ACADEMIA ROMÂNĂ
Şcoala de Studii Avansate a Academiei Române
Institutul de Virusologie "Ștefan S Nicolau"

PhD THESIS SUMMARY

MOLECULAR MECHANISMS ASSOCIATED WITH HUMAN PAPILLOMAVIRUS ONCOPROTEINS IN CERVICAL CANCER

SCIENTIFIC COORDINATOR:
CS I, Dr. Gabriela ANTON

PhD STUDENT:
Alina FUDULU

2024

INTRODUCTION

Over the past two decades, it has become increasingly evident that viruses play a significant role in the development of cancers, with 15% to 20% of cancers being linked to viral infections (McLaughlin-Drubin ME, 2008). Among these viruses, the human papillomavirus (HPV) is notably implicated in cancer statistics, as it is identified as the causative agent of cervical cancer, being found in 99.7% of specific cyto/histological specimens (Walboomers JM, 1999). Within 12 to 24 months of exposure, 90% of genital HPV infections are either cleared or become inactive, while the remaining infections—associated with high-risk viral genotypes—persist and increase the risk of developing cervical intraepithelial neoplasia (CIN) and progressing to cervical cancer (Asiaf A, 2014). Epidemiological studies have established that HPV16 and HPV18 are present in 70% of cervical cancers, making these two genotypes the most prevalent in this disease.

HPV contributes to carcinogenesis primarily through the actions of two viral oncoproteins, E6 and E7. These proteins promote cellular transformation and immortalization by interfering with key tumor suppressor genes, p53 and retinoblastoma, leading to alterations in host DNA, changes in epigenetic mechanisms, and viral DNA methylation (Mittal S, 2017). The E6 and E7 oncoproteins are central to the oncogenesis of cervical cancer, as they drive the replication of the viral genome and induce all the hallmarks of cancer cells, including uncontrolled cell proliferation, angiogenesis, invasion, metastasis, and unrestricted telomerase activity, while also enabling the cells to evade apoptosis and disrupt tumor suppressor functions (Pal A, 2020).

One notable characteristic of cancer cells is the hypermethylation of CpG islands in gene promoters, particularly those of tumor suppressor genes, which leads to gene silencing. Additionally, methylated CpG islands can attract MBD family proteins, which are part of the NuRD chromatin remodeling complex, further contributing to repressive chromatin and gene silencing. Considering these factors, this doctoral thesis aims to explore the role of the viral oncoproteins E6 and E7 in cervical oncogenesis through chromatin remodeling, using an innovative approach. The study involves analyzing the binding patterns of NuRD complex components (MBD2 and MBD3) in the presence of active HPV16 E6 and E7 oncogenes and evaluating the potential for distinguishing HPV-induced lesions in the cervix based on the identified target genes. The research was conducted using modern molecular biology techniques, next-generation sequencing, bioinformatics, and signaling pathway analysis software.

THE SCOPE AND THE OBJECTIVES OF THE THESIS

The nucleosome remodeling and deacetylase complex (NuRD) is one of the four key ATP-dependent chromatin remodeling complexes found in cells. It plays a crucial role in regulating gene transcription, maintaining genome integrity, and controlling cell cycle progression. Like other chromatin remodelers such as the SWI/SNF and polycomb complexes, NuRD is also involved in regulating transcriptional processes that are essential to oncogenesis and cancer progression. Recent molecular insights into how NuRD is recruited to specific loci during development and how its activity is modulated in cancer highlight the potential role of its mislocalization in contributing to tumor biology (Basta J, 2015).

This study aims to

This PhD thesis aimed evaluate the genome-wide binding patterns of the NuRD complex components (MBD2, MBD3) in the presence of HPV16 E6/E7 oncogene activity using ChIP-seq techniques. The findings from this research may pave the way for new directions in both fundamental and applied research. The proposed objectives are as follows:

1. Development of an experimental model to silence the transcripts of the HPV16 E6 and E7 oncogenes;
2. Performing chromatin immunoprecipitation and sequencing (ChIP-Seq) on shRNA-treated samples and evaluating the identified targets;
3. Validating the targets identified in ChIP-Seq analysis in patient samples.

1. Development of an experimental model for silencing the transcripts of the E6 and E7 viral oncoproteins of HPV16

HPV viral oncoproteins E6 and E7 play a crucial role in oncogenesis, as their presence during viral genome replication can induce all the hallmark characteristics of cancer cells, including uncontrolled cell proliferation, angiogenesis, invasion, metastasis, unrestricted telomerase activity, evasion of apoptosis, and inhibition of growth suppressors (Yamato K, 2008). Given the complexity of HPV and its significant impact on public health, exploring the underlying mechanisms is essential. One promising approach is RNA interference, which has been recognized as a powerful method to inhibit viral gene expression through endonuclease cleavage and stable degradation of homologous mRNA. This technique offers a targeted and effective strategy for unravelling the mechanisms associated with HPV (Elbashir SM, 2001).

With these considerations in mind, this chapter presents two studies with the following objectives:

- Development of an in vitro experimental model to silence the viral E6 and E7 oncoproteins by targeting their transcripts using shRNA species;
- Evaluation of the model's efficiency in terms of expression, protein levels, cell cycle progression, and apoptosis.

1.1 Experimental model of *in vitro* silencing of E6 and E7 transcripts

Materials and methods

➢ *Design of HPV16 E6/E7 shRNAs and selection of shRNA sequences for silencing E6/E7 viral oncogene transcripts*. To generate high-efficiency shRNA sequences, Clontech software (In-Fusion Cloning Primer Design Tool v1.0, Takara, San Jose, CA, USA) was used, a specialized program that uses advanced algorithms to facilitate both sequence design, as well as their insertion into a plasmid vector. For inhibition of viral oncogene expression, five shRNA sequences targeting different regions of the HPV16 E6 or E7 transcripts were designed. (Table 1).

Table 1. Specific oligonucleotide sequences that targeted full-length HPV16 E6 and E7 transcripts.

shRNA	Targeted Nucleotide Sequence in Full-Length E6 and E7 HPV16 Transcript	shRNA Nucleotide Sequence (5'-3')
shARNE6_208	272 – 292 E6	GGGAATCCATATGCTGTATGT
shARNE6_439	189– 209 E6	AATGTGTGTACTGCAAGCAAC
shjRNA_23	503 – 523 E6^E7 joint	GGTCGATGTATGTCTTGTTGC
shARNE7_158	702-722 E7	GGACAGAGCCCATTACAATAT
shARNE7_165	709–729 E7	GCCCATTACAATATTGTAACC

➢ *Generation of constructs and transformation of competent E. coli cells, followed by insert analysis via Sanger sequencing:* Initially, shRNA oligonucleotides were designed to specifically target particular regions within the E6 and E7 gene transcripts. Subsequently, these oligonucleotides were incorporated into a plasmid vector (pENTR™/U6) designed for shRNA expression. Using Clontech software (In-Fusion Cloning Primer Design Tool v1.0, Takara), a specialized program with specific algorithms to build shRNA sequences presenting high efficiency and also facilitating their introduction into transfection (vector), specific shRNA sequences for the targeted genes (E7 or/and E6 HPV16) were designed and were subsequently synthesized by LifeTechnology (Waltham). To facilitate the transfection of CaSki cells, the custom-designed shRNA sequences were cloned using the BLOCK-iT U6 RNAi Entry Vector kit (LifeTechnology). The kit allows a fast and efficient cloning of specific shRNA molecules using the Gateway technology (Invitrogen), which is based on the bacteriophage λ property of site-specific recombination to introduce a DNA sequence of interest in a vector by enzymatic methods. Recombinant plasmid DNA was subsequently sequenced using the BigDye™ Terminator v3.1 Cycle Sequencing Reagent kit, and the results were analyzed using BioEdit v7.2 and NCBI Blast software (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

➢ *Transfection of CaSki cells with specific shRNAs to silence E6 and E7 viral oncogenes*: CaSki cells, which are immortalized with HPV 16, were transfected with five selected shRNA constructs (shE6RNA_208, shE6RNA_439, shE7RNA_158, shE7RNA_165, and shjRNA_23) at three different concentrations (4, 6, and 8 µg) using Lipofectamine 2000 (Invitrogen) and OptiMEM (Invitrogen).

Results

At this stage, we generated specific primers necessary to obtain oligo dc and confirmed obtaining oligo dc by electrophoresis, in 2% agarose gel (Figure 1).

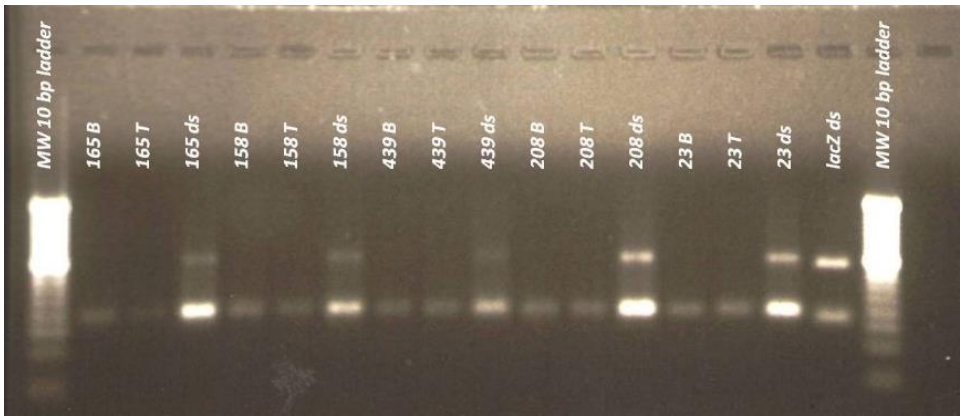

Fig 1. Agarose gel (4%) for highlighting dc oligonucleotide sequences

After vector ligation and transformation reactions, the resulting shRNA constructs were used to transiently transfect CaSki cells with Lipofectamine.

1.2 Evaluation of shRNA transfection efficiency

Materials and Methods

- *Evaluation of Transfection Efficiency at the mRNA Level by Real-Time PCR:* The efficiency of transfection for the five constructs was assessed by measuring the expression levels of HPV16 E6 and HPV16 E7 mRNA in transfected cell cultures compared to non-transfected controls at 24, respectively 48 hours post-transfection. The molecular techniques employed included: RNA isolation, cDNA synthesis, and Real-Time PCR. Specific primers were used to quantify E6 and E7 expression, and the transfection efficiency of the shRNA was calculated using the formula: % knockdown = 100 − 100 × 2−ΔΔCq E7/E6 HPV16; ΔCq = Cq E7/E6 HPV16 − Cq GAPDH; ΔΔCq = ΔCq E7/E6 HPV16 - ΔCq cell controls.
- *Evaluation of Transfection Efficiency at the Protein Level by Western Blot:* Protein extraction for Western blot analysis was performed on transfected and control CaSki cells using M-PER™ reagent (Thermo Scientific, USA). The primary antibodies used included anti-beta actin (ab8227), anti-human papillomavirus 16 (E7) [TVG 701Y] (ab20191), and anti-HPV16 E6 + HPV18 E6 [C1P5] (ab70). Detection was based on the chemiluminescence of the reporter conjugated to the secondary antibody. Results were interpreted through densitometric analysis using ImageJ software (National Institute of Health, USA).

➢ *Evaluation of the Impact of the Constructs on Apoptosis and Cell Cycle:* For cell cycle analysis, transfected and non-transfected cells were treated with propidium iodide and analysed by flow cytometry. The distribution of the cell cycle phases was analysed using ModFIT software (BD Biosciences, San Jose, CA, USA). Apoptosis was evaluated using the Annexin V-FITC detection kit (BD Biosciences, San Jose, CA, USA).

Result

Analysis of RT-PCR data revealed a variation in transfection efficiency in CaSki cells, with values ranging from 30.26% to 80.45%. The most efficient silencing was observed in CaSki cells transfected individually with shE6RNA_208, shE7RNA_158 and shjRNA_23, at a concentration of 6 µg, 24 hours post-transfection (Table 3).

Table 3. Optimization of the transfection protocol for the shRNAs used

	Concentration of shRNA and their efficiency (%)					
Concentration	**C_1 = 4 µg**		**C_2 = 6 µg**		**C_3 = 8 µg**	
Time of action	**24h**	**48h**	**24h**	**48h**	**24h**	**48h**
shE6ARN_208	68,83%	43,72%	80,45%	55,71%	75,01%	51,02%
shE6ARN_439	60,11%	41,85%	75,98%	42,45%	68,11%	44,25%
shjARN_23	74,53% E6 50,35% E7	50,45% E6 30,27% E7	79,51% E6 66,66% E7	43,18% E6 30,26% E7	63,04% E6 47,75% E7	69,34% E6 31,33% E7
shE7ARN_158	55,01%	43,80%	65,27%	49,83%	63,31%	39,81%
shE7ARN_165	53,9%	35,82%	55,81%	44,17%	59,81%	41,74%

Statistical analysis of E6 and E7 gene expression revealed a significant decrease in samples obtained from transfected cells compared to untreated ones at 24 h post-transfection. After 48 hours, the expression efficiency was considerably lower and the results became less significant. Relative expression and p-values are shown in Figure 2.

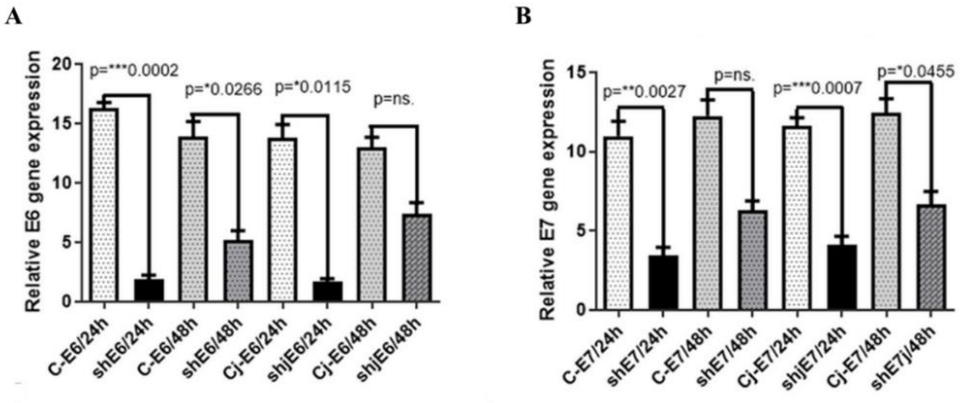

Fig 2. The impact of E6 and E7 shRNAs on the mRNA expression of HPV16 E6 and E7, assessed through RT-PCR for: HPV16 E6 (**A**) and HPV16 E7 after 24 and 48 h for treated and control (**C**) samples (**B**) *—$p < 0.05$; **—$p < 0.01$; ***—$p < 0.001$; ****—$p < 0.0001$; ns—$p > 0.05$.

Evaluation of E6 and E7 protein levels in transfected cells vs. control was performed by Western blot. Densitometric results showed that, 24 hours post-transfection, both oncoproteins were effectively blocked. For E6, statistically significant silencing was observed for both shE6RNA_208 (****p < 0.0001) and shjRNA_23 combination variant (p = 0.0269) compared to the untreated sample (Figure 3).

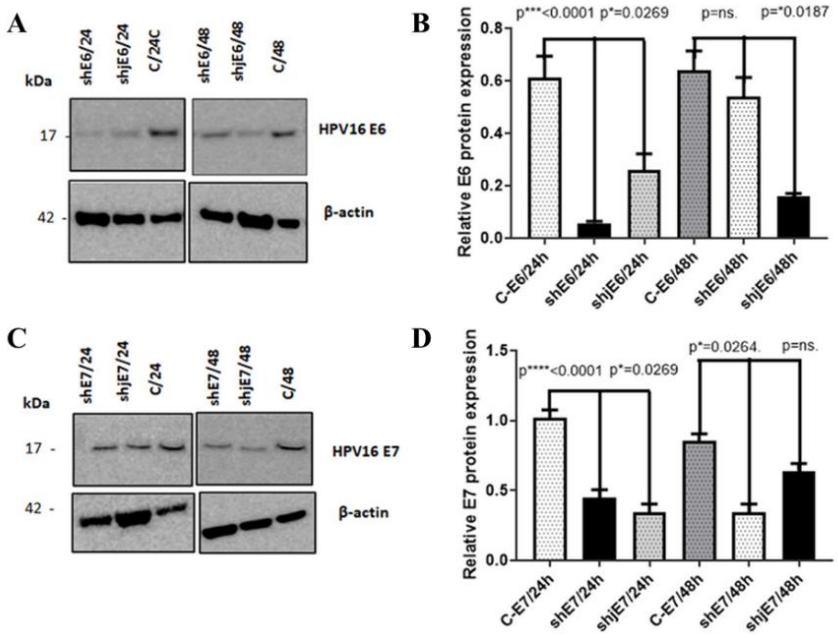

Fig 3. The effect of E6 and E7 shRNAs on protein expression analysed via Western blot analysis (WB) after 24 and 48 h for treated and control (C) samples for: E6HPV16 protein (**A**) and E7HPV16 protein (**C**), (Cj—control joint). Densitometry analysis for E6HPV16 protein (**B**) and E7HPV16 protein (**D**). *—$p < 0.05$; **—$p < 0.01$; ***—$p < 0.001$; ****—$p < 0.0001$; ns—$p > 0.05$.

The silencing of the two oncogenes caused slight changes in the G1 phase and G2+M phase, without influence on the S phase, compared to the untreated control sample at 24 h (shARNE6_208 - G: 62.6%; G2+M: 1, 8%; shARNE7_158: 0.3%; S: 55%; G2+M %; S: 34.2%). However, the significant changes were observed at 48 hours, through the simultaneous blocking of the two oncogenes. The results showed an increase in G1 phase

from 53.5% (Control) to 69.4%, respectively, a dramatic decrease in S phase from 46% (Control) to 17.5% (Table 4).

Table 4. Cell cycle in shRNA-treated and untreated CaSki cells.

shARN	24 hours			48 hours		
	%G1	%G2+M	% S	%G1	%G2+M	% S
shARNE6_208	62,6	1,8	35,5	53,1	0,3	46,6
shARNE7_158	62,6	0,3	37,1	56	4	40
shjRNA_23	57,3	2,5	40,2	69,4	17,5	14,1
Control	55	10,8	34,2	53,5	0,5	46

At 24 h post-transfection, individual oncogene silencing was noted to increase apoptosis in shARNE6_208 (16.6%) and shARNE7_158 (15.3%) compared to control (1.8%). CaSki cells treated for synergic inhibition of both E6 and E7 oncogenes (shjRNA_23) entered apoptosis in a percentage of 32.5% (Table 5).

Table 5. Distribution of total apoptosis in shRNA-treated and untreated CaSki cells.

shARN	Apoptoză totală (Q2+Q4)	
	24 ore	48 ore
shARNE6_208	1,7+ 14,9 = 16,6%	10,7 +42=52,7%
shARNE7_158	12,8+ 2,5=15,3%	24,2 +12,9=37,1%
shjARN_23	10,2+ 22,3=32,5%	19 +48,2=67,2%
Control	1,5+0,3=1,8%	8,5+1,3=9,8%

CONCLUSIONS

Three of the 5 shRNA-containing constructs (shE6ARN_208, shE7ARN_158 and E6^E7 joint shjARN_23) were validated for the presence, sequence and sense of the insert (Sanger sequencing), and showed increased efficiency in blocking the expression of oncogenes in transfected versus control CaSki cells, according to data obtained at testing by RT-PCR, Western blot and flow cytometry.

2. Chromatin immunoprecipitation and sequencing (ChIP-Seq). Evaluation of identified targets

2.1 Chromatin immunoprecipitation in the context of MBD2/MBD3, components of the chromatin remodeling complex (NuRD); Sequencing (ChIP-seq)

Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) is a method to investigate, at the genome level, the profile of DNA-binding proteins, histone or nucleosome modifications. Thanks to the progress in next-generation sequencing technology, ChIP-Seq offers higher resolution and coverage, becoming an indispensable tool for studying gene regulatory processes and implicitly epigenetic mechanisms. ChIP-Seq experiments generate large amounts of data, and efficient computational analysis is important for deciphering biological mechanisms (Park 2009).

The present study aims to evaluate by ChIP-seq the binding mode of NuRD complex components (MBD2, MBD3) at the genome level, in the presence/absence of HPV16 E6/E7 oncogene activity.

Materials and methods

- *Chromatin immunoprecipitation*. The study was performed on DNA extracted from CaSki cells transfected (4×10^7 cells for each immunoprecipitation) with the three constructs that demonstrated efficiency, at 24, respectively 48 hours post-transfection. Immunoprecipitation was performed using antibodies specific for MBD2 and MBD3 and the SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling Technology). For optimal results, 10 μg of digested and cross-linked DNA was used in the immunoprecipitation process. A positive control, 10 μL of Histone H3 (D2B12) XP Rabbit mAb #4620 and as a negative control, 2 μL of IgG Normal Rabbit #2729 were used. The study also included non-transfected CaSki cells (control).
- *ChIP-Seq sequencing*. This was performed on the Mi-Seq system (Illumina), following the TruSeq ChIP Library Preparation Kit protocol. Data processing was carried out with the help of bio-informatics tools such as: Bowtie v1.3.1, Picard v2.27.0 or MACS2 v2.2.7.1. Genome coverage for both treated and untreated samples was performed using the Easeq browser, and peak annotation was performed with HOMER.

Result

ChIP-Seq sequencing data indicated that shARNE6_208 treatment induces NuRD complex deposition mainly in the intergenic region (89%), shjRNA_23 in the intron region (57%), and shRNA E7 in the intergenic regions (40%), in the intron regions (30%), exon (20%), and along with TSS (10%) (Figure 4).

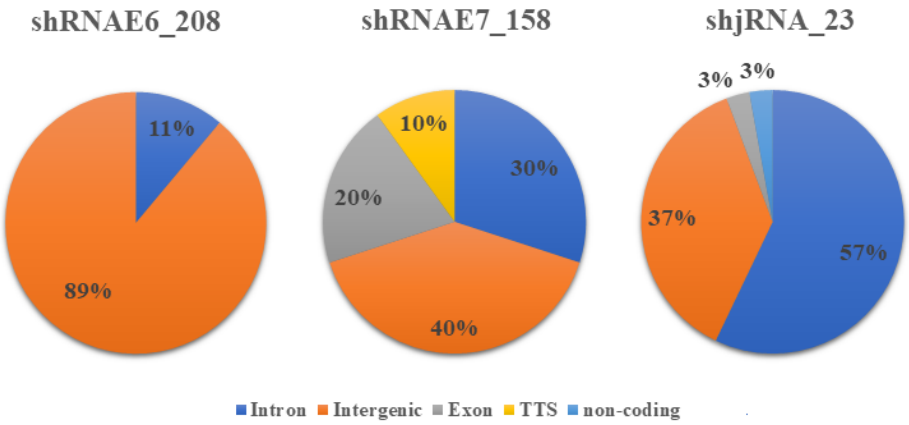

Fig 4. Peaks specific to the MBD2/MBB3-NuRD complex-enriched region.

The distribution of the distances from the peaks to the nearest TSS indicated that shjRNA_23 transfection showed the highest percentages at all investigated levels (TSS - 83%, +20 kB – 57%, - 20 kB – 60%) (Figure 5).

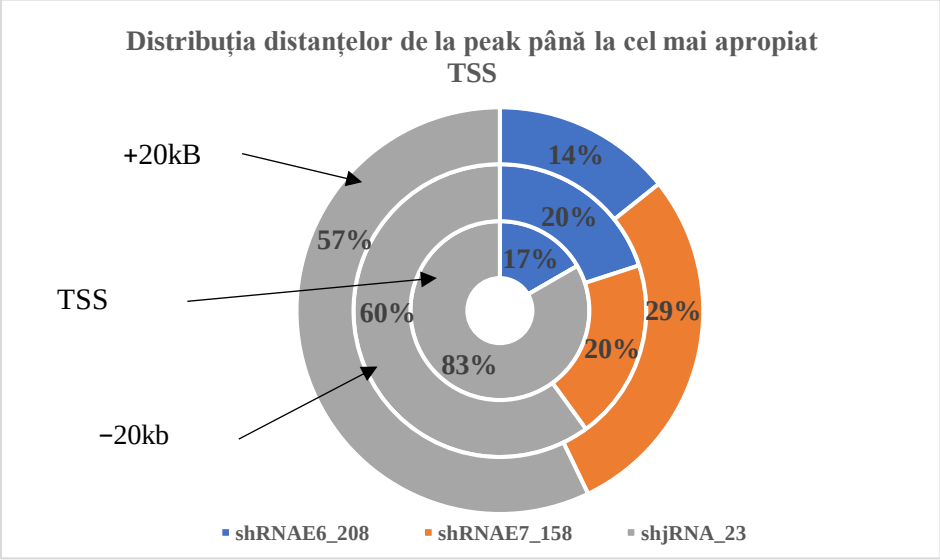

Fig 5. Distribution of the distances from the peaks to the nearest TSS in each sample

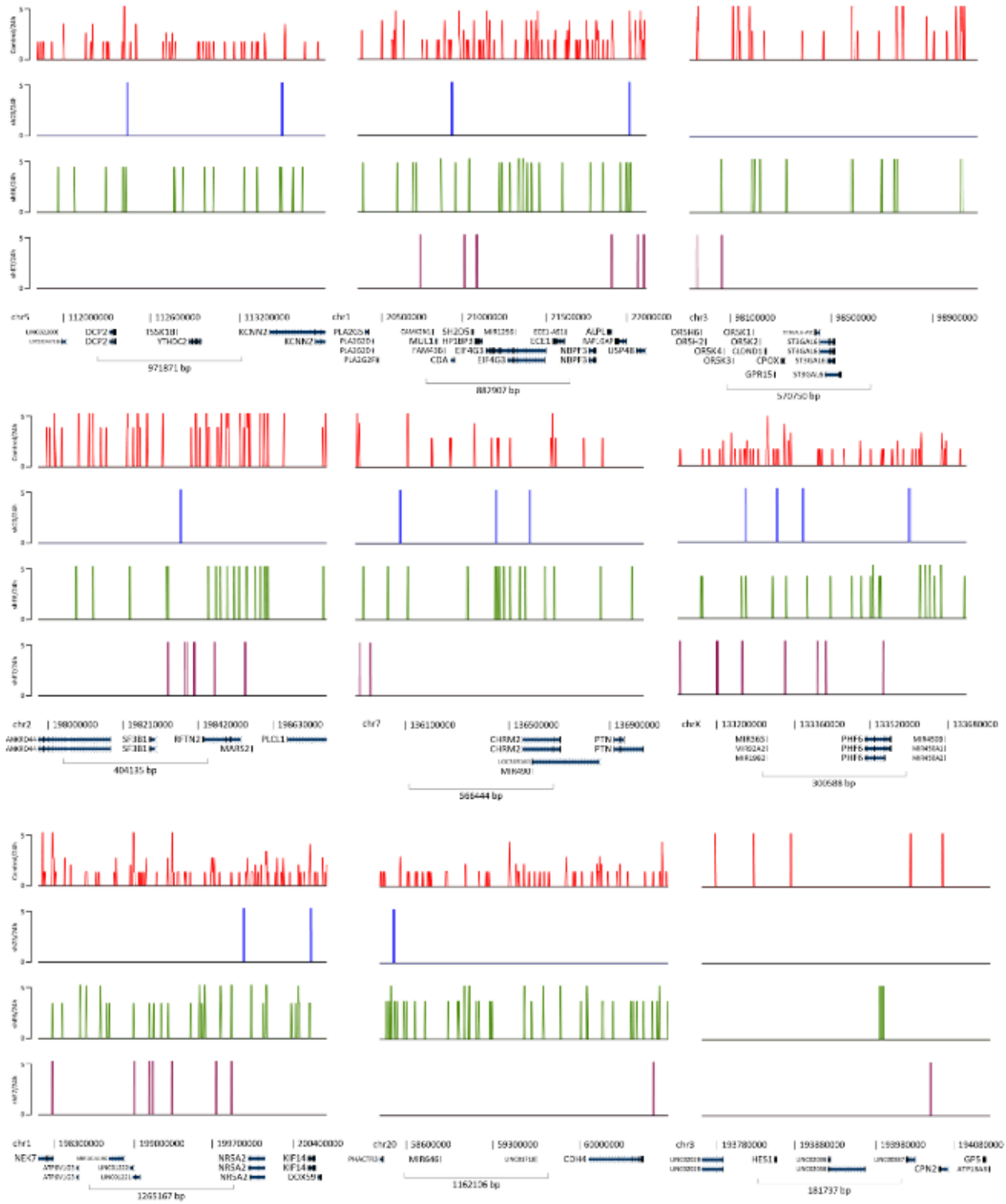

Fig 6. Distribution of peaks along chromosomes in CaSki transfected with shRNAE6_208, shRNAE7_158 and shjRNA_23 versus untreated CaSki cell line (control).

Annotation of peaks with nearby genes or shared genomic features, performed with the HOMER interpretation software, led to the identification of transcription factor (TF)-specific binding sites for 54 target genes involved in transcriptional regulation, cell

differentiation, proliferation, immune response, apoptosis, cell cycle and oxidative stress responses, indicating a much more subtle oncological process with a greater number of host genes affected by HPV16 E6 and E7 oncogenes.

The selected target genes were: PHF6, SF3B1, FGF16, MAGEB17, LINC02036, EIF4G3, LINC02720, EQTN, KCNJ3, LGMN, DDHD1, LINC01222, PSEN2, CHMR3, ZFPM2, THEMIS, CDK6, TRIM60, LINC00936, GIPC2, DCP2, DLC1, SEMA5A-AS1, CCDC138, NRIP1, ARHGEF28, CLEC16A, CDH17, ATXN10, EPHA3, LRRC4C, OR13F1, FAM71D, LOC101928446, PTGDR, GPR15, FABP6, MIR490, LRRC52, CHEK2P2, DSG2, KANK3, EPCAM-DT, STHG4, PROKR2, LINC01718, CCDC134, IL1RAP, FAT1, RHOH, MRPS30, TTC33, FAXC.

2.3 Evaluation of targeted genes in ChIP analysis by RT-PCR in CaSki cells transfected with shRNA

Materials and methods

- *Molecular techniques used*: RNA isolation, reverse transcription (cDNA synthesis) and Real-Time PCR. Expression evaluation was performed using specific primers for each of the 54 target genes, in transfected and non-transfected (control) samples at 24, respectively 48 hours after transfection.

Results

- PHF6, SF3B1, EIF4G3, DCP2, SEMA5A-AS1, CCDC138, NRIP, DSG2, FAT1 and FAM71D genes had the lowest expression when cells were transfected with shjRNA_23, compared to those transfected with shRNAE6_208 and shRNAE7_158, respectively, control at 24 hours after transfection. A decrease in expression due to E6 gene silencing was observed for LINC02036, LINC02720, LINC01222, LINC01718, miR-490, CHMR3, GPR15, OR13F1 and DSG2 genes, 24 h after transfection with shRNAE6_208. Other changes observed at 24 h post-transfection was increased expression of LGMN, CDK6, DCP2, SEMA5A-AS1, CLEC16A, EPCAM-DT, STHG4, MRPS30, and ATXN10 genes after transfection with shRNAE7_158. At 48 h post-transfection, increases in LINC02036, LINC02720, LINC01718, miR490, OR13F1, CHMR3, EPCAM-DT, GPR15 and CHEK2P2 gene expression were observed after transfection with shRNAE6_208 and shRNAE7_158 (Figure 6).

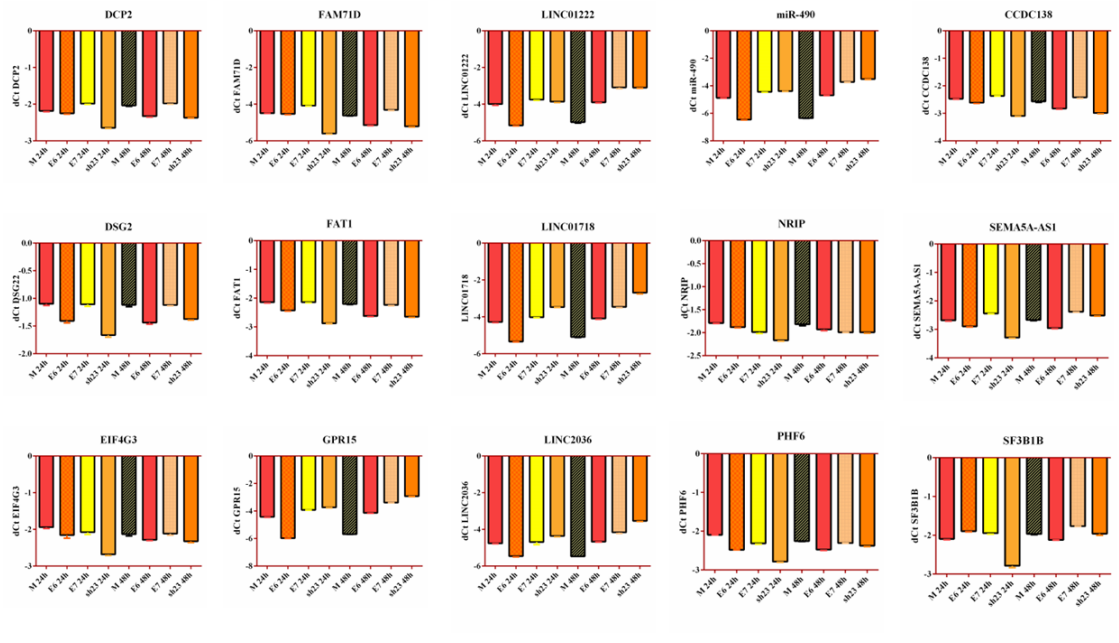

Fig 7. Expression profile of some investigated genes in cells treated with shRNA versus control, at 24 and 48 hours, respectively.

CONCLUSIONS

The RT-PCR analysis results confirm that the most significant changes in gene expression occurred 24 hours post-transfection. Furthermore, we observed the impact of oncogene inhibition, both individually—using shRNAE6_208 or shRNAE7_158—and synergistically through transfection with shjRNA_23.

The genes exhibiting the most notable changes (PHF6, SF3B1, EIF4G3, DCP2, GPR15, SEMA5AAS1, CCDC138, NRIP, DSG2, FAT1, FAM71D, ATXN10, miR-490, LINC01222, LINC01718, LINC02036) were selected for subsequent validation in patient samples to assess their potential in distinguishing between precursor lesions of cervical cancer.

3.1 Validation of target gene expression in patient samples

- The study consisted in evaluating the capacity of the genes selected by ChIP-seq to discriminate between cervical cancer precursor lesions, with the aim of evaluating their potential as diagnostic biomarkers;

Materials and methods

- *Sample collection*. The study included 74 vaginal lavage samples from women who presented for gynecological examinations in the Obstetrics and Gynecology Clinical Hospital "Cuza Vodă" in Iasi.
- *Molecular biology techniques used*: DNA isolation, HPV genotyping, RNA isolation and purification, cDNA synthesis and RT-PCR reaction.
- *Statistical analysis*. It was made with GraphPad Prism version 9.3 software. The results were presented as medians and further, we used one-way non-parametric ANOVA (Kruskal-Wallis), non-parametric (Mann-Whitney) test and paired t-test. Values of $p < 0.05$ were considered statistically significant.

Results

Patient samples were classified into seven groups according to cytology and the presence or absence of HPV infection as follows: NILM (Negative for Intraepithelial Lesion or Malignancy) positive and negative; ASCUS (Atypical squamous cells of undetermined significance); ASCH (Atypical Squamous Cells); LGSIL (Low-grade squamous intraepithelial lesion); HGSIL (High-grade squamous intraepithelial lesion); SCC (squamous cervical carcinoma) (Table 6).

Table 6. Classification of patient samples according to cytological analysis, HPV status and age.

	Total number of cases, n = 74 (100%)						
	NILM HPV-	**NILM HPV+**	**ASCUS HPV+**	**LGSIL HPV+**	**ASCH HPV+**	**HGSIL HPV+**	**SCC HPV+**
Number of cases	*n* = 11	*n* = 9	*n* = 12	*n* = 12	*n* = 11	*n* = 10	*n* = 9
(% from total number of cases)	(14.86%)	(12.16%)	(16.21%)	(16.21%)	(14.86%)	(13.51%)	(12.16%)
Age (Mean ± SD)	33.88 ± 4.224	36.22 ± 7.496	32.00 ± 6.663	31.45 ± 8.042	36.00 ± 8.790	37.13 ± 9.062	62.44 ± 14.910

HPV testing revealed that 85.13% (63 out of 74) of the samples were positive for the virus. Among NILM cases, HPV infection was found in 45% (9 out of 20) of the

samples. The HPV16 genotype had the highest prevalence, appearing in 52.38% (33 out of 63) of the positive samples. Other high-risk HPV genotypes identified included HPV45 in 12.70% (8 out of 63) of samples, and HPV18 and HPV33, each detected in 7.94% (5 out of 63). The remaining 12 samples contained high-risk genotypes, such as HPV51, HPV52, HPV53, HPV56, HPV58, and HPV66 (Figure 8).

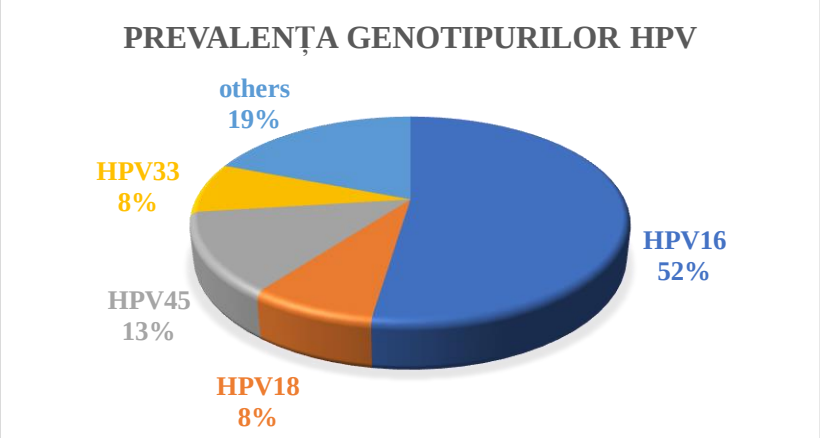

Fig 8. Prevalence of HPV genotypes in the study group

All nine selected genes showed altered expression profiles across the evaluated groups. The expression levels of miR-490, LINC01222, LINC01718, and LINC02036 were significantly elevated in all study groups compared to the control group. Specifically, the expression of LINC02036 was statistically higher in all groups relative to NILM (-) ($p < 0.0001$). In the case of miR-490, the highest expression was observed in the NILM (+) group (median = -0.036; $p < 0.0001$) compared to NILM (−) (median = -1.819), with the difference being statistically significant.

EIF4G3 and GPR15 also exhibited increased expression in all groups except for ASCUS and NILM (+). The SF3B1 gene showed significantly elevated expression in the LGSIL (median = -2.220; $p = 0.0113$), SCC (median = -2.820; $p = 0.0421$), and NILM(+) groups (median = -2.574; $p = 0.0005$) compared to the control group (median = −3.719).

A similar expression pattern was observed for the PHF6 gene in the LGSIL (median = -3.414; $p = 0.0169$) and SCC (median = -2.970; $p < 0.0001$) groups compared to the NILM (-) group (median = -4.451). For the DCP2 gene, significant changes were found in the ASCUS (median = -2.961; $p = 0.0207$), LGSIL (median = -2.595; $p = 0.0031$), and NILM (+) groups (median = -2.724; $p = 0.0127$). The remaining genes did not exhibit significant expression changes in the studied groups. Figure 9 illustrates the results for the genes with significantly altered expression.

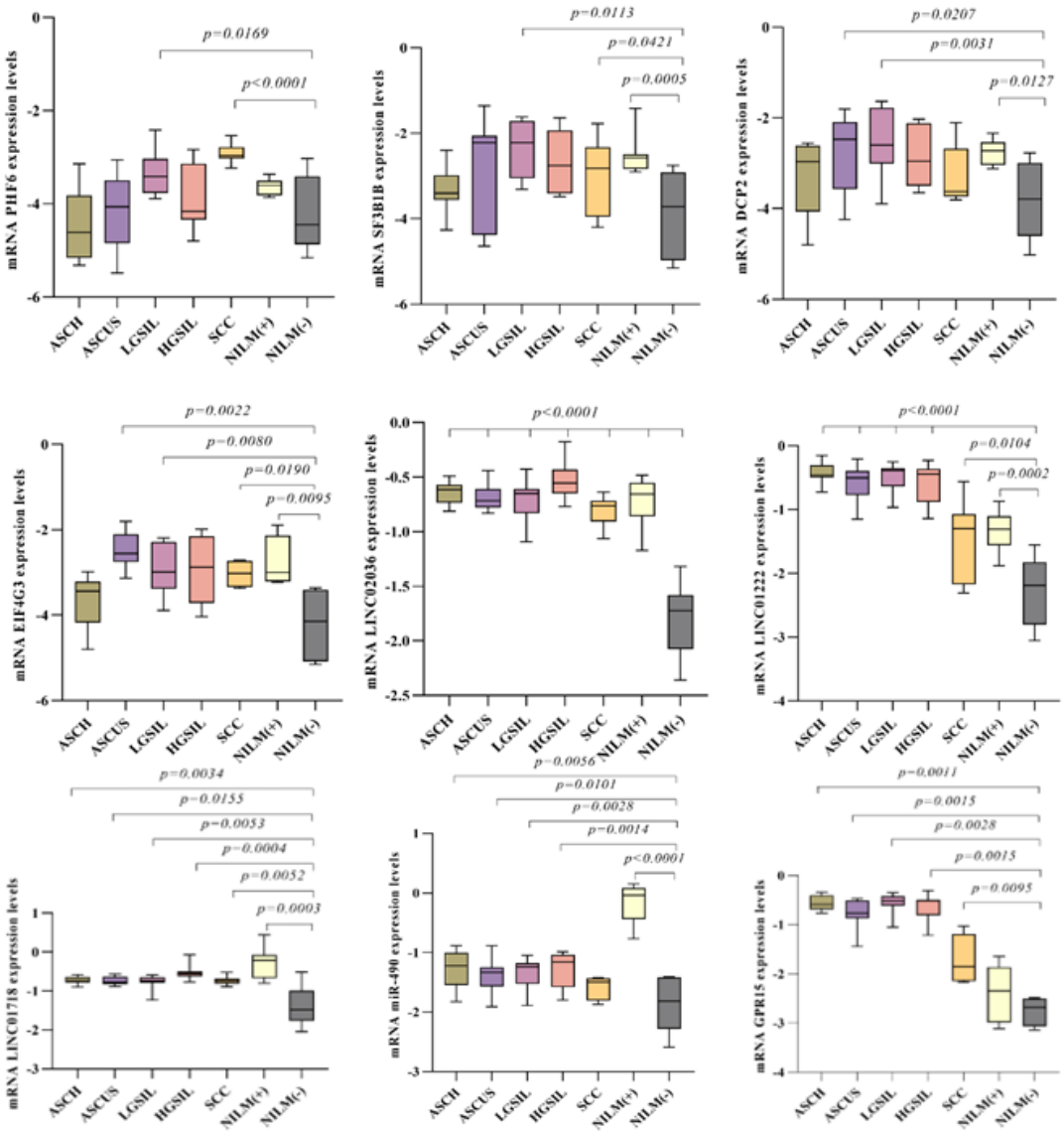

Fig 9. Expression levels of target genes significantly altered in patient samples: LINC02036, DCP2, EIF4G3, LINC01222, PHF6, SF3B1, LINC01718, GPR15 and miR-490.

ChIP analysis revealed that the transcriptional activity of several genes is modulated by the HPV16 E6/7 viral oncogenes through the MBD2/MBD3 NuRD complex. Targeting both viral oncogenes with shRNA_23 significantly affects gene transcription and enhances apoptosis. Moreover, inhibiting both E6 and E7 oncogenes results in a higher percentage of cells undergoing apoptosis compared to inhibiting either oncogene alone.

3.2 Signaling pathways involving target genes validated in patient samples

- This study aimed to investigate the molecular mechanisms by which the NuRD complex influences gene expression, implicitly about how these signaling pathways are modulated in the context of carcinogenesis.

Methods and materials

- *Analysis with Ingenuity Pathway Analysis (IPA) (Qiagen)*. The software was used for data mining and the generation of connectivity mapping between the most significant genes in patient samples, based on the manually curated publications stored in the QIAGEN Ingenuity Knowledge Base (QKB). The functions accessed during the analysis were: "*Grow*", "*Build*" - "*Grow to Diseases and Functions*", "*Build*" - "*Grow to Canonical Pathways*" and "*Build*" - "*Path Explorer*".

Result

To assess how the 9 target genes are involved in various signaling pathways, we constructed a network where we included them in our dataset list; it was noted that all genes were identified in the QKB database except LINCs. When we used the "*Grow to Disease and Functions*" function on the genes found in the QKB database, we noticed that they were associated with a network containing 163 diseases and their associated functions. Among the most important we list: cellular neoplasia, various types of cancer (bladder, bone marrow, breast, gastric, prostate) and viral infection, being the most important function (Figure 9).

Selecting the *'Grow to Canonical Pathways'* function and applying filters such as '*Cancer*', *'Infectious Disease'*, *'Reproductive Disease System'* and the species '*Human'* resulted in a dataset consisting of lists of molecules of various types, localization and functions, associated with target genes (Figure 10) (Table 7).

Table 7. Distribution of molecules associated with target genes according to location.

Location	SF3B1	DCP2	EIF4G3	MIR-490	PHF6	GPR15
nucleus	20	11	4	1	3	-
cytoplasm	18	6	4	3	-	3

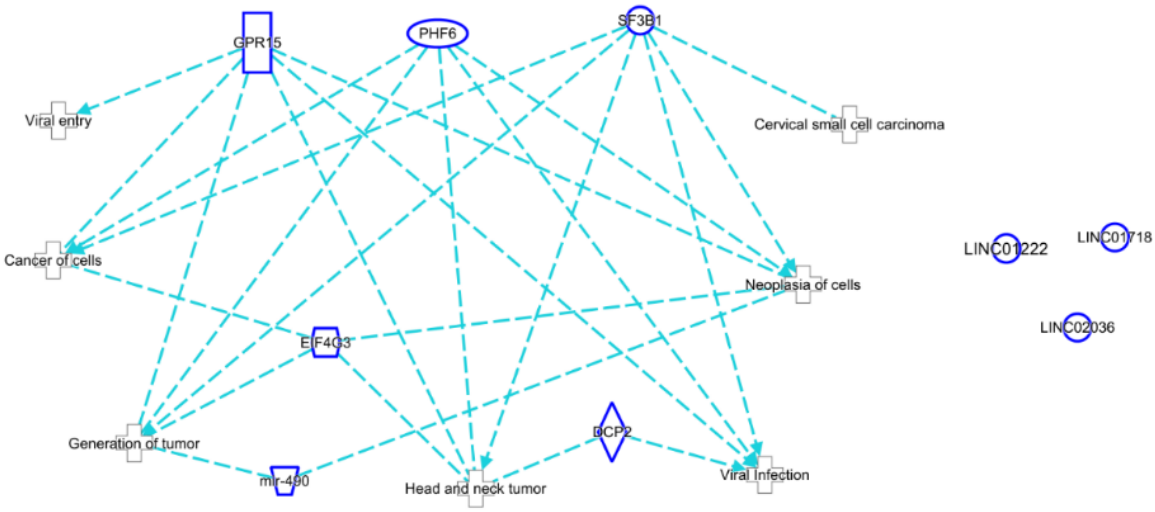

Fig 9. Graphic representation of the "Disease and functions" analysis for the selected genes

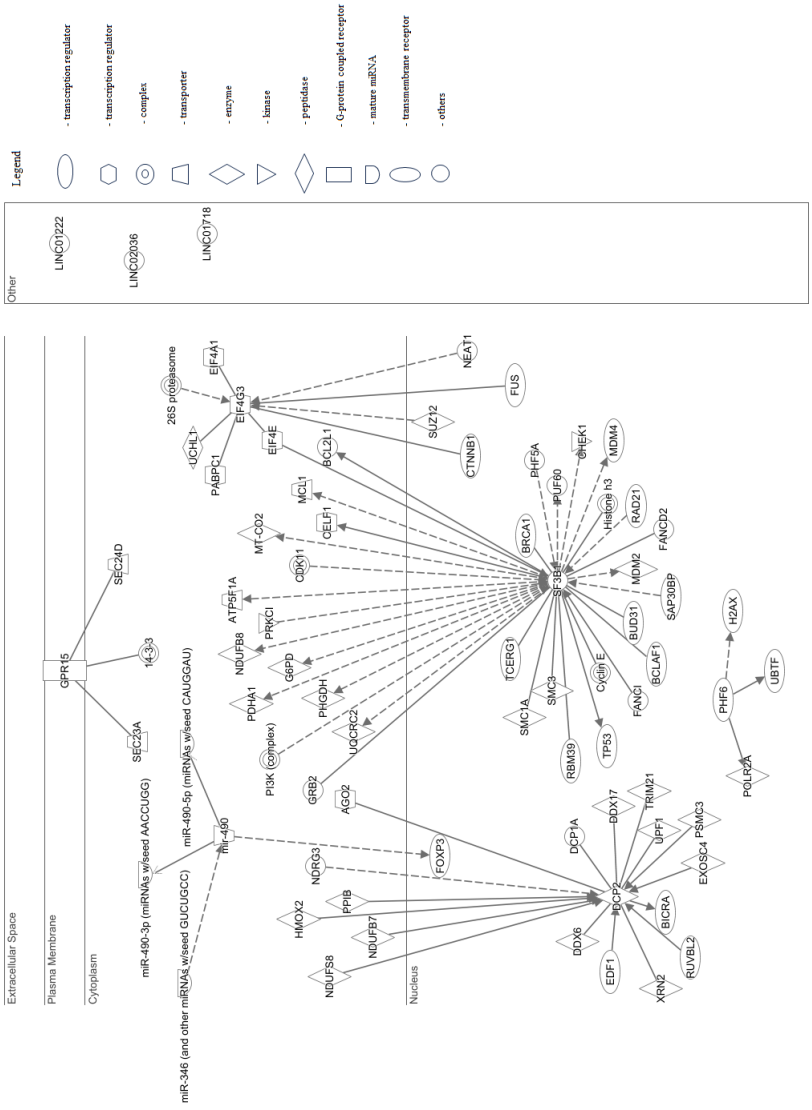

Fig 10. Localization of cellular compartments and molecular interaction between proteins encoded by selected genes

Using the "*Path Explorer*" function allowed the identification of the shortest paths between the inserted nodes. We used this tool to create custom, literature-supported signaling pathways starting from the same set of genes when we defined Set A and Set B in the setup menu. 265 pathways were obtained which were classified by relationship type into protein-protein interactions (n=187), protein-RNA interactions (n=7) and RNA-RNA interactions (targeting microRNA species) (n=71).

We designed a Venn diagram illustrating the number of common and different factors in the generated paths, and the results can be found in Figure 11.

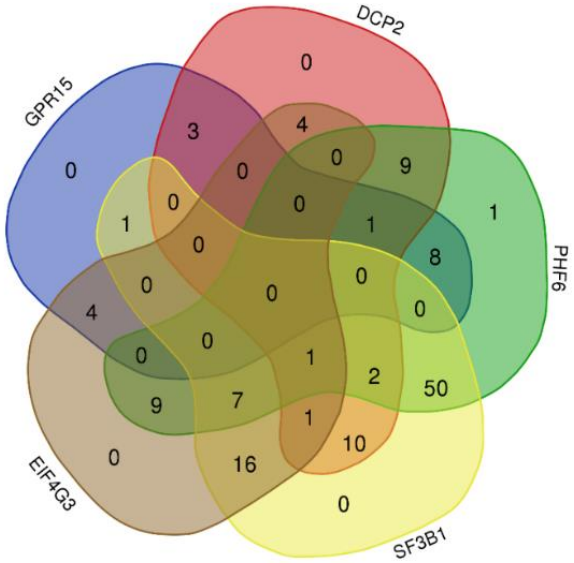

Fig 11. Venn diagram showing the overlap between molecules targeting DCP2 (30), EIF4G3 (42), PHF6 (88), SF3B1 (88), and GPR15 (17).

Finally, the NuRD complex was also added as a new molecule and the "*Path Explorer*" function was used to analyze the connections between Set A (NuRD) and Set B (SF3B1B, PHF6, EIF4G3, DCP2, GPR15, miR-490, LINC01222, LINC01718, LINC02036), using parameters for conditions: "*Cancer*", "*Infectious Disease*" and "*Reproductive Disease System*". The results revealed 9 nodes connecting the NuRD complex with three of the nine genes (SF3B1B, PHF6 and EIF4G3). The inclusion of these molecules led to the identification of 20 pathways in the network (Figure 12).

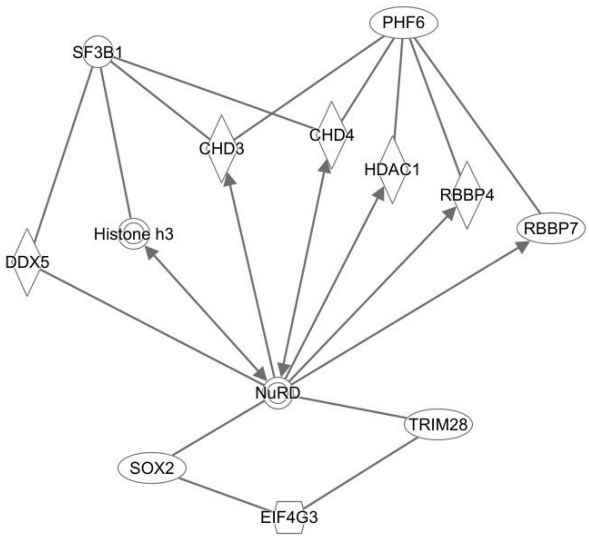

Fig 12. The main target molecules in the network formed by the NuRD complex and target genes

After the NuRD complex was added to the network, the MBD2 and MBD3 genes were also included. The "Grow to Canonical Pathways" function was then employed, using filters for "*Cancer*", "*Infectious Disease*", "*Reproductive Disease System*" and the species "*Human*". The results revealed various molecules of different types and functions associated with the target genes. It was found that the SF3B1, EIF4G3, and PHF6 genes share similar pathways that interact with the same factors, and all three genes also interact with both the NuRD complex and its components, MBD2 and MBD3.

3.3 Methylation of target gene promoters

The methylation status of target gene promoters was investigated based on the analysis of the signaling pathways in which they are involved and their associated molecules

Materials and methods

- *Investigation of CpG islands in gene promoters and primer design.* Primer design was conducted using MethPrimer software (http://www.urogene.org), where the promoter sequences of the three target genes (SF3B1, EIF4G3, and PHF6) were inputted.
- *Sample collection*. 62 patient samples were included in the study, classified as follows: LGSIL (n = 10), HGSIL (n = 10), ASCUS (n = 10), ASCH (n = 8), tissue

specimens from cervical squamous cell carcinomas (SCC) (n = 14) and control samples, NILM HPV negative lavages (n = 10).

Molecular techniques used: DNA and RNA isolation, cDNA synthesis, RT-PCR reaction, qMS-PCR reaction.

- *Conversion with bisulfite.* It was performed using the EpiTect Bisulfite kit (Qiagen). The working protocol includes: conversion of unmethylated cytosine with bisulfite; binding of converted single-stranded DNA to the column membrane; washing desulfonation of membrane-bound DNA; washing the DNA to remove the desulfonating agent; and elution of pure, converted DNA suitable for qMS-PCR.
- *Quantitative methylation-specific PCR (qMS-PCR).* Standard curves were generated using the positive (fully methylated) and negative (fully unmethylated) controls, both serially diluted to concentrations of 50 µg, 500 µg, 5 ng, and 50 ng. The reaction mixture consisted of: Maxima SYBR Green/ROX qPCR Master Mix (2X) (Thermo Fisher Scientific), 0.30 µM of each primer, and 50 ng of bisulfite-treated target DNA. The percentage of methylation (%M) was calculated according to the formula described by Fackler et al. (% methylation = 100 · [ng methylated gene A/(ng methylated gene A + ng unmethylated gene A)] (Fackler MJ, 2004). Unmethylated (U) and methylated (M) DNA concentrations for each sample were extrapolated using standard curves.

Results

Following qMS-PCR analysis, we found that only two of the three selected target genes showed visible and statistically significant changes (EIF4G3 and SF3B1). The highest methylation percentages were observed in control samples, with medians of 88.76% (range: 81.20–98.35%) for EIF4G3 and 87.53% (range: 78.95–98.22%) for SF3B1. In the EIF4G3 gene promoter, CpG islands showed increased methylation percentages in ASCUS and LGSIL patients, with means of 42.83% and 36.23%, respectively, but lower than in the control group. Samples from ASCH and HGSIL showed similar patterns with mean values of 13.22% and 15.21%, respectively, and methylation percentages ranged from 3.31–25.57% and 6.39–22.56 %. The lowest values were observed in SCC tissue samples with a median of 1.49% (Table 8).

Table 8. Statistical parameters of promoter methylation levels for EIF4G3 and SF3B1 genes in the studied groups.

Gene	*EIF4G3*		*SF3B1*	
***t*-test**	***Median* (*min; max*)**	***p-value***	***Median* (*min; max*)**	***p-value***
ASCUS	42.83 (29.65; 79.78)	< 0.0001	56.09 (38.61; 73.89)	< 0.0001
LGSIL	36.23 (18.19; 49.54)	< 0.0001	28.09 (15.97; 36.60)	< 0.0001
ASCH	15.21 (6.39; 22.56)	< 0.0001	11.87 (6.43; 16.50)	< 0.0001
HGSIL	13.22 (3.31; 25.57)	< 0.0001	6.28 (1.70; 26.96)	< 0.0001
SCC	1.49 (0.18; 7.52)	< 0.0001	0.75 (0.00; 3.08)	< 0.0001
NILM (-)	88.76 (81.20; 98.35)	-	87.52 (78.95; 98.22)	-

For the SF3B1 promoter, the lowest methylation values were found in SCC samples (0–3.08%), followed by slight increases in the HGSIL (range: 38.61–73.89%) and ASCH (6.43 -16.50%). Higher percentages were observed in LGSIL (15.97-36.60%) and ASCUS (38.61-73.89%) lesions. All results indicate a significant decrease in promoter methylation percentage in all studied groups compared to controls ($p < 0.0001$). The methylation profiles of EIF4G3 and SF3B1 in the studied groups are shown in Figure 13.

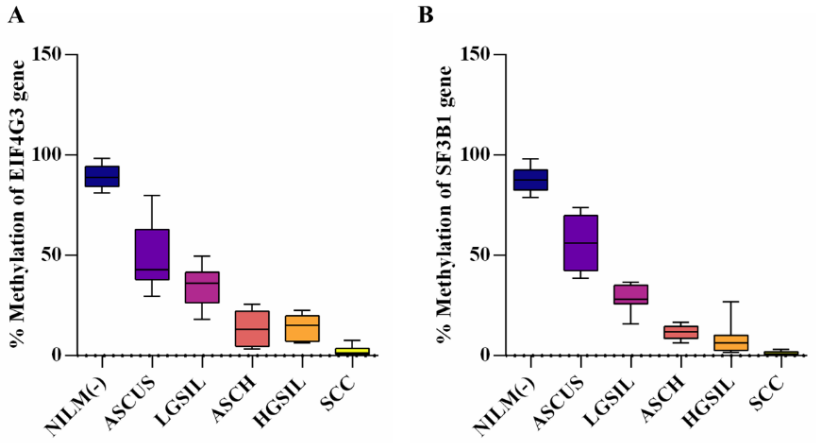

Fig 13. Methylation status profiles in all studied groups for both EIF4G3 (A) and SF3B1 (B) genes

For both genes, a comparison of methylation level between ASCUS and HGSIL showed a significant difference, with lower values observed in patients with HGSIL ($p < 0.0001$). In addition, comparison of the ASCH group with the SCC group revealed significant differences for both EIF4G3 ($p=0.0004$) and SF3B1 ($p <0.0001$). However, no statistical difference was found between the ASCH and HGSIL groups. Biopsy results showed that 60% of ASCH patients were diagnosed with CINIII, this group showing a similar percentage of methylation to HGSIL.

qRT-PCR results indicated that the expression levels of EIF4G3 and SF3B1 genes were significantly increased in all studied groups compared with the control group, except for the ASCH group. Moreover, the SCC group showed the most significant increases for both genes ($p < 0.0001$), with mean values of -2.227 and -1.540, respectively, compared to the NILM (-) group, which had mean values of -4.649 and -4.242.

In addition, significant results were observed for the EIF4G3 gene in the ASCUS group ($p < 0.0001$, median = -2.221) and for the SF3B1 gene in the LGSIL group ($p < 0.0001$, median = -2.220) compared to the control. All results and p-values are detailed in Table 9 and Figure 14.

Table 9. Statistical parameters of EIF4G3 and SF3B1 gene expression in the studied groups

Gene	***EIF4G3***		***SF3B1***	
***t*-test**	***Median (min; max)***	***p-value***	***Median (min; max)***	***p-value***
ASCH	-3.443 (-4.801; -2.983)	0.0545	-3.403 (-4.265; -2.400)	0.0831
ASCUS	-2.552 (-3.138; -1.805)	< 0.0001	-2.217 (-4.644; -1.353)	0.0021
LGSIL	-2.989 (-3.890; -2.191)	0.0011	-2.220 (-3.315; -1.619)	< 0.0001
HGSIL	-2.872 (-4.036; -1.981)	0.0021	-2.755 (-3.483; -1.637)	0.0005
SCC	-2.507 (-3.460; -1.523)	< 0.0001	-1.568 (-3.196; -0.863)	< 0.0001
NILM (-)	-4.146 (-5.125; -3.365)	-	-3.719 (-5.155; -2.747)	-

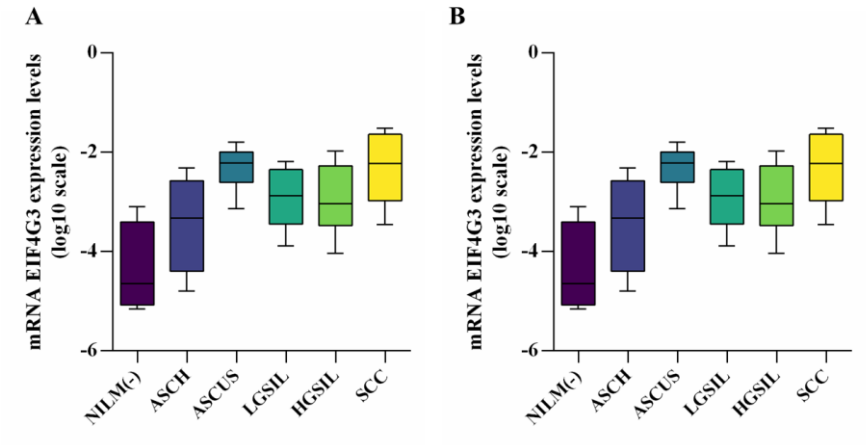

Fig 14. The mRNA expression levels of the EIF4G3 (A) and SF3B1 (B) gene promoters in the studied groups

Comparing the expression levels between the ASCH and SCC groups, there was a significantly increased expression of both genes in SCC patients ($p = 0.0102$ for EIF4G3 and $p < 0.0001$ for SF3B1). A significant increase in expression for the SF3B1 gene in the HGSIL group was observed when the ASCH and HGSIL groups were compared. In contrast, when we compared the ASCUS and HGSIL groups, only EIF4G3 expression levels were significantly increased in HGSIL patients ($p=0.0033$). We can say that the SF3B1 gene expression level could better discriminate between ASCH and SCC groups.

We also investigated the correlation between mRNA expression levels and methylation status for both genes and observed a significant inverse correlation. For EIF4G3, the correlation had a p-value of 0.0016 ($Y = -0.01252X - 2.543$) and for SF3B1, the p-value was less than 0.0001 ($Y = -0.01782X - 2.240$) (Fig 15).

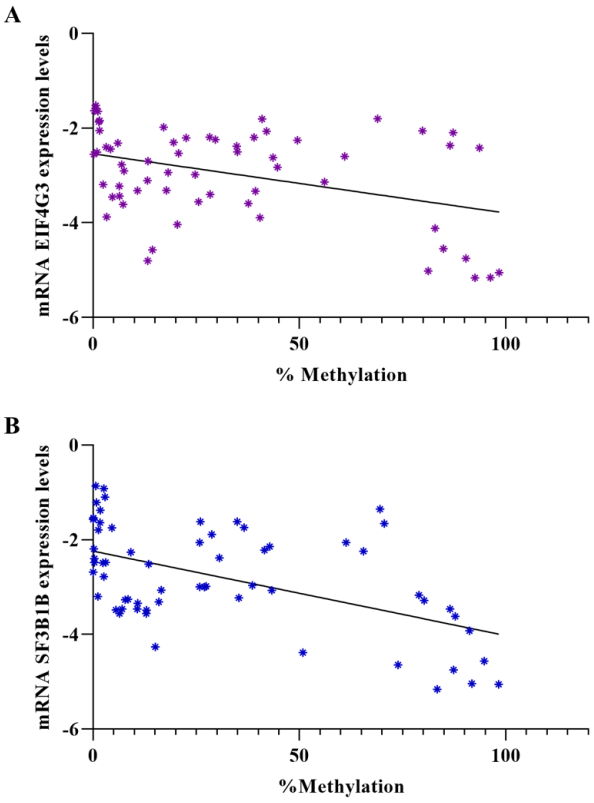

Fig 15. Correlation between the expression level and the methylation percentage of the EIF4G3 (A) and SF3B1 (B) gene promoters

CONCLUSIONS

Treatment with shRNAE6_208 causes a decrease in the expression of LINC02036, LINC01222, LINC01718 and miR-490 genes. The significantly increased level of expression of these genes in all HPV patient groups compared to the control group suggests their potential as markers for hrHPV infection. Although the function of the identified LINCs is not fully understood, this study is the first to highlight their involvement in cervical oncogenesis, opening new research directions.

The analysis of the signaling pathways associated with the target genes showed that they are involved in protein-protein, protein-RNA, and RNA-RNA interactions with the factors and molecules in the network. This also allowed analysis of common factors targeted by genes of interest. Thus, we noted that the most common factors are between PHF6 and SF3B1 (n=50), followed by EIF4G3 and SF3B1 (n=16). Moreover, the interaction between the three genes, PHF6, EIF4G3 and SF3B1, was also highlighted, they have 7 molecules in common.

Starting from this result, we analyzed the interaction between the target genes and the NuRD complex, where we noted that the same three genes have common factors with the complex: CHD3, CHD4, HDAC1, RBBP4 and RBBP7, which are actually components of the NuRD complex. Finally, we also included the two important components of the NuRD complex in the network: MBD2 and MBD3 and generated the common pathway analysis, where we confirmed once again that PHF6, EIF4G3 and SF3B1, target both components of the NuRD complex (MBD2 and MBD3), as well as the factors associated with them,

The study highlights the importance of epigenetic changes, particularly aberrant methylation, in the development of cervical cancer. Methylation is evidence that SF3B1 and EIF4G3 are not also subject to other epigenetic regulations, observing a decreased trend from NILM to SCC patients. Therefore, we can conclude that decreased methylation at the promoter level for both genes can be an indicator for abnormal cytology.

FINAL CONCLUSIONS

The investigation into the role of HPV16 E6 and E7 in cervical oncogenesis, utilizing ChIP-seq techniques to identify the binding patterns of NuRD complex components (MBD2, MBD3) in the presence of viral oncogenes, led to the following conclusions:

- Three of the 5 constructs containing shRNA (shE6RNA_208, shE7RNA_158, and E6^E7 joint shjRNA 23) validated for the presence, sequence and meaning of the insert (Sanger sequencing), showed increased efficiency in blocking the expression of oncogenes in transfected CaSki cells versus control, according to obtained data during testing by RT-PCR, Western blot and flow cytometry;

- ChIP-Seq sequencing data indicated that shARNE6_208 treatment causes deposition of the NuRD complex mainly in the intergenic region (89%), shjRNA_23 in the intron region (57%), and shRNA E7 in the intergenic regions (40%), in the intron regions (30%) exon (20%), along with TSS (10%);

- The distribution of the distances from the peaks to the nearest TSS indicated that shjRNA_23 transfection presented the highest percentages at all investigated levels (TSS - 83%, +20 kB – 57%, - 20 kB – 60%);

- Peak annotation using HOMER software identified specific transcription factor (TF) binding sites for 54 target genes involved in various cellular processes, including transcription regulation, cell differentiation, proliferation, immune response, apoptosis, cell cycle, and oxidative stress. This suggested a more nuanced oncological process with a broader impact on host genes by HPV16 E6 and E7. Of these, 31 genes with altered expression profiles in treatments versus controls were selected for validation in patient samples;

- In the patient cohort, 9 of the 31 selected target genes (PHF6, SF3B1, EIF4G3, DCP2, GPR15, SEMA5AAS1, CCDC138, NRIP, DSG2, FAT1, FAM71D, ATXN10, miR-490, LINC01222, LINC01718, LINC02036) exhibited statistically significant expression changes in precursor lesions and cervical cancer;

- IPA analysis confirmed for the first time that LINC01222, LINC01718, and LINC02036 are involved in cervical oncogenesis;

- IPA analysis also revealed that among the 9 target genes, PHF6, EIF4G3, and SF3B1 interact with the NuRD complex through factors CHD3, CHD4, HDAC1, RBBP4, and RBBP7, as well as with MBD2, MBD3, and associated molecular factors;

- The epigenetic analysis of PHF6, EIF4G3, and SF3B1 gene promoters in HPV-positive and -negative samples (including washes and tissue from cervical squamous carcinomas) showed statistically significant methylation changes associated with viral infection, specifically for EIF4G3 and SF3B1.

- Correlation analysis between methylation percentages and gene expression revealed statistically significant results for both EIF4G3 ($p = 0.0016$; $Y = -0.01252X - 2.543$) and SF3B1 ($p < 0.0001$; $Y = -0.01782X - 2.240$), indicating the involvement of these genes in cervical oncogenesis.

SELECTIVE BIBLIOGRAPHY:

Asiaf A, Ahmad ST, Mohammad SO, Zargar MA. 2014. „Review of the current knowledge on the epidemiology, pathogenesis, and prevention of human papillomavirus infection. .” *Eur J Cancer Prev.* 23 (3): 206-224.

Basta J, Rauchman M. 2015. „The nucleosome remodeling and deacetylase complex in development and disease.” *Transl Res.* 165 (1): 36-47.

Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. 2001. „Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. .” *EMBO J.* 20 (23): 6877-6888.

Fackler MJ, McVeigh M, Mehrotra J, Blum MA, Lange J, Lapides A, Garrett E, Argani P, Sukumar S. 2004. „Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer.” *Cancer Res.* 64 (13): 4442-4452.

McLaughlin-Drubin ME, Munger K. 2008. „Viruses associated with human cancer.” *Biochim Biophys Acta.* 1782 (3): 127-150.

Mittal S, Banks L. 2017. „Molecular mechanisms underlying human papillomavirus E6 and E7 oncoprotein-induced cell transformation.” *Mutat Res Rev Mutat Res.* 772: 23-35.

Pal A, Kundu R. 2020. „Human Papillomavirus E6 and E7: The Cervical Cancer Hallmarks and Targets for Therapy.” *Front Microbiol.* 10: 3116.

Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Muñoz N. 1999. „Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. .” *J Pathol.* 189 (1): 12-9.

Yamato K, Yamada T, Kizaki M, Ui-Tei K, Natori Y, Fujino M, Nishihara T, Ikeda Y, Nasu Y, Saigo K, Yoshinouchi M. 2008. „New highly potent and specific E6 and E7 siRNAs for treatment of HPV16 positive cervical cancer.” *Cancer Gene Ther.* 15 (3): 140-153.