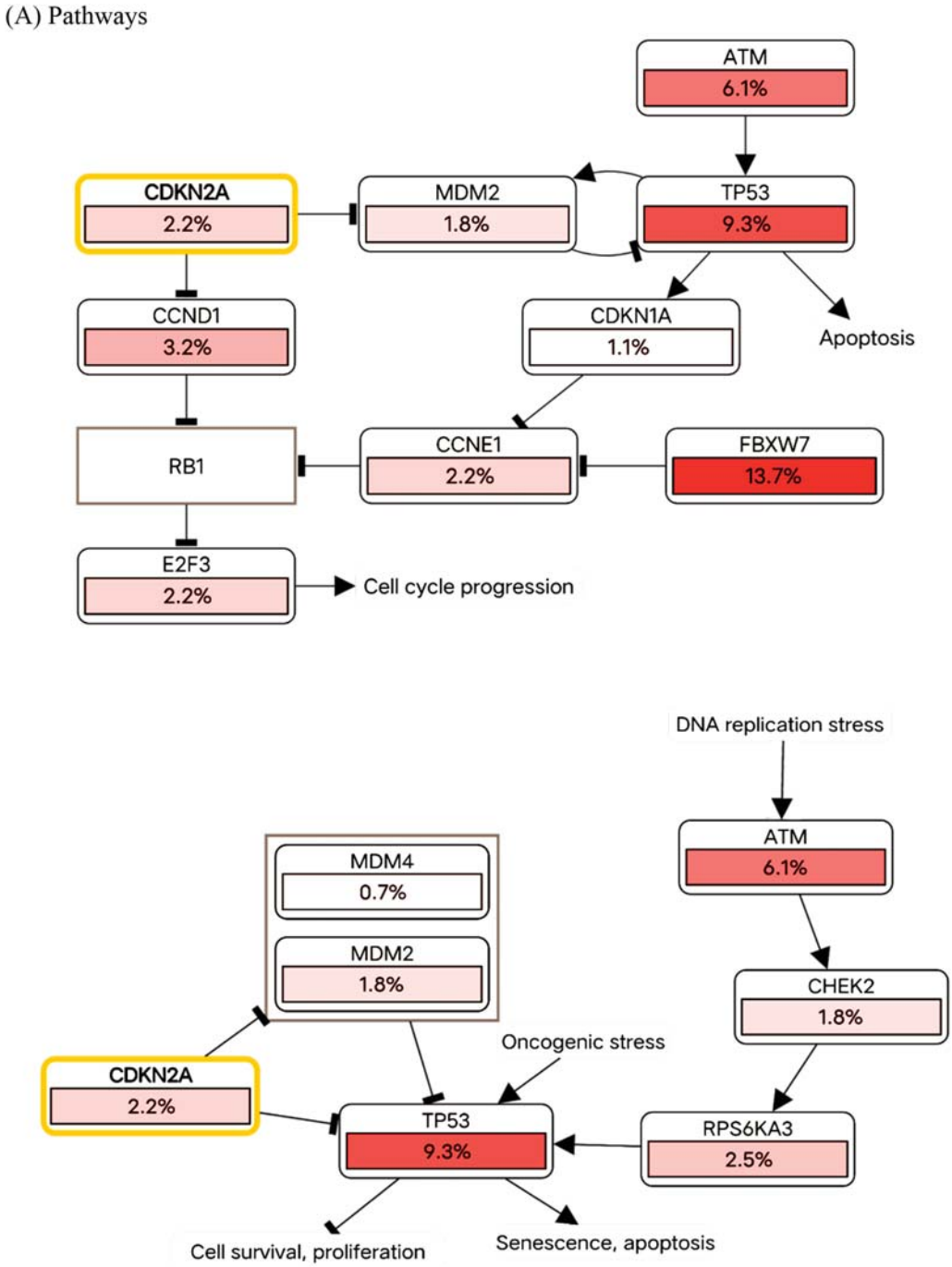
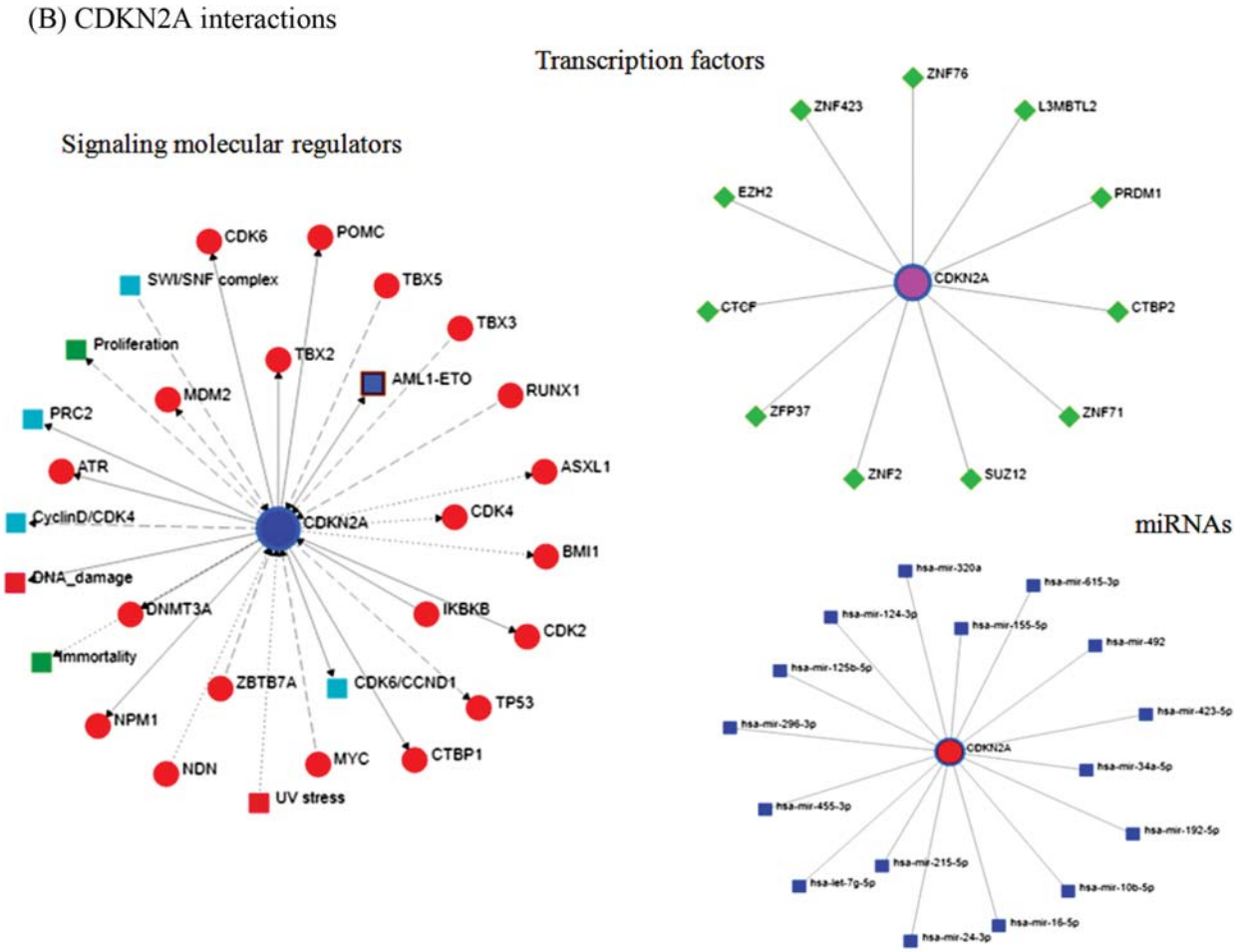


Fig. 5. Network modelling for prediction of CDKN2A associated pathways (A), interactions (B), and its gene expression correlation analysis (C).
(A) Pathways.
(B) CDKN2A interactions.
(C) Gene expression correlation analysis.



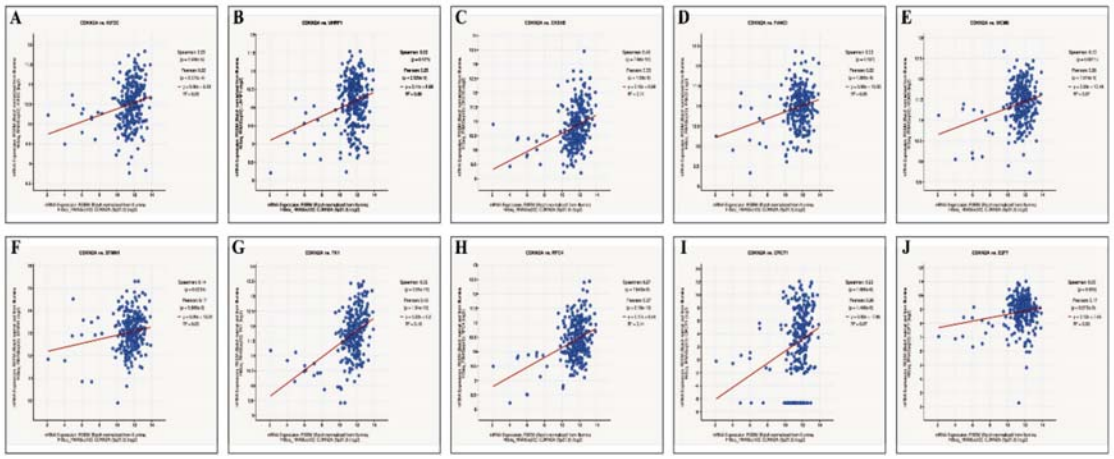


Fig. 5. (*continued*).

Data availability statement

Datasets analyzed during the current study are available in the GSE67522, GSE138080, and GSE75132. These datasets were retrieved from the NCBI database.

CRediT authorship contribution statement

Balraj Sudha: Data curation, Resources, Writing – original draft. **Arumugam Poornima:** Methodology. **Kanagaraj Suganya:** Methodology. **Kandasamy Swathi:** Data curation. **Nachimuthu Senthil Kumar:** Writing – review & editing, Investigation. **Sundaravadivelu Sumathi:** Supervision. **Paulchamy Chellapandi:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgement

The authors are thankful to the Department of Biotechnology (DBT), New Delhi, Govt. of India for the support in the form of Advanced State Biotech Hub, Mizoram University, Grant No: No.BT/NER/143/SP44475/2021.

References

Abbas, M., Srivastava, K., Imran, M., Banerjee, M., 2019. Genetic polymorphisms in DNA repair genes and their association with cervical cancer. Br. J. Biomed. Sci. 76 (3), 117–121.

Abdel-Fatah, T.M.A., Green, A.R., Lemetre, C., Moseley, P., Chan, S., Ellis, I.O., Balls, G., 2011. P4–09-11: Kinesin Family Member 2C (KIF2C) Is a New Surrogate Prognostic Marker in Breast Cancer (BC).

Aftab, A., Shahzad, S., Hussain, H.M.J., Khan, R., Irum, S., Tabassum, S., 2019. CDKN2A/P16INK4A variants association with breast cancer and their in-silico analysis. Breast Cancer 26 (1), 11–28. https://doi.org/10.1007/s12282-018-0894-0.

Anderson, L., Henderson, C., Adachi, Y., 2001. Phosphorylation and rapid relocalization of 53BP1 to nuclear foci upon DNA damage. Mol. Cell. Biol. 21 (5), 1719–1729. https://doi.org/10.1128/MCB.21.5.1719-1729.2001.

Babion, I., Miok, V., Jaspers, A., Huseinovic, A., Steenbergen, R.D., van Wieringen, W.N., Wilting, S.M., 2020. Identification of deregulated pathways, key regulators, and novel miRNA-mRNA interactions in HPV-mediated transformation. Cancers 12 (3), 700. https://doi.org/10.3390/cancers12030700.

Basic, V., Zhang, B., Domert, J., Pellas, U., Tot, T., 2021. Integrative meta-analysis of gene expression profiles identifies FEN1 and ENDOU as potential diagnostic biomarkers for cervical squamous cell carcinoma. Oncol. Lett. 22 (6), 1–18. https://doi.org/10.3892/ol.2021.13101.

Bowe, S., Butler, R., Halliday, D., Geremias, R., Cerbon, D., Huang, M., George, S., 2019. 240 First Year Experience of Hereditary Testing in Gynecological Cancer Patients in a Clinical Setting in the Bahamas.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68 (6), 394–424. https://doi.org/10.3322/caac.21492.

Chandrashekar, D.S., Bashel, B., Balasubramanya, S.A.H., Creighton, C.J., Ponce-Rodriguez, I., Chakravarthi, B.V., Varambally, S., 2017. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 19 (8), 649–658.

Cohen, P.A., Jhingran, A., Oaknin, A., Denny, L., 2019. Cervical cancer. Lancet 393 (10167), 169–182. https://doi.org/10.1016/S0140-6736(18)32470-X.

Das, M., Prasad, S.B., Yadav, S.S., Govardhan, H.B., Pandey, L.K., Singh, S., Pradhan, S., Narayan, G., 2013. Over expression of minichromosome maintenance genes is clinically correlated to cervical carcinogenesis. PLoS One 8 (7), e69607. https://doi.org/10.1371/journal.pone.0069607.

Dizon, D.S., Mackay, H.J., Thomas, G.M., Werner, T.L., Kohn, E.C., Hess, D., Covens, A. L., 2014. State of the science in cervical cancer: where we are today and where we need to go. Cancer 120 (15), 2282–2288. https://doi.org/10.1002/cncr.28722.

Gao, J., Aksoy, B.A., Dogrusoz, U., et al., 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. 2 (6) (pl1).

Gheit, T., 2019. Mucosal and cutaneous human papillomavirus infections and cancer biology. Front. Oncol. 9, 355. https://doi.org/10.3389/fonc.2019.00355.

Gurram, L., Kalra, B., Mahantshetty, U., 2020. Meeting the global need for radiation therapy in cervical cancer—an overview. Semin. Radiat. Oncol. 30 (4), 348–354. https://doi.org/10.1016/j.semradonc.2020.05.004.

Hatzistergos, K.E., Williams, A.R., Dykxhoorn, D., Bellio, M.A., Yu, W., Hare, J.M., 2019. Tumor suppressors RB1 and CDKN2a cooperatively regulate cell-cycle progression and differentiation during cardiomyocyte development and repair: implications for stimulating neomyogenesis with cell-based therapy. Circ. Res. 124 (8), 1184–1197. https://doi.org/10.1161/CIRCRESAHA.118.314063.

Hosseini, E.S., Meryet-Figuiere, M., Sabzalipoor, H., Kashani, H.H., Nikzad, H., Asemi, Z., 2017. Dysregulated expression of long noncoding RNAs in gynecologic cancers. Mol. Cancer 16 (1), 1–13. https://doi.org/10.1186/s12943-017-0671-2.

Jagarlamudi, K.K., Shaw, M., 2018. Thymidine kinase 1 as a tumor biomarker: technical advances offer new potential to an old biomarker. Biomark. Med 12 (9), 1035–1048. https://doi.org/10.2217/bmm-2018-0157.

Jiao, X., Sherman, B.T., Huang, D.W., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A., 2012. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 28 (13), 1805–1806. https://doi.org/10.1093/bioinformatics/bts251.

Jiao, Y., Feng, Y., Wang, X., 2018. Regulation of tumor suppressor gene CDKN2A and encoded p16-INK4a protein by covalent modifications. Biochem. Mosc. 83 (11), 1289–1298. https://doi.org/10.1134/S0006297918110019.

Li, Q., Wei, X., Zhou, Z.W., Wang, S.N., Jin, H., Chen, K.J., Shan, J.L., 2018. GADD45α sensitizes cervical cancer cells to radiotherapy via increasing cytoplasmic APE1 level. Cell Death Dis. 9 (5), 1–11. https://doi.org/10.1038/s41419-018-0452-x.

Liu, X., Liu, X., Han, X., 2021. FANCI may serve as a prognostic biomarker for cervical cancer: A systematic review and meta-analysis. Medicine 100 (51). https://doi.org/10.1097/MD.0000000000027690.

Luan, Y., Zhang, W., Xie, J., Mao, J., 2021. CDKN2A inhibits cell proliferation and invasion in cervical cancer through LDHA-mediated AKT/mTOR pathway. Clin. Transl. Oncol. 23 (2), 222–228. https://doi.org/10.1007/s12094-020-02409-4.

Manawapat-Klopfer, A., Thomsen, L.T., Martus, P., Munk, C., Russ, R., Gmuender, H., Iftner, T., 2016. TMEM45A, SERPINB5 and p16INK4A transcript levels are predictive for development of high-grade cervical lesions. Am. J. Cancer Res. 6 (7), 1524.

Mandal, P., Saha, S.S., Sen, S., Bhattacharya, A., Bhattacharya, N.P., Bucha, S., Sengupta, S., 2019. Cervical cancer subtypes harbouring integrated and/or episomal HPV16 portray distinct molecular phenotypes based on transcriptome profiling of mRNAs and miRNAs. Cell Death Dis. 5 (1), 1–13. https://doi.org/10.1038/s41420-019-0154-x.

Mizuarai, S., Machida, T., Kobayashi, T., Komatani, H., Itadani, H., Kotani, H., 2011. Expression ratio of CCND1 to CDKN2A mRNA predicts RB1 status of cultured cancer cell lines and clinical tumor samples. Mol. Cancer 10 (1), 1–11. https://doi.org/10.1186/1476-4598-10-31.

Narayan, G., Bourdon, V., Chaganti, S., Arias-Pulido, H., Nandula, S.V., Rao, P.H., Gissmann, L., Dürst, M., Schneider, A., Pothuri, B., Mansukhani, M., 2007. Gene dosage alterations revealed by cDNA microarray analysis in cervical cancer: identification of candidate amplified and overexpressed genes. Genes Chromosom. Cancer 46 (4), 373–384. https://doi.org/10.1002/gcc.20418.

Niu, G., Wang, D., Pei, Y., Sun, L., 2017. Systematic identification of key genes and pathways in the development of invasive cervical cancer. Gene 618, 28–41. https://doi.org/10.1016/j.gene.2017.03.018.

Ohtani, N., Zebedee, Z., Huot, T.J., Stinson, J.A., Sugimoto, M., Ohashi, Y., Hara, E., 2001. Opposing effects of Ets and Id proteins on p16 INK4a expression during cellular senescence. Nature 409 (6823), 1067–1070. https://doi.org/10.1038/35059131.

Pal, A., Potjer, T.P., Thomsen, S.K., Ng, H.J., Barrett, A., Scharfmann, R., Gloyn, A.L., 2016. Loss-of-function mutations in the cell-cycle control gene CDKN2A impact on glucose homeostasis in humans. Diabetes 65 (2), 527–533. https://doi.org/10.2337/db15-0602.

Qiu, H.Z., Huang, J., Xiang, C.C., Li, R., Zuo, E.D., Zhang, Y., Shan, L., Cheng, X., 2020. Screening and discovery of new potential biomarkers and small molecule drugs for cervical cancer: a bioinformatics analysis. Technol. Cancer Res. Treatment. 19 https://doi.org/10.1177/1533033820980112, 1533033820980112.

Qu, X., Shi, Z., Guo, J., Guo, C., Qiu, J., Hua, K., 2021. Identification of a novel six-gene signature with potential prognostic and therapeutic value in cervical cancer. Cancer Med. 10 (19), 6881–6896. https://doi.org/10.1002/cam4.4054.

Rajput, M., Kumar, M., Kumari, M., Bhattacharjee, A., Awasthi, A.A., 2020. Identification of key genes and construction of regulatory network for the progression of cervical cancer. Gene Rep. 21, 100965.

Robertson, K.D., Jones, P.A., 1999. Tissue-specific alternative splicing in the human INK4a/ARF cell cycle regulatory locus. Oncogene 18 (26), 3810–3820. https://doi.org/10.1038/sj.onc.1202737.

Saha, S.S., Chowdhury, R.R., Mondal, N.R., Roy, S., Sengupta, S., 2017. Expression signatures of HOX cluster genes in cervical cancer pathogenesis: Impact of human papillomavirus type 16 oncoprotein E7. Oncotarget 8 (22), 36591.

Shi, W., Huang, Q., Xie, J., Wang, H., Yu, X., Zhou, Y., 2020. CKS1B as drug resistance-inducing gene—a potential target to improve cancer therapy. Front. Oncol. 10, 1978. https://doi.org/10.3389/fonc.2020.582451.

Shibata, T., Lieblonga, B.J., Sasagawab, T., Nakagawa, M., 2019. The promise of combining cancer vaccine and checkpoint blockade for treating HPV-related cancer. Cancer Treat. Rev. 78, 8–16. https://doi.org/10.1016/j.ctrv.2019.07.001.

Shimo, A., Nishidate, T., Ohta, T., Fukuda, M., Nakamura, Y., Katagiri, T., 2007. Elevated expression of protein regulator of cytokinesis 1, involved in the growth of breast cancer cells. Cancer Sci. 98 (2), 174–181. https://doi.org/10.1111/j.1349-7006.2006.00381.x.

Stanca, M., Căpîlna, M.E., 2021. Prognostic factors associated with 5-year overall survival in cervical cancer patients treated with radical hysterectomy followed by adjuvant concurrent chemoradiation therapy at a tertiary care center in Eastern Europe. Diagnostics 11 (3), 570. https://doi.org/10.3390/diagnostics11030570.

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 Cancer J. Clin. 71 (3), 209–249. https://doi.org/10.3322/caac.21660.

Tang, Z., Kang, B., Li, C., Chen, T., Zhang, Z., 2019. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 47 (W1), W556–W560. https://doi.org/10.1093/nar/gkz430.

Tewari, K.S., Monk, B.J., 2014. Development of a platform for systemic antiangiogenesis therapy for advanced cervical cancer. Clin. Adv. Hematol. Oncol. 12 (11), 1–12.

van Dam, P.A., Rolfo, C., Ruiz, R., Pauwels, P., Van Berckelaer, C., Trinh, X., Van Laere, S., 2018. Potential new biomarkers for squamous carcinoma of the uterine cervix. ESMO open 3 (4), e000352. https://doi.org/10.1136/esmoopen-2018-000352.

Wang, S., Chen, X., 2018. Identification of potential biomarkers in cervical cancer with combined public mRNA and miRNA expression microarray data analysis. Oncol. Lett. 16 (4), 5200–5208. https://doi.org/10.3892/ol.2018.9323.

Wang, W.X., Zhang, W.J., Peng, Z.L., Yang, K.X., 2009. Expression and clinical significance of CDC6 and hMSH2 in cervical carcinoma. J Sichuan Univ 40 (5), 857–860.

Wang, J., Ma, S., Ma, R., Qu, X., Liu, W., Lv, C., Gong, Y., 2014. KIF2A silencing inhibits the proliferation and migration of breast cancer cells and correlates with unfavorable prognosis in breast cancer. BMC Cancer 14 (1), 1–9. https://doi.org/10.1186/1471-2407-14-461.

Wei, J., Wang, Y., Shi, K., Wang, Y., 2020. Identification of core prognosis-related candidate genes in cervical cancer via integrated bioinformatical analysis. Biomed. Res. Int. 2020.

Wu, B., Xi, S., 2021. Bioinformatics analysis of differentially expressed genes and pathways in the development of cervical cancer. BMC Cancer 21 (1), 1–5.

Wu, K., Yi, Y., Liu, F., Wu, W., Chen, Y., Zhang, W., 2018. Identification of key pathways and genes in the progression of cervical cancer using bioinformatics analysis. Oncol. Lett. 16 (1), 1003–1009. https://doi.org/10.3892/ol.2018.8768.

Xia, J., Benner, M.J., Hancock, R.E., 2014. NetworkAnalyst-integrative approaches for protein–protein interaction network analysis and visual exploration. Nucleic Acids Res. 42 (W1), W167–W174. https://doi.org/10.1093/nar/gku443.

Xue, H., Sun, Z., Wu, W., Du, D., Liao, S., 2021. Identification of hub genes as potential prognostic biomarkers in cervical cancer using comprehensive bioinformatics analysis and validation studies. Cancer Manag. Res. 13, 117.

Yuan, L., Li, G., Sun, G., Cao, D., 2021. Identification and Verification of a Four-Gene Signature Predicting Overall Survival for Cervical Cancer.

Zhang, Q., Qiao, L., Wang, X., Ding, C., Chen, J.J., 2018. UHRF1 epigenetically down-regulates UbcH8 to inhibit apoptosis in cervical cancer cells. Cell Cycle 17 (3), 300–308. https://doi.org/10.1080/15384101.2017.1403686.

Zhao, F., Siu, M.K., Jiang, L., Tam, K.F., Ngan, H.Y., Le, X.F., Cheung, A.N., 2014. Overexpression of forkhead box protein M1 (FOXM1) in ovarian cancer correlates with poor patient survival and contributes to paclitaxel resistance. PLoS One 9 (11), e113478. https://doi.org/10.1371/journal.pone.0113478.

Zhao, L., Zhang, Z., Lou, H., Liang, J., Yan, X., Li, W., Ou, R., 2018. Exploration of the molecular mechanisms of cervical cancer based on mRNA expression profiles and predicted microRNA interactions. Oncol. Lett. 15 (6), 8965–8972. https://doi.org/10.3892/ol.2018.8494.

Zhao, Q., Li, H., Zhu, L., Hu, S., Xi, X., Liu, Y., Liu, J., Zhong, T., 2020. Bioinformatics analysis shows that TOP2A functions as a key candidate gene in the progression of cervical cancer. Biomed. Rep. 13 (4), 1–10.

Zhou, G., Soufan, O., Ewald, J., Hancock, R.E., Basu, N., Xia, J., 2019. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 47 (W1), W234–W241. https://doi.org/10.1093/nar/gkz240.

Indian Journal of Gynecologic Oncology (2022)20:40
https://doi.org/10.1007/s40944-022-00639-8

ORIGINAL ARTICLE

Identification of Key Candidate Genes in the Progression of Cervical Cancer: An in Silico Analysis

Balraj Sudha[1] · Kandasamy Swathi[1] · Kanagaraj Suganya[1] · Arumugam Poornima[1] · Nachimuthu Senthil Kumar[2] · Sundaravadivelu Sumathi[1]

Received: 9 February 2022 / Revised: 9 February 2022 / Accepted: 5 September 2022

Abstract

Purpose Cervical cancer is one of the most widespread gynaecological tumours in women, and the molecular pathogenesis and chances of recurrence of the illness are still not clear. It is critical to identify new biomarkers to detect cervical cancer sooner to minimize women's incidence and mortality rates.

Methods Differentially expressed genes were screened from the GSE64517 transcriptome profile. A protein–protein interaction network analysis was used to find the hub genes associated with this condition. Experimental data from TCGA cervical cancer patients was used to confirm the statistical significance of important genes.

Results Totally twenty differentially expressed genes were retrieved. Their biological functions, connections, and their expression in cervical cancer were analysed virtually.

Conclusion This study suggested that *IGF2BP3* and *PTPRZ1* could be targeted for cervical cancer treatment.

Introduction

Globally, cancer is the leading cause of death. Cervical cancer (CC) is the most common malignant tumour among women and it is the fourth most frequent gynecological cancer, with 604,000 cases and 342,000 deaths reported in 2020 [1]. India and China account for more than one-third of the global CC burden, with an incidence rate of 97,000 and 106,000 new cases and 60,000 and 48,000 deaths, respectively. Worldwide, the average age of diagnosis and death from CC is 53 and 59 years, respectively [2]. CC is one of the three leading malignancies afflicting women under 45 years old in 146 of the 185 countries analysed [3].

Over time, cervical cancer develops gradually. The cervix cells undergo changes called dysplasia until cancer occurs inside the cervix, in which irregular cells will become cancer cells and begin to expand and spread more widely into the cervix and the surrounding areas [4]. Human papillomavirus (HPV) infection, many sexual partners, cigarette smoking, long-term use of contraceptive pills, and first sexual practices at an early age are all possible causes for cervical cancer [5]. With chemotherapy, radiotherapy and surgery, numerous approaches are used to treat cervical cancer. The efficacy of chemotherapy and radiotherapy is particular to cancer cells, and the entire normal cells can be killed [6].

Even though early detection can minimize CC mortality, most CC cases are identified at a later stage. Invasive CC identified early has 92% chance of a 5-year survival rate [7]. The high morbidity rate of CC is due to its late diagnosis and lack of effective treatment when the disease has progressed. Hence, it is a dire need to find molecular markers and processes for early identification and devise novel treatment techniques for effective CC management.

Researchers have used large-scale genome-wide analyses in normal and cancer samples to find and establish the molecular complexity driving cancer cells' biological behaviour and transform this information into molecular markers [2]. Despite this, current molecular marker identification studies require additional validation because the markers discovered may be inconsistent due to the heterogeneity of tumour cells and the highlighted regulatory systems in CC. As a result, re-analysing the data could

✉ Sundaravadivelu Sumathi
sumathi_bc@avinuty.ac.in

[1] Department of Biochemistry, Biotechnology and Bioinformatics, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu 641043, India

[2] Department of Biotechnology, Mizoram University, Aizawl, Mizoram 796004, India

Published online: 22 September 2022

reveal new information on the regulatory processes and signalling pathways linked to CC. One of the key causes of poor survival in CC is treatment resistance [7]. Many prior cancer in silico research has proven to be very valuable and reliable, implying that integrated bioinformatic analysis can be utilised to aid in the discovery of new biomarkers, networks, and mechanisms involved in carcinogenesis. Part of the sample data from the GSE64217 transcriptome profile was analysed by the Yao and Liu, 2018 [8]. Their study focused on normal (GSM1566485, GSM1566488) and cervical intraepithelial neoplasia (GSM1566486, GSM1566489) sample data only. They have reported that *EGR1* was connected to cervical intraepithelial neoplasia.

In this study, we have selected normal (GSM1566485, GSM1566488) and cervical cancer tissue sample (GSM1566487, GSM1566489) data from GSE64217 dataset for the identification of differentially expressed genes related to cervical cancer. To the best of our knowledge, this is the first study conducted with cervical cancer tissue and normal sample expression profile data from the GSE64217 dataset.

Methodology

Microarray Data

The cervical cancer gene expression data (GSE64217) was downloaded from Gene Expression Omnibus. It contains two normal and two cervical cancer samples. GPL10558 Illumina Human HT-12 V4.0 expression bead chip was employed for this study.

Screening of Differentially Expressed Genes (DEGs)

GEO2R (ncbi.nlm.nih.gov/geo/geo2r) was used to distinguish differentially expressed genes in cervical cancer and healthy cervical samples. To establish a balance between detecting statistically important genes and decreasing false positives, the adjusted P-values (adj.P) and Benjamin and Hochberg false discovery rates were used. DEGs with a log fold-change more than and less than 2 and an adj.P-value less than 0.05 were considered statistically significant.

Functional Annotation and Pathway Enrichment Analysis

PANTHER bioinformatics software was used to perform Gene Ontology enrichment analysis for the chosen DEGs. Biological process, cellular component, and molecular functions were identified in the Gene Ontology categories for the top 20 significant genes. P 0.05 was used as the significant level [9].

Protein–protein Interaction of Selected DEGs

Using network analyst and the IMEx interactome, the protein–protein interaction networks of newly found differentially expressed genes were created. Network Analyst can access a large, high-quality PPI database due to InnateDB, a member of International Molecular Exchange project [10]. The selected differentially expressed genes were used for constructing the protein–protein interaction.

Validation with TCGA Samples

The top ranked significant gene expression profile was validated and compared with samples in the TCGA database using UALCAN [11]. The expression of selected genes was validated on the basis of cervical cancer stages. The TCGA database contained the expression information of normal, stage 1, stage 2, stage 3 and stage 4 in the sample size of 3, 161, 69, 46 and 22, respectively. The statistically significant genes were selected and identified as a therapeutic target.

Survival Analysis

The overall analysis of selected genes in cervical cancer was analysed with survival data of 304 cervical cancer samples in KM plotter database. Log-rank $P < 0.05$ was considered as significant [12].

Statistics

The adj.P-value was corrected for multiple comparisons to retrieve the significant differentially expressed genes using the Benjamini and Hochberg's false discovery rate. The mandatory statistical analysis was done for all the outputs obtained.

Results

Identification of Differentially Expressed Genes

In GSE64517 dataset, 29,406 microarray probe identifiers were retrieved. In that, the missing values and values without gene names were removed and significant genes were retrieved. Among them, 970 were up-regulated genes and 621 genes were down-regulated (Fig. 1). The top 10 significant upregulated (Table 1) and down-regulated (Table 2) genes were screened for further study.

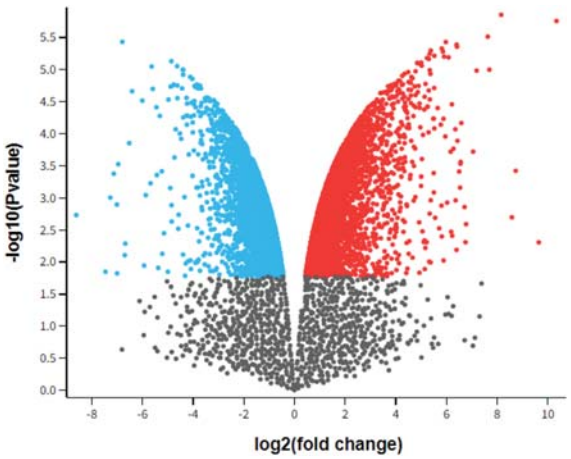


Fig. 1 Differentially expressed genes (red – up regulated genes; blue – down regulated genes)

Gene Enrichment Analysis

Gene enrichment analysis was done using PANTHER to determine to overlap of the biological function. The topmost significantly supplemented DEGs in the molecular functions are binding, catalytic activity, molecular function regulator, structural molecular activity, and transporter activity (Fig. 2). The biological processes are biological regulation, cellular process, development process, immune system process, interspecies interaction between organisms, localization, metabolic process, multicellular organismal process and response to stimulus (Fig. 3); the cellular components are cellular anatomical entity, protein-containing complex and intracellular (Fig. 4).

Protein–Protein Interaction

Using selected twenty DEGs, protein–protein interaction network was assembled using Network Analyst. The subnetwork possessed 13 seeds, 261 nodes (proteins) and 283

Table 1 Significant up regulated genes

adj.P.Val	*P* value	T	logFC	Gene symbol	Gene title
0.00659	1.4E-06	81.8983	8.163822	*HP*	Haptoglobin
0.00659	1.75E-06	76.54745	10.34803	*SCGB3A1*	secretoglobin family 3A Member 1
0.00659	3.08E-06	64.50635	7.637129	*SAA2*	Serum amyloid A2
0.00659	3.73E-06	60.86968	5.978707	*PIGR*	polymeric Immunoglobulin receptor
0.00659	4.1E-06	59.1231	6.401727	*TGM3*	Transglutaminase 3
0.00659	4.43E-06	57.7909	6.419478	*PIANP*	PILR alpha associated Neural protein
0.00659	0.00000472	56.6914465	5.82025006	*SAA4*	serum amyloid A4, constitutive
0.00659	5.27E-06	54.82307	5.865289	*C1orf194*	Chromosome 1 open reading frame 194
0.00659	5.71E-06	53.49634	5.339603	*WFDC2*	WAP four-disulfide core Domain 2
0.00659	6.12E-06	52.393	5.523774	*CCNA1*	Cyclin A1

Table 2 Significant down regulated genes

adj P value	*P* value	T	logFC	Gene symbol	Gene title
0.00659	7.44E-06	− 49.3822	− 4.85925	*SERPINB7*	Serpin family B member 7
0.00659	8.81E-06	− 46.9139	− 4.63112	*FBN2*	Fibrillin 2
0.00659	8.93E-06	− 46.7099	− 5.63316	*MMP12*	Matrix metallopeptidase 12
0.00659	1E-05	− 45.1254	− 4.40051	*CA2*	Carbonic anhydrase 2
0.00659	1.01E-05	− 45.0582	− 4.39883	*IGF2BP3*	Insulin like growth factor 2 mRNA binding protein 3
0.00659	1.39E-05	− 40.8554	− 4.02718	*HORMAD1*	HORMA domain containing 1
0.00659	1.71E-05	− 38.4027	− 3.84693	*APOC1*	Apolipoprotein C1
0.00659	1.71E-05	− 38.3732	− 4.76296	*PNCK*	Pregnancy up-regulated nonubiquitous CaM kinase
0.00659	1.76E-05	− 38.0267	− 4.62014	*PTPRZ1*	Protein tyrosine phosphatase, receptor type Z1
0.00659	1.77E-05	− 37.9648	− 4.80722	*KRT1*	Keratin 1

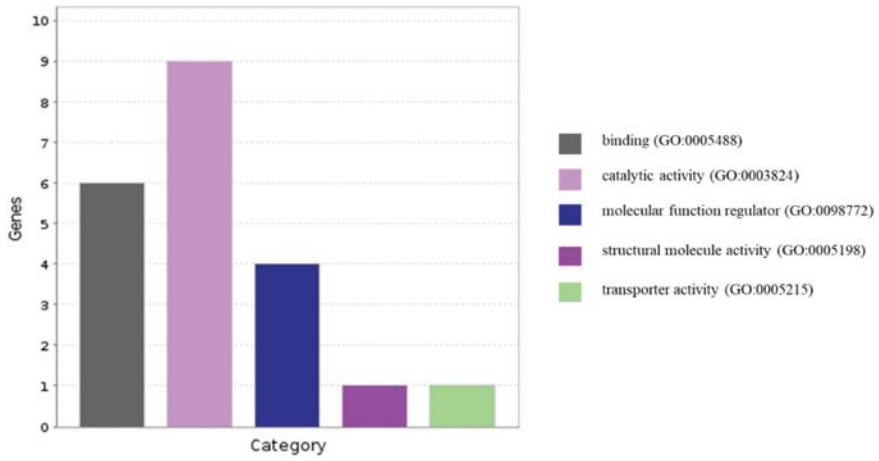


Fig. 2 Molecular Function of DEGs

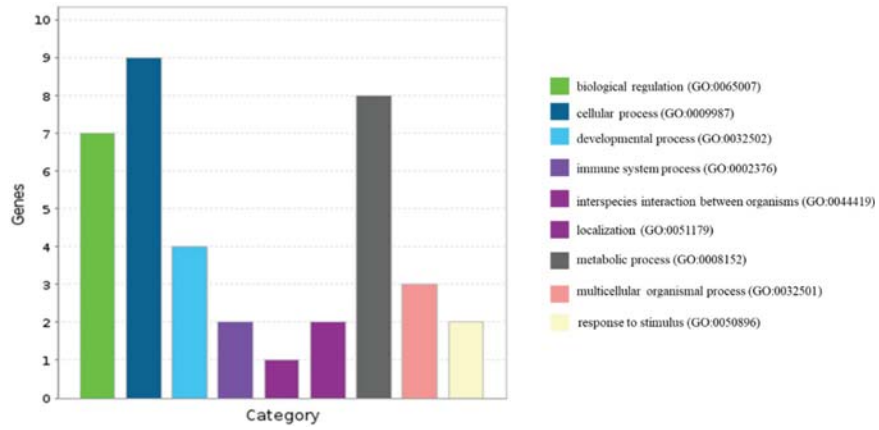


Fig. 3 Biological Process of DEGs

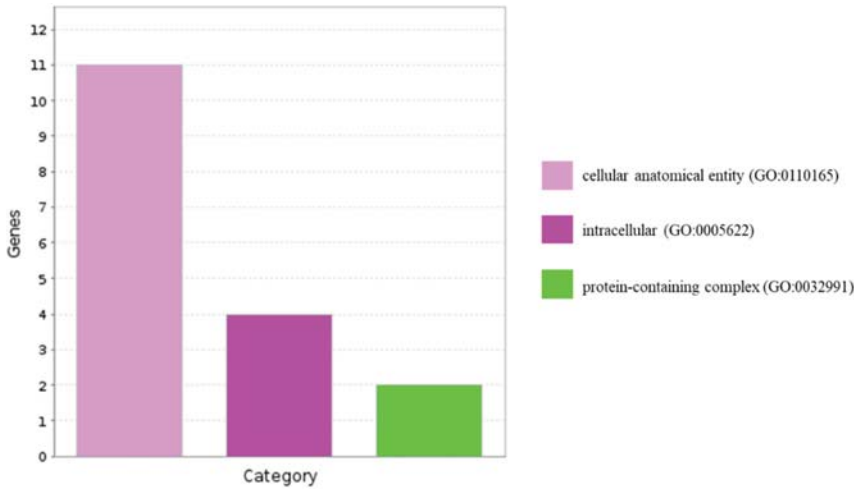


Fig. 4 Cellular Component of DEGs

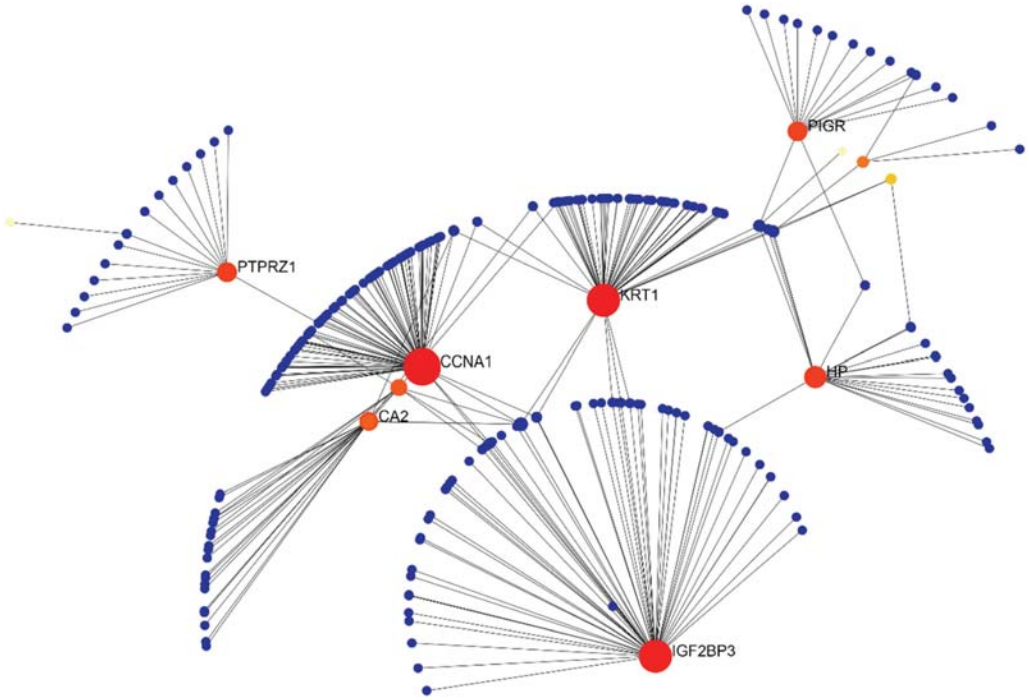


Fig. 5 Protein—Protein Interaction of DEGs

edges (interactions) (Fig. 5). *CCNA1* (degree centrality = 80; betweenness centrality = 19,252.94), *KRT1* (degree centrality = 58; betweenness centrality = 16,536.16), *IGF2BP3* (degree centrality = 57; betweenness centrality = 12,249.35), *HP* (degree centrality = 20; betweenness centrality = 4048.443) and *PTPRZ1* (degree centrality = 15; betweenness centrality = 3779) were the top-ranking nodes in the combined datasets derived from subnetwork topology measures. The selected DEGs genes involved KEGG pathways are tabulated (Table 3).

Table 3 Pathway associated with DEGs

Pathway	*P* value	Adj *P* value
Cell cycle	2.10E-23	6.69E-21
Cellular senescence	6.64E-13	1.06E-10
Small cell lung cancer	1.37E-12	1.45E-10
Viral carcinogenesis	1.23E-11	9.75e-10
Chronic myeloid leukemia	9.49E-10	5.76E-08
Pathways in cancer	1.09E-09	5.76E-08

Validation of DEGs with TCGA Samples

The top-ranking nodes *CCNA1, KRT1, IGF2BP3, HP* and *PTPRZ1* were validated with the TCGA samples using ALCAN. The selected gene expression was validated for different cervical cancer stages. From that, *IGF2BP3* and *PTPRZ1* were statistically significant for normal vs stage1, normal vs satge2, normal vs stage3, normal vs stage4 (Table 4).

Survival Analysis of *IGF2BP3* and *PTPRZ1*

The prognostic significance of *PTPRZ1* was virtually examined for cervical cancer patients' survival time. The KM plotter demonstrated that *PTPRZ1* might be used to distinguish between high and low-risk cervical cancer patients. In cervical cancer samples, the level of *PTPRZ1* mRNA was considerably lower (Fig. 6). Furthermore, lower expression levels of *PTPRZ1* have been linked to a poor prognosis for relapse-free survival in women with cervical cancer. However, *IGF2BP3* was not linked to overall survival in cervical cancer patients.

Table 4 Clinical validation of selected genes with TCGA samples

Comparison	Statistical significance				
	CCNA1	*KRT1*	*IGF2BP3*	*HP*	*PTPRZ1*
Normal-vs-Stage1	9.68E-01	1.12E-01	3.33E-16*	5.99E-01	1.63E-12*
Normal-vs-Stage2	6.63E-01	3.28E-02	1.07E-11*	5.09E-01	3.24E-06*
Normal-vs-Stage3	6.44E-01	9.33E-02	5.82E-07*	4.90E-01	7.74E-07*
Normal-vs-Stage4	6.48E-01	9.09E-02	4.19E-04*	4.15E-01	5.74E-03*
Stage1-vs-Stage2	2.17E-01	2.68E-01	8.79E-01	4.82E-02*	6.47E-01
Stage1-vs-Stage3	1.96E-01	1.70E-01	4.54E-01	5.46E-02	5.95E-01
Stage1-vs-Stage4	2.94E-01	3.55E-01	3.72E-01	4.69E-01	9.90E-01
Stage2-vs-Stage3	8.19E-01	3.30E-01	4.05E-01	9.99E-01	9.72E-01
Stage2-vs-Stage4	9.84E-01	7.96E-02	3.80E-01	3.29E-01	8.04E-01
Stage3-vs-Stage4	8.54E-01	1.17E-01	8.25E-01	3.29E-01	7.40E-01

* Statistically significant

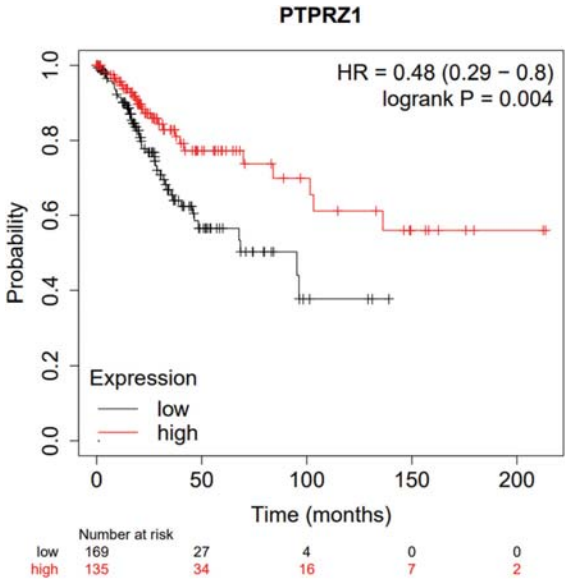


Fig. 6 Survival analysis of *PTPRZ1* with cervical cancer patients

Discussion

Cervical cancer is the worst type of cancer, affecting the cell lining of the cervix, which undergoes a series of alterations. Because precancerous cells can turn cancerous, early diagnosis of cervical cell alterations and appropriate therapy can help prevent cervical cancer from progressing. Researchers have recently been attempting to decipher the molecular mechanism of the disease. Medications, surgery, and radiation therapy have all been widely used to treat cervical cancer in the past [13]. As a result, studies to establish processes underlying the disease's aetiology and linked genes on cervical cancer growth are critical. We employed microarray-based transcriptome analysis to find dysregulated genes in cervical cancer. Microarray has been widely utilised to research genetic abnormalities in cancer and uncover disease-specific prognostic biomarkers and therapeutic targets; hence, we used a microarray-based transcriptome framework to analyse dysregulated genes in cervical cancer.

Other studies have also mined the GSE64217 dataset [14–17]. The previous study with the GSE64217 dataset covered Cervical intraepithelial neoplasia (CIN) and normal samples alone [8]. The GSE64217 dataset had not yet been examined with cervical cancer patients and normal samples. Therefore, the GEO2R was used to perform the analyses of GSE64217 dataset to determine any potentially relevant factors.

The study's findings showed that twenty significant genes were differentially expressed (10 upregulated and 10 downregulated genes) between patients' samples compared to the normal cervix. The key candidate genes were significantly enriched in different biological processes, molecular functions and cellular components which were analysed using PANTHER. Protein–protein interaction of selected genes showed 283 interactions with 261 proteins for 13 seeds (our selected genes of interest). Through pathway analysis, Network Analyst showed that selected DEGs were associated with the following pathways: cell cycle, cellular senescence, small cell lung cancer, viral carcinogenesis, chronic myeloid leukaemia and pathways in cancer. Based on the degree and betweenness centrality, we have selected the top 5 genes namely *CCNA1, KRT1, IGF2BP3, HP* and *PTPRZ1*.

Further, we have validated the expression of five genes with TCGA samples. Significant expression was observed for *IGF2BP3* and *PTPRZ1*. Both the genes showed similar kind of results for normal vs all the four stages of cervical cancer. Many studies have been published on the clinical significance of *IGF2BP3* and *PTPRZ1* in human cancer

[18]. But very limited study was reported so far regarding the clinical importance of *IGF2BP3* and *PTPRZ1* in cervical cancer. Zhu et al., 2021 reported that decreased expression of *DARS-AS1* could prevent the development of cervical cancer cells by downregulating *IGF2BP3* [19]. It means that *IGF2BP3* overexpression may play a role in cervical cancer carcinogenesis. Meng et al. 2000, found that the secreted growth factor pleiotrophin (PTN), which is one of *PTPRZ1's* key ligands, enhanced tumour development and angiogenesis by signalling through ligand-dependent receptor inactivation of *PTPRZ1* [20]. Ma et al., 2011 reported that the expression of *PTPRZ1* in cervical carcinoma was substantially higher than in normal cervical epithelium. These are the findings supports our results that *PTPRZ1* may be involved in the formation or progression of cervical cancer. Although - *IGF2BP3* and *PTPRZ1* has been found to have a role in tumour cell invasion, migration, and adhesion [21], few investigations on its role in tumour growth have been performed. Very few studies were reported about the role of *IGF2BP3* and *PTPRZ1* in cervical cancer. Collectively, our result shows that *IGF2BP3* and *PTPRZ1* expression is linked to cervical cancer.

Conclusion

IGF2BP3 and *PTPRZ1* could be targeted because they interact with proteins contributing to cervical cancer. Because of the heterogeneity in the size and information accessible in the datasets, the current study had a few limitations related to varying sample sizes and the correspondence data updated in the online databases. The expression levels of *IGF2PB3* and *PTPRZ1* mRNA have been discovered to be diagnostic for cervical cancer prognosis. Larger sample sizes are required to confirm these findings and identify potential cervical cancer diagnosis and treatment targets. Overall, we believe that *IGF2BP3* and *PTPRZ1* could be exploited as a diagnostic and therapeutic target for cervical cancer diagnosis and treatment. And also, selected DEGs were highly involved in the cell cycle. So, additional study is needed to investigate and demonstrate the use of DEGs for cervical cancer diagnosis, prognosis, and therapy.

Acknowledgements We thank the Department of Biochemistry, Biotechnology, and Bioinformatics, Avinashilingam Institute for Home Science and Higher Education for women, Coimbatore, Tamil Nadu, for encouraging and providing the required facility for completion of this work.

Author's contributions BS contributed to data curation, resources, and writing original draft preparation; AP contributed to methodology; KS contributed to methodology; KS contributed to data curation; NSK contributed to writing—reviewing and editing, and investigation; SS contributed to writing—reviewing and editing, investigation, and supervision.

Declarations

Conflict of interests The authors declare that they have no competing interests.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J Clin. 2021;71(3):209–49. https://doi.org/10.3322/caac.21660.
2. Sriharikrishnaa S, Shukla V, Khan GN, Eswaran S, Adiga D, Kabekkodu SP. Integrated bioinformatic analysis of miR-15a/16-1 cluster network in cervical cancer. Reprod Biol. 2021;21(1): 100482. https://doi.org/10.1016/j.repbio.2021.100482.
3. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, Bray F. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. Lancet Glob Health. 2020;8(2):e191-203. https://doi.org/10.1016/S2214-109X(19)30482-6.
4. Tarney CM, Han J. Postcoital bleeding: a review on etiology, diagnosis, and management. Obstet Gynecol Int. 2014;17:2014. https://doi.org/10.1155/2014/192087.
5. Wu J, Zhao Y, Li F, Qiao B. MiR-144-3p: a novel tumor suppressor targeting MAPK6 in cervical cancer. J Physiol Biochem. 2019;75(2):143–52. https://doi.org/10.1007/s13105-019-00681-9.
6. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019;393(10167):169–82. https://doi.org/10.1016/S0140-6736(18)32470-X.
7. Adiga D, Eswaran S, Pandey D, Sharan K, Kabekkodu SP. Molecular landscape of recurrent cervical cancer. Crit Rev Oncol Hematol. 2021;157: 103178. https://doi.org/10.1016/j.critrevonc.2020.103178.
8. Yao S, Liu T. Analysis of differential gene expression caused by cervical intraepithelial neoplasia based on GEO database. Oncol Lett. 2018;15(6):8319–24. https://doi.org/10.3892/ol.2018.8403.
9. Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. In: Nikolsky Y, Bryant J, editors. Protein networks and pathway analysis. UK: Humana Press; 2009. p. 123–40.
10. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ. InnateDB: systems biology of innate immunity and beyond–recent updates and continuing curation. Nucleic Acids Res. 2013;41:D1228-33. https://doi.org/10.1093/nar/gks1147.
11. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi BV, Varambally S. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia. 2017;19(8):649–58. https://doi.org/10.1016/j.neo.2017.05.002.
12. Lánczky A, Győrffy B. Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. J Med Internet Res. 2021;23(7): e27633. https://doi.org/10.2196/27633.
13. Sudhalkar N, Rathod NP, Mathews A, Chopra S, Sriram H, Shrivastava SK, Goda JS. Potential role of cancer stem cells as biomarkers and therapeutic targets in cervical cancer. Cancer Rep. 2019;2(2): e1144. https://doi.org/10.1002/cnr2.1144.

14. Feng D, Lin J, Wang W, Yan K, Liang H, Liang J, Yu H, Ling B. Wnt3a/β-Catenin/CBP activation in the progression of cervical intraepithelial neoplasia. Pathol Oncol Res. 2021. https://doi.org/10.3389/pore.2021.609620.
15. Liu S, Gu L, Wu N, Song J, Yan J, Yang S, Feng Y, Wang Z, Wang L, Zhang Y, Jin Y. Overexpression of DTL enhances cell motility and promotes tumor metastasis in cervical adenocarcinoma by inducing RAC1-JNK-FOXO1 axis. Cell Death Dis. 2021;12(10):1–9. https://doi.org/10.1038/s41419-021-04179-5.
16. Wu B, Xi S. Bioinformatics analysis of differentially expressed genes and pathways in the development of cervical cancer. BMC Cancer. 2021;21(1):1–5.
17. Guo Y, Luo S. miR-140-5p inhibits cervical cancer cell phenotypes via downregulating FEN1 to halt the cell cycle. Mol Med Rep. 2020;22(6):4919–30. https://doi.org/10.3892/mmr.2020.11552.
18. Xia Z, Ouyang D, Li Q, Li M, Zou Q, Li L, Yi W, Zhou E. The expression, functions, interactions and prognostic values of PTPRZ1: a review and bioinformatic analysis. J Cancer. 2019;10(7):1663. https://doi.org/10.7150/jca.28231.
19. Zhu J, Han S. Downregulation of LncRNA DARS-AS1 Inhibits the tumorigenesis of cervical cancer via inhibition of IGF2BP3. Onco Targets Ther. 2021;14:1331. https://doi.org/10.2147/OTT.S274623.
20. Meng K, Rodríguezpeña A, Dimitrov T, Chen W, Yamin M, Noda M. et al. (2000) Pleiotrophin signals increased tyrosine phosphorylation of β-catenin through inactivation of the intrinsic catalytic activity of the receptor-type protein tyrosine phosphatase β/ζ. In: Proceedings of the national academy of sciences of the United States of America. 97: 2603–8. https://doi.org/10.1073/pnas.020487997
21. Ma Y, Ye F, Xie X, et al. Significance of PTPRZ1 and CIN85 expression in cervical carcinoma. Arch Gynecol Obstet. 2011;284:699–704. https://doi.org/10.1007/s00404-010-1693-9.

Indian Journal of Gynecologic Oncology (2024) 22:64
https://doi.org/10.1007/s40944-024-00831-y

ORIGINAL ARTICLE

Translation of Gene Expression Data Into Personalized Treatment in Cervical Cancer: Machine Learning Approach

Balraj Sudha[1] · M. Krishnaveni[2] · Sundaravadivelu Sumathi[1]

Received: 31 October 2023 / Revised: 18 February 2024 / Accepted: 7 April 2024 / Published online: 28 April 2024

Abstract

Background The majority of cervical cancers have been linked to the infection by human papillomavirus (HPV). There is a need to identify genes which play a role in the final manifestation of cervical cancers following HPV infection.

Objective To identify a number of genetic markers associated with cervical cancer that may aid in the disease's diagnosis or prognosis using machine learning methods.

Methods To do this, we will assess numerous gene expression profiles with integrative machine learning approaches such as random forest (RF) and support vector machine-based recursive feature elimination (SVMRFE). The conceptual analysis consists of following steps: (i) gene expression analysis and (ii) machine learning analysis for predicting genes.

Result The selected datasets were GSE75132 and GSE39001 for this study. Accuracy and cross validation were carried for both SVM-RFE and RF model for the gene identification purpose. R Bioconductor packages "GEOquery," "limma," and "umap" were utilized. The selected genes of machine learning methods were combined. The SVM model was the best for predicting the gene expression microarray profile based on the accuracy this study was able to get.

Conclusion The SVM model indicated that genes might be used as biomarkers to identify biological processes. The identified genes were considered as potential gene signatures in cervical cancer detection, and their interactions were studied.

Keywords Cervical cancer · Machine learning · GEO dataset · Biomarkers · Gene expression

Introduction

Cervical carcinoma is the fourth most common cancer in women and the leading cause of death from cancer. More than 600,000 new cases and 342,000 deaths have been reported up until 2020. In twenty-three countries, cervical cancer is the most common cancer, and in thirty-six countries, it is the leading cause of cancer death [1]. Following surgery or radiotherapy, some patients experience a reoccurrence. Patients who relapse have few treatment options and a poor prognosis [2]. Cervical cancer treatment has been found to be ineffective. As a result, new approaches to increasing the efficacy of cervical cancer treatment are urgently required.

Microarray is a technique for monitoring the expression of a large number of genes in real time. In recent years, it has become one of the most important tools for global gene expression analysis in molecular biology research. Because of the large amount of expression data generated by this technology, it is now possible to study some complex biological problems, and machine learning methods are playing an important role in the analysis process. Many machine learning methods have been applied to major areas of gene expression analysis, or have the potential to be applied in the future. Clustering, classification, dynamic modeling, and reverse engineering are some of these areas. Microarray technology was used to define a significant number of gene expressions that could be linked to a specific cancer. Meta-analysis is promoted as a way to solve of low predictive significance in studies, as well as the ability to analyze overlapped genes between datasets, extract optimum values across multiple datasets, and identify the most commonly expressed genes [3, 4]. Biology research has entered the

✉ Sundaravadivelu Sumathi
sumathi_bc@avinuty.ac.in

[1] Department of Biochemistry, Biotechnology, and Bioinformatics, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India

[2] Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India

post-genome era with the completion of the Human Genome Project. Despite the fact that biologists have amassed a massive amount of DNA sequence data, the specifics of how these sequences work are still largely unknown. Even the simplest organisms' genomes are extremely complex [5]. A molecular biologist's focus was on a few genes or proteins. The behavior of genes can be studied globally using large-scale biological information quantification methods such as microarray and DNA sequencing. Automatic analysis of the overall relationship hidden behind a large number of genes from their expression is currently in high demand [6].

Differentially co-expressed gene modules can provide information about disease formation, progression, and their regulatory patterns [7]. The differential co-expression networks were applied in the past to common complex diseases such as endometrial cancer, breast cancer, types 1 and 2 diabetes and ovarian cancer. Support vector machine is used in cancer including cervical cancer [8, 9]. Though Random Forest is more used in such researches in cervical cancer recursive feature elimination is also used [10].

Machine learning may play an important role in the analysis process due to the complex nature of biological data. Moreover, until now, the majority of previous studies either used statistical analysis methods or machine learning-based methods [10–13]. With this background, this study aimed to identify key genes from the gene expression dataset.

Methodology

Data Collection

Gene expression data of individual cervical cancer microarray dataset was taken from publicly available source. The datasets GSE75132 [14] and GSE39001[15] were selected which belongs to *homo sapiens*. GSE75132 included twenty-one normal and one cancer samples. GSE39001 included twelve normal sample and thirty-nine cancer samples expression data. The raw data were retrieved. R Bioconductor packages "GEOquery," "limma," and "umap" were utilized for this data retravel process.

Data Pre-Processing

For the accurateness and purity of the resultant about more than 1000 microprobe identifiers were normalized. Missing values were removed.

df [] < —lapply (df, function(x) {x[is.na(x)] < —mean (x, na.rm = TRUE) x}).

Data Identification

The differentially expressed genes between CeCx and normal cervical samples were partitioned using GEO2R. The adjusted *P*-values (adj.P) and Benjamini and Hochberg false discovery rates was used to provide a balance between the discoveries of statistically significant genes and limit false—positives. Only genes with a log fold-change > 1 and adj.*P*-value of < 0.05 which were regarded as statistically significant DEGs were selected [16].

Data Categorization

The upregulated and downregulated genes were identified based on log 2 value (greater than log 2 and less than log 2). The upregulated and downregulated genes were prepared as an input for the variable prediction.

Feature Selection and Machine Learning Methods

Feature selection and machine learning approaches was attained identify the common optimum genes. This is also to fully utilize the measurements of different machine learning methods (supervised and unsupervised) by collaborating them in selecting the significant genes. Support vector machine-based recursive feature elimination (SVM-RFE) and Random Forest (RF), which embeds variable selection as part of their learning process and to evaluate them according to the classification of the results of specific machine learning methods [17].

GSEA Analysis for Predicted Genes

The predicted genes were subjected for their biological function prediction which involved in the cervical cancer progression. The predicted output was retrieved as microarray probe identifiers. The probe id was converted into gene symbol. The biological process, molecular function and cellular components were reported using DAVID tool.

Results and Discussion

Data Extraction

In our study, two mRNA profiles GSE39001 and GSE75132 were downloaded from Gene Expression Omnibus. mRNA profile GSE39001 included 43 cervical cancer biopsy and 12 healthy subjects. The mRNA profiles of GSE39001 were detected by Affymetrix Human HG-Focus Target Array. mRNA profile GSE75132 was downloaded from the Gene Expression Omnibus which included one cervical cancer

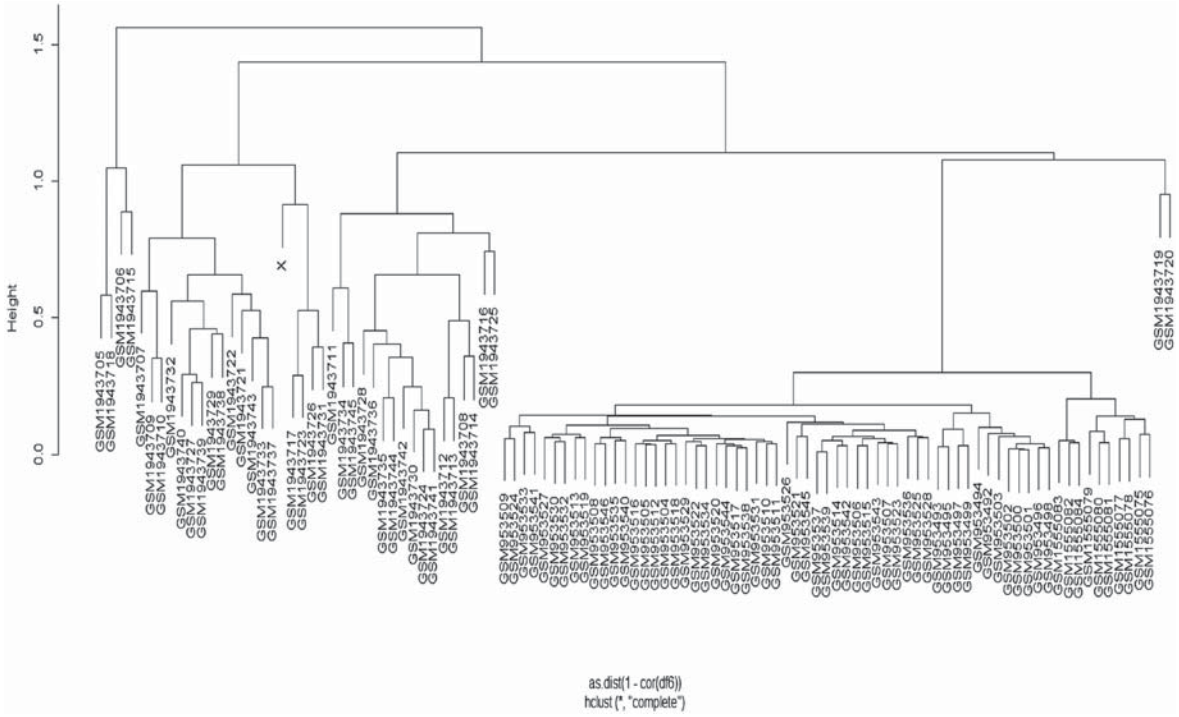


Fig. 1 Hierarchical clustering

biopsy and twenty-one healthy subjects. The mRNA profiles of GSE39001 were detected by Affymetrix Human HG-Focus Target Array. Hierarchical clustering graph showed the connection between the tumor and samples (Fig. 1).

Data Pre-Processing

The raw data was extracted from CEL files. It contains 55 samples (normal and cancer), and the corresponding expression of 8793 unique features were retrieved for GSE39001, while 22 samples (normal and cancer) and the corresponding expression of 16,383 unique features were retrieved for GSE75132. Linear model for microarray (LIMMA) data was used to perform the analysis in order to select out the differentially expressed genes. LIMMA tests the average difference in log expression levels of two groups per gene by fold change. The control denoted as "0", the cancer sample denoted as "1". The *P* value and fold value were assigned for the statistical significance and biological significance of expression dataset. Here, the *P* value cutoff is 0.05 and fold cut off is 0.5. Raw data was imputed to remove the missing values and was normalized. After the normalization step, both data-set were merged together as a metadata.

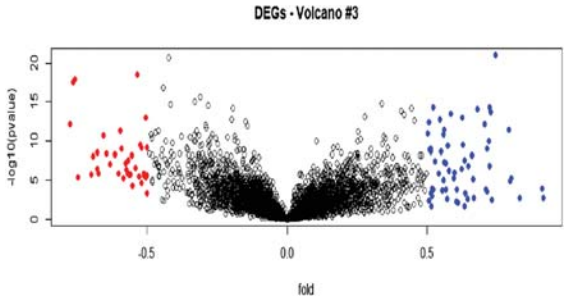


Fig. 2 Volcano plot for differentially expressed genes

Data Categorization

Significant differential expressed genes were classified into upregulated (13) and downregulated genes (169) from the meta-data based on the statistical value (*p* value) and biological value (fold change) shown in Fig. 2 (volcano plot) and Fig. 3 (heatmap). The categorized up and downregulated genes were taken as an input for the construction of SVM and Random Forest model. The gene count was very low for upregulated genes. Hence, the downregulated genes were selected for the next process.

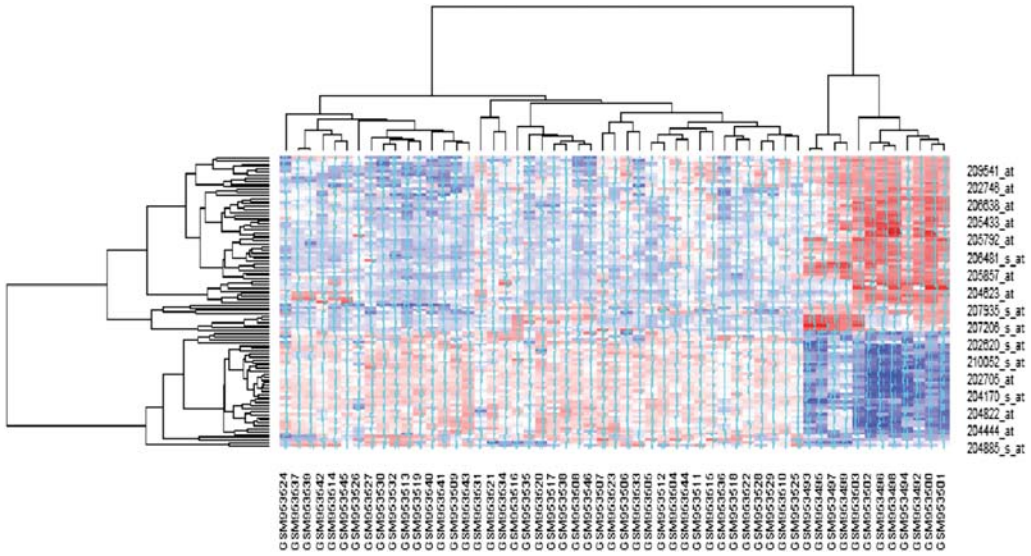


Fig. 3 Heatmap of DEGs

Machine Learning Model for Predicting Gene

Machine learning approaches were applied in order to identify the common optimum genes. This is also to fully utilize the measurements of different machine learning methods by collaborating them in selecting the significant genes. The machine learning methods chosen in this study are based on the literature review done which obtained good result in the related studies. Machine learning methods recursive partitioning, and Random Forest (RF) and support vector machine (SVM) which embeds variable selection as part of their learning process and evaluates them according to the classification result of specific machine learning methods. The results (selected genes) were determined and counted by the frequency of the genes that was being selected by machine learning methods.

Random Forest Prediction

Here, the down regulated genes were subjected for the prediction. In the training set 168 gene predictors from 42 samples were taken. Resampling was done with cross-validation of 50-fold and repeated one time. The results of resampling across tuning parameters were shown in Table 1.

Table 2 Random forest predicted genes for training set with accuracy

Probe set ID	Symbol	Accuracy
212186_at	ACACA	100
202338_at	TK1	84.6
219004_s_at	MIS18A	80.28
209520_s_at	NCBP1	79.01
204170_s_at	CKS2	78.12
219148_at	PBK	76.23
201930_at	MCM6	75.71
208079_s_at	AURKA	73.42
219649_at	ALG6	73.32
202954_at	UBE2C	73.2
207039_at	CDKN2A	70.51
204822_at	TTK	70.25
209642_at	BUB1	70.03
203418_at	CCNA2	69.79
205024_s_at	RAD51	69.34
204092_s_at	AURKA	69.33
204146_at	RAD51AP1	69.3
219429_at	FA2H	68.98
41037_at	TEAD4	66.95
209773_s_at	RRM2	63.25

Table 1 Training set

mtry	Accuracy	Kappa
2	0.9761905	0
168	0.9761905	0

Table 3 Test set

mtry	Accuracy	Kappa
2	0.96	0
168	0.96	0

Accuracy was used to select the optimal model using the largest value. The final value used for the model was mtry = 2. Only 20 most important variables shown (out of 168) in Table 2. In the test set, 168 gene predictors from 25 samples were taken. Resampling was done with cross-validation of 50-fold and repeated one time. Table 3 shows the outcomes of resampling across tuning parameters.

Accuracy was used to select the optimal model using the largest value. The final value used for the model was mtry = 2. Only 20 most important variables shown (out of 168) in Table 4.

Table 4 Random forest predicted genes for test set with accuracy

Probe set ID	Symbol	Accuracy
205046_at	CENPE	100
218350_s_at	GMNN	99.94
200613_at	AP2M1	96.8
200876_s_at	PSMB1	95.8
202338_at	TK1	94.67
202613_at	CTPS1	91.93
202983_at	HLTF	89.24
203576_at	BCAT2	84.77
41037_at	TEAD4	80.34
219306_at	KIF15	79.93
201066_at	CYC1	79.79
202954_at	UBE2C	79.25
204170_s_at	CKS2	78.64
208972_s_at	ATP5G1	77.93
203554_x_at	PTTG1	76.71
201897_s_at	CKS1B	76.51
201252_at	PSMC4	75.28
218355_at	KIF4A	74.2
205053_at	PRIM1	73.63
210125_s_at	BANF1	68.06

SVM Prediction

In the training set, 48 samples and their 168 genes were taken for the prediction. In SVM, the pre-processing was done with the parameter of centered (168), scaled (168). Resampling was cross-validated (20-fold, repeated 3 times). Tuning parameter 'C' was held constant at a value of 1. The predicted variables are tabulated in Table 5.

Compared to Random Forest, the significant accuracy was obtained for cancer gene prediction while using SVM method. The SVM model predicted genes were analyzed for their biological role in cervical cancer.

GSEA Analysis for Predicted Genes

The selected significant genes were subjected for Gene Set Enrichment analysis using GSEA. Selected genes using SVM model were interacted and involved in the

Table 5 SVM predicted genes with accuracy

Training set			Test set		
Probe Set ID	SYMBOL	Accuracy	Probe Set ID	SYMBOL	Accuracy
204023_at	RFC4	100	210125_s_at	BANF1	100
218009_s_at	PRC1	99.22	204170_s_at	CKS2	99.69
204170_s_at	CKS2	99.18	200876_s_at	PSMB1	99.43
200853_at	H2AFZ	99.06	202954_at	UBE2C	99.22
207039_at	CDKN2A	97.62	209520_s_at	NCBP1	99.21
203209_at	RFC5	97.03	201897_s_at	CKS1B	98.3
202107_s_at	MCM2	96.91	209773_s_at	RRM2	98.09
203755_at	BUB1B	96.7	202705_at	CCNB2	97.98
202705_at	CCNB2	95.91	201292_at	TOP2A	97.6
201930_at	MCM6	95.81	202983_at	HLTF	97.46
219148_at	PBK	95.78	202107_s_at	MCM2	97.29
204962_s_at	SLC35F6	95.72	200853_at	H2AFZ	96.94
218755_at	KIF20A	94.69	202870_s_at	CDC20	96.91
201897_s_at	CKS1B	94.05	218009_s_at	PRC1	96.55
208079_s_at	AURKA	93.8	204962_s_at	SLC35F6	96.44
202338_at	TK1	93.6	200052_s_at	ILF2	96.28
204709_s_at	KIF23	93.57	201136_at	PLP2	95.75
203213_at	CDK1	93.4	203213_at	CDK1	95.72
203418_at	CCNA2	92.74	202338_at	TK1	95.66
201202_at	PCNA	92.66	209714_s_at	CDKN3	95.22

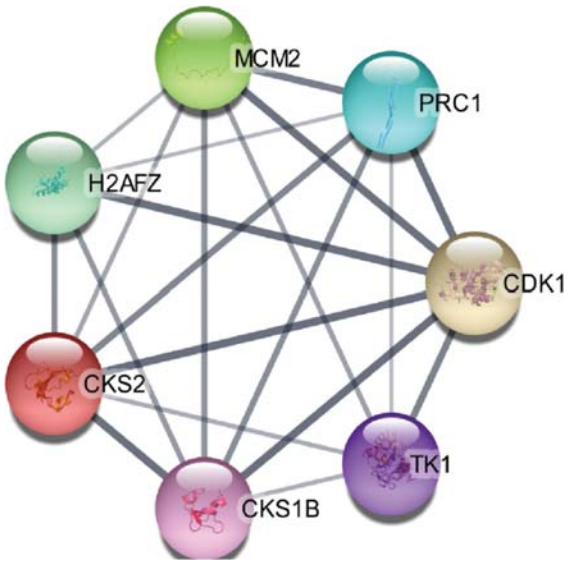


Fig. 4 Protein–protein interaction of predicted genes

following functional enrichments, namely, mitotic cell cycle phase transition, mitotic cell cycle process, cyclin-dependent protein serine/threonine kinase activator activity, cyclin-dependent protein kinase holoenzyme complex. The gene–gene interaction analyzed using cystoscope (Fig. 4).

Summary and Conclusion

Cancer is a disease with abnormal cell growth and uncontrolled multiplication of the cells within the body. Cervical cancer is the second most common cancer among Indian women accounts with new cases about 96,922 and deaths accounts for 60,078 per year in India. The management of cervical cancer represents a challenge due to its worse prognosis, non-developed treatments, and lack of targeted therapy. There are various treatment methods employed that are surgery, chemotherapy, and radiation therapy. Following surgery or radiotherapy, some patients experience a reoccurrence. Patients who relapse have few treatment options and a poor prognosis. Cervical cancer treatment has been found to be ineffective. As a result, new approaches to increasing the efficacy of cervical cancer treatment are urgently required. With this background this study aimed to develop a tool for predicting key genes from the gene expression dataset. Random forest model and SVM model were utilized for the prediction of the significant genes in the combined dataset (GSE75132 and GSE39001) which were involved in the prognosis of cervical cancer. Based on the accuracy obtained by this study, the SVM model was best for the prediction of gene expression microarray profile. The SVM model predicted genes were considered as a potential biomarker which were involved in the biological process. But this approach was carried out with two microarray profile. For the better understanding, *in-vitro* study should be carried out in future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40944-024-00831-y.

Acknowledgements This work is supported by Centre for Machine Learning and Intelligence No. DST – CURIE – AI- Phase II – Proj. 17., Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore – 43.

Declarations

Conflict of interest Nil.

References

1. Stanca M, Căpîlna ME. Prognostic factors associated with 5-year overall survival in cervical cancer patients treated with radical hysterectomy followed by adjuvant concurrent chemoradiation therapy at a tertiary care center in Eastern Europe. Diagnostics. 2021;11(3):570.
2. Dizon DS, Mackay HJ, Thomas GM, Werner TL, Kohn EC, Hess D, Rose PG, Covens AL. State of the science in cervical cancer: where we are today and where we need to go. Cancer. 2014;120(15):2282–8. https://doi.org/10.1002/cncr.28722.
3. Warnat P, Eils R, Brors B. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. BMC Bioinformatics. 2005;6:1–5.
4. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat Genet. 2001;29(4):365–71.
5. Brazma A, Vilo J. Gene expression data analysis. FEBS Lett. 2000;480(1):17–24.
6. Yang YH, Dudoit S, Luu P, Speed TP, Bittner ML, Chen Y, Dorsel AN, Dougherty ER. Microarrays: optical technologies and informatics. InSPIE, Bellingham, WA. 2001
7. Li J, Li YX, Li YY. Differential regulatory analysis based on coexpression network in cancer research. Biomed Res Int [Epub ahead of print]. 2016. https://doi.org/10.1155/2016/4241293.
8. Liu Y. Active learning with support vector machine applied to gene expression data for cancer classification. J Chem Inf Comput Sci. 2004;44(6):1936–41. https://doi.org/10.1021/ci049810a. (**PMID: 15554662**).
9. Njoroge E, Alty SR, Gani MR, Alkatib M. Classification of cervical cancer cells using FTIR data. Conf Proc IEEE Eng Med Biol Soc. 2006;2006:5338–41. https://doi.org/10.1109/IEMBS.2006.260024. (**PMID: 17945893**).
10. Cai S, Yu X, Gu Z, Yang Q, Wen B, Sheng J, Guan R. A 10-gene prognostic methylation signature for stage I-III cervical cancer. Arch Gynecol Obstet. 2020;301(5):1275–87. https://doi.org/10.1007/s00404-020-05524-3. (**Epub 2020 Apr 9 PMID: 32274635**).
11. Alles MC, Gardiner-Garden M, Nott DJ, Wang Y, Foekens JA, Sutherland RL, Musgrove EA, Ormandy CJ. Meta-analysis and gene set enrichment relative to er status reveal elevated activity of

MYC and E2F in the "basal" breast cancer subgroup. PLoS ONE. 2009;4(3): e4710.
12. Chow YP, Alias H, Jamal R. Meta-analysis of gene expression in relapsed childhood B-acute lymphoblastic leukemia. BMC Cancer. 2017;17:1.
13. Dong H, Hong S, Xu X, Xiao Y, Jin L, Xiong M. Meta-analysis and network analysis of five ovarian cancer gene expression dataset. In 2010 Third International Joint Conference on Computational Science and Optimization (Vol. 1, pp. 242–246). IEEE. 2010
14. Manawapat-Klopfer A, Thomsen LT, Martus P, Munk C, Russ R, Gmuender H, Frederiksen K, Haedicke-Jarboui J, Stubenrauch F, Kjaer SK, Iftner T. TMEM45A, SERPINB5 and p16INK4A transcript levels are predictive for development of high-grade cervical lesions. Am J Cancer Res. 2016;6(7):1524.
15. Espinosa AM, Alfaro A, Roman-Basaure E, Guardado-Estrada M, Palma Í, Serralde C, Medina I, Juárez E, Bermúdez M, Márquez E, Borges-Ibáñez M. Mitosis is a source of potential markers for screening and survival and therapeutic targets in cervical cancer. PLoS ONE. 2013;8(2): e55975.
16. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2012;41(D1):D991–5.
17. Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng. 2014;40(1):16–28.

Plagiarism Report

Avinashilingam Institute for Home Science and Higher Education for Women
(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category A by MHRD)
Re-accredited with 'A++' Grade by NAAC.CGPA 3.65/4, Category I by UGC
Coimbatore – 641 043, Tamil Nadu, India

PLAGIARISM CHECK REPORT (THESIS)

1.	Name of the Research Scholar	B. Sudha
2.	Roll No. and Year of Registration	19PHBTF001, 2019
3.	Department	Biotechnology
4.	Name of the Research Guide	Dr. S. Sumathi
5.	Title of the Thesis / Dissertation	Genomic and epidemiological profile of cervical cancer patients - identifying risk factors, pathways and novel variants through integrated survey and Whole Exome Sequencing strategies
6.	Similarity Content (%) Identified	**5%**
7.	Software Used	Turnitin
8.	Date of Verification	05-01-2025

Note : The report is excluding 14 Consecutive words, Review of Literature and Quoted Materials.

Checked by :

5/1/25
Information Scientist

B. Sudha
Research Scholar

05.01.2025
Assistant Librarian

5/1/2025
Research Guide

Date: 05-01-2025

Genomic and epidemiological profile of cervical cancer patients - identifying risk factors, pathways and novel variants through integrated survey and Whole Exome Sequencing strategies

ORIGINALITY REPORT

5%	2%	5%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Aliete Wan, Emily Place, Eric A. Pierce, Jason Comander. "Characterizing variants of unknown significance in rhodopsin: a functional genomics approach", Human Mutation, 2019 Publication	1%
2	link.springer.com Internet Source	1%
3	Balraj Sudha, Nachimuthu Senthil Kumar, Sundaravadivelu Sumathi. "Absence of Knowledge and Awareness About Cervical Cancer Among Educated Women: A Need for Education About Cervical Cancer", Indian Journal of Gynecologic Oncology, 2022 Publication	<1%
4	www2.mdpi.com Internet Source	<1%
5	www.ecco-org.eu	

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Central Library Avinashilingam
Assignment title:	Paper 2024
Submission title:	Genomic and epidemiological profile of cervical cancer patie...
File name:	Last_Sudha_Thesis_-_AI_checked_1.docx
File size:	5.58M
Page count:	134
Word count:	35,894
Character count:	213,097
Submission date:	05-Jan-2025 12:19PM (UTC+0530)
Submission ID:	2347454336

1. Introduction

Cervical cancer (CC) is a malignancy that originates at the cellular level within the epithelium of the cervix and progresses through a series of pre-cancerous stages before becoming invasive (Singh *et al.*, 2023). According to GLOBOCAN in 2022, cervical cancer is the fourth most common cancer in women around the world, with an estimated 660,000 new cases and 350,000 deaths annually (Bray *et al.*, 2024). Cervical cancer represents a significant global health challenge, especially for women in low- and middle-income countries (LMICs), where the disease burden is notably higher (Sudha *et al.*, 2025). This high mortality rate is particularly concerning because cervical cancer is largely preventable through screening and vaccinations. However, there is a stark contrast in outcomes between high-income countries (HICs) and LMICs, revealing an alarming disparity in cervical cancer prevention, diagnosis, and treatment options (Hull *et al.*, 2020).

In high-income countries (HICs), the widespread implementation of cervical cancer screening programs, including Pap smear tests and HPV DNA testing, combined with the introduction of the HPV vaccine, has significantly decreased the incidence and mortality of cervical cancer. Countries like the United States, Canada, and Western Europe have significantly declined cervical cancer rates due to these preventive measures (Siegel *et al.*, 2021). In contrast, LMICs face numerous barriers to effective cervical cancer prevention and control, leading to an overwhelming majority—nearly 90%—of cervical cancer deaths occurring in these regions (Maluf *et al.*, 2022; Duncan *et al.*, 2021). The lack of robust healthcare infrastructure, limited access to affordable and reliable screening programs, and insufficient availability of the HPV vaccine are significant challenges in these settings. In sub-Saharan Africa, cervical cancer is often diagnosed at advanced stages, where treatment options are scarce and survival rates are significantly lower (Perkins *et al.*, 2023). Even when screening is available, follow-up and treatment services may be inadequate, further contributing to high mortality rates.

The financial constraints of LMICs also limit the widespread implementation of HPV vaccination programs. Despite the proven efficacy of HPV vaccines in preventing infection with the high-risk HPV strains responsible for most cervical cancer cases, coverage remains low in resource-poor settings (Gültekin *et al.*, 2020). The Global Vaccine Alliance (GAVI) has made efforts to subsidize HPV vaccines in LMICs.