Вавиловский журнал генетики и селекции
Vavilov Journal of Genetics and Breeding. 2024;28(3):342-350

DOI 10.18699/vjgb-24-39

mRNA-lncRNA gene expression signature in HPV-associated neoplasia and cervical cancer

E.D. Kulaeva, E.S. Muzlaeva, E.V. Mashkina

Southern Federal University, Rostov-on-Don, Russia
ked05685@gmail.com

Abstract. Cervical cancer is one of the most frequent cancers in women and is associated with human papillomavirus (HPV) in 70 % of cases. Cervical cancer occurs because of progression of low-differentiated cervical intraepithelial neoplasia through grade 2 and 3 lesions. Along with the protein-coding genes, long noncoding RNAs (lncRNAs) play an important role in the development of malignant cell transformation. Although human papillomavirus is widespread, there is currently no well-characterized transcriptomic signature to predict whether this tumor will develop in the presence of HPV-associated neoplastic changes in the cervical epithelium. Changes in gene activity in tumors reflect the biological diversity of cellular phenotype and physiological functions and can be an important diagnostic marker. We performed comparative transcriptome analysis using open RNA sequencing data to assess differentially expressed genes between normal tissue, neoplastic epithelium, and cervical cancer. Raw data were preprocessed using the Galaxy platform. Batch effect correction, identification of differentially expressed genes, and gene set enrichment analysis (GSEA) were performed using R programming language packages. Subcellular localization of lncRNA was analyzed using Locate-R and iLoc-LncRNA 2.0 web services. 1,572 differentially expressed genes (DEGs) were recorded in the "cancer *vs*. control" comparison, and 1,260 DEGs were recorded in the "cancer *vs*. neoplasia" comparison. Only two genes were observed to be differentially expressed in the "neoplasia *vs*. control" comparison. The search for common genes among the most strongly differentially expressed genes among all comparison groups resulted in the identification of an expression signature consisting of the *CCL20*, *CDKN2A*, *CTCFL*, *piR-55219*, *TRH*, *SLC27A6* and *EPHA5* genes. The transcription level of the *CCL20* and *CDKN2A* genes becomes increased at the stage of neoplastic epithelial changes and stays so in cervical cancer. Validation on an independent microarray dataset showed that the differential expression patterns of the *CDKN2A* and *SLC27A6* genes were conserved in the respective gene expression comparisons between groups.
Key words: human papillomavirus; neoplasia; cervical cancer; transcriptome analysis; lncRNA; *CDKN2A*; *CCL20*.

For citation: Kulaeva E.D., Muzlaeva E.S., Mashkina E.V. mRNA-lncRNA gene expression signature in HPV-associated neoplasia and cervical cancer. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding*. 2024;28(3): 342-350. DOI 10.18699/vjgb-24-39

Funding. The study was carried out with the financial support of the Ministry of Science and Higher Education of Russian Federation within the state assignment framework in the field of scientific activity No. FENW-2023-0018.

Профиль экспрессии мРНК-днРНК при неоплазии и цервикальном раке, ассоциированными с ВПЧ-инфекцией

Е.Д. Кулаева, Е.С. Музлаева, Е.В. Машкина

Южный федеральный университет, Ростов-на-Дону, Россия
ked05685@gmail.com

Аннотация. Рак шейки матки является одним из наиболее частых онкологических заболеваний у женщин и в 70 % случаев связан с вирусом папилломы человека (ВПЧ). Рак шейки матки развивается в результате прогрессии цервикальной интраэпителиальной неоплазии через поражения второй и третьей степени. Помимо белок-кодирующих генов, важную роль в развитии злокачественной трансформации клеток играют длинные некодирующие РНК. Хотя вирус папилломы человека широко распространен, в настоящее время нет хорошо охарактеризованных транскриптомных признаков, позволяющих предсказать злокачественную трансформацию клеток эпителия при наличии связанной с ВПЧ неоплазии эпителия шейки матки. Изменения генной активности в опухолях отражают биологическое разнообразие клеточного фенотипа и физиологических функций и могут быть важным диагностическим маркером. Используя открытые данные секвенирования РНК, мы провели сравнительный анализ транскриптома для оценки дифференциально экспрессируемых генов в образцах нормальной ткани, эпителия с диспластическими изменениями и раком шейки матки. Первичные данные были предварительно обработаны с использованием платформы Galaxy. Коррекция пакетного эффекта, идентификация дифференциально экспрессируемых генов и анализ обогащения набора генов выполнены в пакетах языка программирования R. Субклеточная локализация днРНК была проанализирована с помощью веб-сервисов Locate-R и iLoc-LncRNA 2.0. В сравнении «рак *vs*. контроль» зарегистрировано 1572 дифференциально экспрессируемых гена, в сравнении «рак *vs*. неоплазия» – 1260. Только два дифференциально экспрессируемых гена выявлено при сравнении контро-

ля и неоплазии. Поиск общих среди наиболее сильно дифференциально экспрессируемых генов во всех группах сравнения привел к выявлению сигнатуры экспрессии, состоящей из генов *CCL20*, *CDKN2A*, *CTCFL*, *piR-55219*, *TRH*, *SLC27A6* и *EPHA5*. Повышенный уровень транскрипции генов *CCL20* и *CDKN2A* возникает на стадии неопластических изменений эпителия и сохраняется при раке шейки матки. Валидация на независимом наборе данных микрочипа показала, что паттерны дифференциальной экспрессии генов *CDKN2A* и *SLC27A6* сохраняются в соответствующих сравнениях экспрессии генов между группами.
Ключевые слова: вирус папилломы человека; неоплазия; рак шейки матки; транскриптомный анализ; lncRNA; *CDKN2A*; *CCL20*.

Introduction

Cervical cancer is the fourth most common cancer in women worldwide after breast cancer, colorectal cancer, and lung cancer. The World Health Organization (WHO) estimates that 604,127 new cases and 341,831 deaths from the disease worldwide were registered in 2020 (Sung et al., 2021; Gebrie, 2022). Cervical cancer occurs as a result of progression of low-differentiated cervical intraepithelial neoplasia (CIN1) through grade 2 and 3 lesions (CIN2 and CIN3). Inflammatory responses are rarely observed in persistent low-grade lesions and are thought to be due to the inflammation-suppressing activity of high-risk HPV oncoproteins (Walch-Ruckheim et al., 2015).

Although HPV is the most significant factor in cervical cancer, the development of cervical cancer is considered multifactorial. Common risk factors for cervical cancer also include smoking, a high number of sexual partners, low social and/or economic status and its consequences, and immune suppression caused by infection such as human immunodeficiency virus (HIV) or the use of immunosuppressants after organ transplantation (Walch-Ruckheim et al., 2015).

According to the International Human Papillomavirus Reference Center data (Eklund et al., 2020), only 12 out of 220 HPV strains have the greatest impact on cancer development (these strains include HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59). About 70 % of cervical cancer and precancerous lesions of the cervix cases are specifically associated with HPV types 16 and 18 (Okunade, 2020).

Eighty percent of sexually active women become infected with HPV during their lifetime, but the infection persists in only 5–10 % of those initially infected and leads to cervical cancer in only 3 % (Schubert et al., 2023). In the absence of a clearly persistent HPV infection, the risk of developing cervical cancer is extremely low. However, virus persistence may be associated with many factors. Host genetic factors are thought to play an important role in the response to HPV infection and further development of oncology.

Along with the protein-coding genes, long noncoding RNAs (lncRNAs) play an important role in the development of malignant cell transformation. Results of the TCGA project showed that approximately the same number of protein-coding genes and lncRNA genes carried mutations in more than 5,000 different tumor samples. However, at the same time, 60 % of lncRNAs showed tumor type specificity and are superior to protein-coding genes in terms of specificity to the type of cancer (Yan et al., 2015). Consequently, lncRNAs can be a good class of biomarkers for cancer prognosis and early diagnosis.

Both protein-coding and lncRNAs can be analyzed as efficiently as possible by high-throughput RNA sequencing (RNA-seq). Profiling the entire transcriptome can identify genes that are differentially expressed in related tissues. Changes in gene activity in tumors reflect the biological diversity of cellular phenotype and physiological functions and can be an important diagnostic marker (Martin, Wang, 2011; Bao et al., 2019). A significant change in the expression of both protein-coding and non-coding parts of the genome may be a consequence of local chromatin remodeling in the region of the virus integration site, which plays a somewhat spontaneous but often important role in oncogenesis (Karimzadeh et al., 2023).

The aim of this work was to perform bioinformatics analysis of RNA sequencing data from epitheliocytes of women with cervical epithelial neoplasia and cervical cancer based on open data from three different studies (Royse et al., 2014; Hu et al., 2015; Qi et al., 2022).

Materials and methods

Datasets. The study material was raw RNA sequencing data of cervical epithelial samples from three separate studies analyzing the transcriptome in cervical cancer, neoplasia, and normal tissue. Neoplasia grade data were also available. The main characteristics of the studies used are summarized in Table 1.

Data preprocessing. Raw RNA-seq data (fastq format) were processed using the Galaxy platform (https://usegalaxy.org/). Read quality was assessed with FastQC, adapter trimming was performed with TrimGalore, transcript alignment and mapping was performed with RNA STAR, and transcript counting was performed with featureCounts, respectively.

Data variability analysis and batch effect correction. Analysis of data variability and assessment of the batch effect (effect of the subsample/sequencing platform rather than biological variability) were performed using principal component analysis (PCA) with the plotPCA function of the DESeq2 v.1.42.0 package for R. Based on the results of the variability analysis, a conclusion was made about the inclusion/exclusion of samples in further analysis.

Differential gene expression analysis was performed using the DESeq2 package in R. Genes were filtered by $\log_2$FC >2, $\log_2$FC < (–2), and adjusted p-value < 0.05 (as visualized in the R package EnchancedVolcano). Genes encoding mRNAs and lncRNAs were categorized using Ensembl Ids. To identify differentially expressed genes (DEGs) between three biological states (neoplasia *vs.* control; cancer *vs.* neoplasia; cancer *vs.* control), comparisons were performed and the top 10 genes with statistically significant increased and decreased expression were identified. Heatmap for common DEGs was plotted with the pheatmap package in R.

Gene set enrichment analysis (GSEA) to estimate activated and repressed biological pathways in the comparison groups was performed using the clusterProfiler v.4.10.0 package for R.

Table 1. Main characteristics of the used studies

No.	Study ID	Samples count			Sample type	Reference
		Control	Neoplasia	Cervical cancer		
1	SRP048735	6	12	–	Cervical biopsy (FFPE)	Royse et al., 2014
2	SRA189004	–	–	7	Cervical smear; cell lines	Hu et al., 2015
3	GSE149763	3	3	3	Cervical biopsy (FF)	Qi et al., 2022

lncRNAs subcellular localization analysis. The web services Locate-R (Ahmad et al., 2020) and iLoc-LncRNA 2.0 (Su et al., 2018) were used to determine the subcellular localization of lncRNAs.

Validation on an independent dataset. An independent microarray dataset from the GEO database (GSE63514; 24 normal samples, 40 CIN3 samples, and 28 cancer samples) was used to validate the obtained results. Gene expression was obtained using GEO2R GUI available on the sample panel in GEO (all comparison groups were likened to those performed on the original dataset). Differential expression analysis was performed within GEO2R.

Results

The variability analysis of RNA sequencing data and the PCA assessment of the batch effect presented in Figure 1 showed that study No. 1 was significantly different from studies No. 2 and No. 3 (Fig. 1*a*), which may indicate the presence of a batch effect. After it was excluded from the analysis, the variability of samples from different studies decreased significantly (Fig. 1*b*). In further analysis, only data from studies 2 and 3 were used. Samples from these two datasets were combined for analysis into a single dataset without normalization due to the overall low batch effect.

The results of the differential gene expression analysis are shown in Figure 2. 1,572 DEGs were recorded in the "cancer *vs.* control" comparison, also 1,260 DEGs were recorded in the "cancer *vs.* neoplasia" comparison. It is important to note, that only 2 genes were observed to be differentially expressed in the "neoplasia *vs.* control" comparison.

The genes with the largest difference in the expression level for all comparisons are shown in Table 2. The top 10 genes

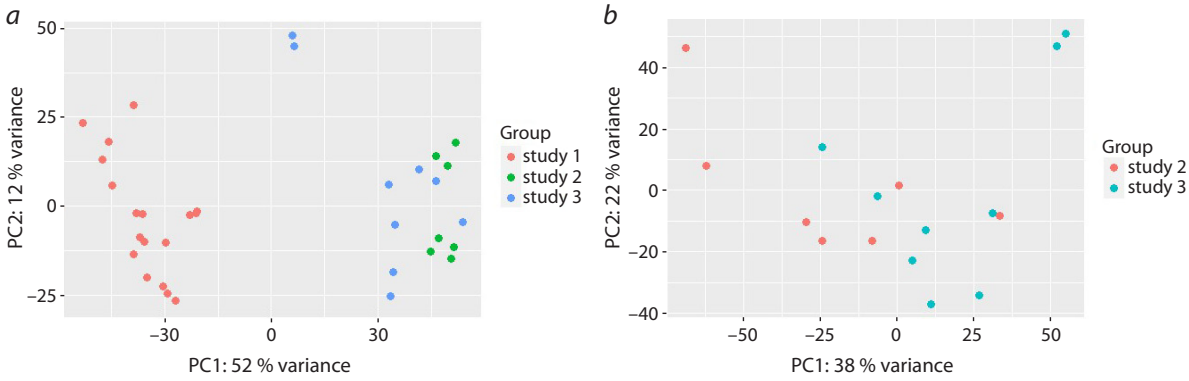

Fig. 1. RNA sequencing data variability analysis using PCA before the exclusion of study No. 1 (*a*) and after its exclusion (*b*).

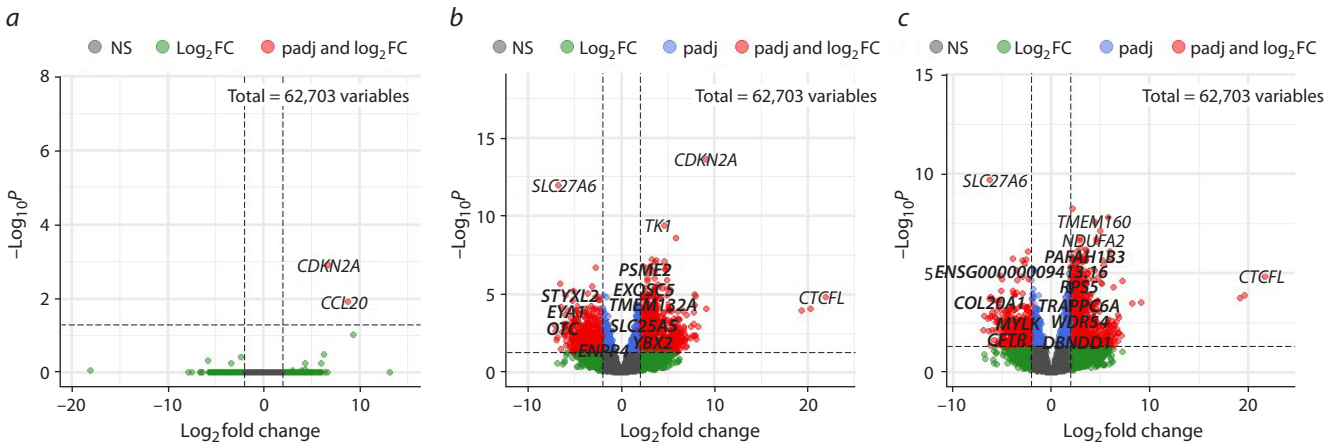

Fig. 2. Volcano plots for differentially expressed genes in the "neoplasia *vs.* control" (*a*), "cancer *vs.* control" (*b*), and "cancer *vs.* neoplasia" (*c*) comparison groups.

Table 2. Top differentially expressed genes in the "neoplasia *vs.* control", "cancer *vs.* control", and "cancer *vs.* neoplasia" comparisons

Comparison	ENSEMBL ID	Gene name	Transcript type	log_2FC
Neoplasia *vs.* control, increased expression	ENSG00000115009.13	*CCL20*	Protein-coding	8.78
	ENSG00000147889.18	*CDKN2A*	Protein-coding	6.75
Cancer *vs.* control, increased expression	ENSG00000124092.13	*CTCFL*	Protein-coding	21.91
	ENSG00000290242.1	*piR-55219*	piRNA	20.25
	ENSG00000176165.13	*FOXG1*	Protein-coding	19.33
	ENSG00000147889.18	*CDKN2A*	Protein-coding	9.16
	ENSG00000019186.10	*CYP24A1*	Protein-coding	9.15
	ENSG00000149968.12	*MMP3*	Protein-coding	8.29
	ENSG00000119547.6	*ONECUT2*	Protein-coding	8.04
	ENSG00000196611.6	*MMP1*	Protein-coding	7.99
	ENSG00000118156.13	*ZNF541*	Protein-coding	7.98
	ENSG00000163064.7	*EN1*	Protein-coding	7.92
Cancer *vs.* control, decreased expression	ENSG00000253105.6	*AP003548.1*	lncRNA	–7.15
	ENSG00000124205.18	*EDN3*	Protein-coding	–7.06
	ENSG00000170893.4	*TRH*	Protein-coding	–6.91
	ENSG00000259458.1	*MGC15885*	lncRNA	–6.83
	ENSG00000113396.13	*SLC27A6*	Protein-coding	–6.72
	ENSG00000285336.1	*LOC101928882*	lncRNA	–6.58
	ENSG00000145242.14	*EPHA5*	Protein-coding	–6.50
	ENSG00000185069.2	*KRT76*	Protein-coding	–6.49
	ENSG00000279030.1	*AC007336.3*	Uncategorized transcript	–6.39
	ENSG00000280650.1	*KCNIP4-IT1*	lncRNA	–6.10
Cancer *vs.* neoplasia, increased expression	ENSG00000124092.13	*CTCFL*	Protein-coding	21.74
	ENSG00000282122.1	*IGHV7-4-1*	Protein-coding	19.65
	ENSG00000290242.1	*piR-55219*	piRNA	19.22
	ENSG00000127129.10	*EDN2*	Protein-coding	9.09
	ENSG00000213058.3	*RPS14*	Pseudogene	8.26
	ENSG00000107159.14	*CA9*	Protein-coding	7.22
	ENSG00000241749.4	*RPSAP52*	Pseudogene	7.12
	ENSG00000133328.4	*PLAAT2*	Protein-coding	6.85
	ENSG00000287929.1	*lnc-LAMC1-1*	lncRNA	6.84
	ENSG00000181617.6	*FDCSP*	Protein-coding	6.79
Cancer *vs.* neoplasia, decreased expression	ENSG00000170893.4	*TRH*	Protein-coding	–6.91
	ENSG00000289337.1	*piR-52324-054*	piRNA	–6.79
	ENSG00000145808.10	*ADAMTS19*	Protein-coding	–6.47
	ENSG00000248698.6	*LINC01085*	lncRNA	–6.34
	ENSG00000113396.13	*SLC27A6*	Protein-coding	–6.24
	ENSG00000178115.12	*GOLGA8Q*	Protein-coding	–6.21
	ENSG00000279622.2	*lnc-ZDHHC7-3*	lncRNA	–6.16
	ENSG00000145242.14	*EPHA5*	Protein-coding	–6.13
	ENSG00000143536.7	*CRNN*	Protein-coding	–6.13
	ENSG00000249421.2	*ADAMTS19-AS1*	lncRNA	–6.09

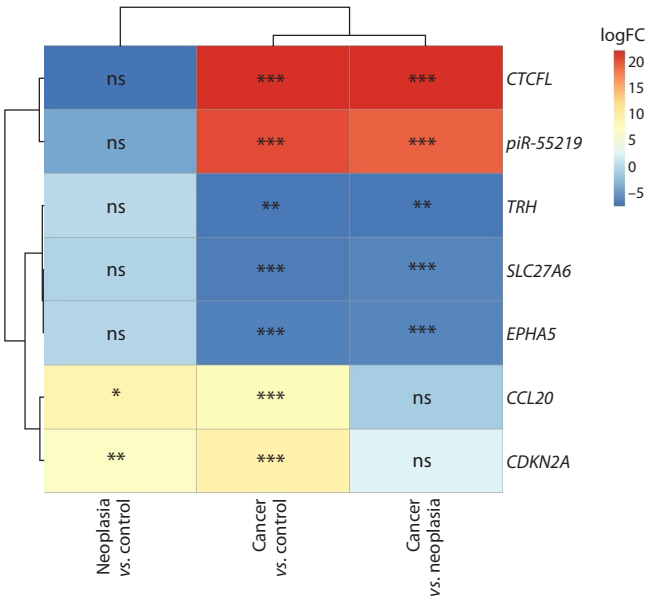

Fig. 3. Heatmap of the expression change patterns by logFC of selected genes.

The stars are intended to flag levels of significance for *p*-adjusted >0.05 (ns), <0.05 (*), <0.01 (**), <0.001 (***).

with increased and decreased expression for the "cancer *vs.* neoplasia" and "cancer *vs.* control" comparisons and 2 DEGs for the "neoplasia *vs.* control" comparison were presented. Out of the top 10 genes with increased expression for the "cancer *vs.* neoplasia" comparison, 1 belongs to the lncRNA class, 1 belongs to the piwi-interacting RNA (piRNA) class, 2 belong to the pseudogene class, 6 belong to the protein-coding gene class; out of the genes with decreased expression, 3 genes belong to the lncRNA class, 1 gene belongs to the piRNA class, and the remaining 6 belong to the protein-coding gene class. In turn, out of the top 10 genes with increased expression for the "cancer *vs.* control" comparison, 1 belongs to the piRNA class, and 9 belong to the protein-coding gene class; out of the genes with decreased expression, 4 belong to the lncRNA class, 1 belongs to the uncategorized transcript class, and the remaining 5 belong to the protein-coding gene class.

Search for common genes among the most strongly differentially expressed genes among all comparison groups resulted in the identification of an expression signature consisting of the genes *CCL20*, *CDKN2A*, *CTCFL*, *piR-55219*, *TRH*, *SLC27A6* and *EPHA5*. The expression patterns of these genes are shown in Figure 3.

The gene enrichment analysis shown in Figure 4 demonstrated that in the "cancer *vs.* control" comparison, the molecular pathways associated with the cell cycle, DNA packaging, replication and translational mRNA base-pairing repression were activated, and the pathways associated with the membrane structure and cell-cell adhesion were repressed. Conversely, in the "cancer *vs.* neoplasia" comparison, molecular pathways related to the immunoglobulin production, antigen binding, respiratory chain and respirasome were activated, while pathways related to the translational mRNA base-pairing repression, post-transcriptional silencing, RISC and RNAi effector complexes were repressed.

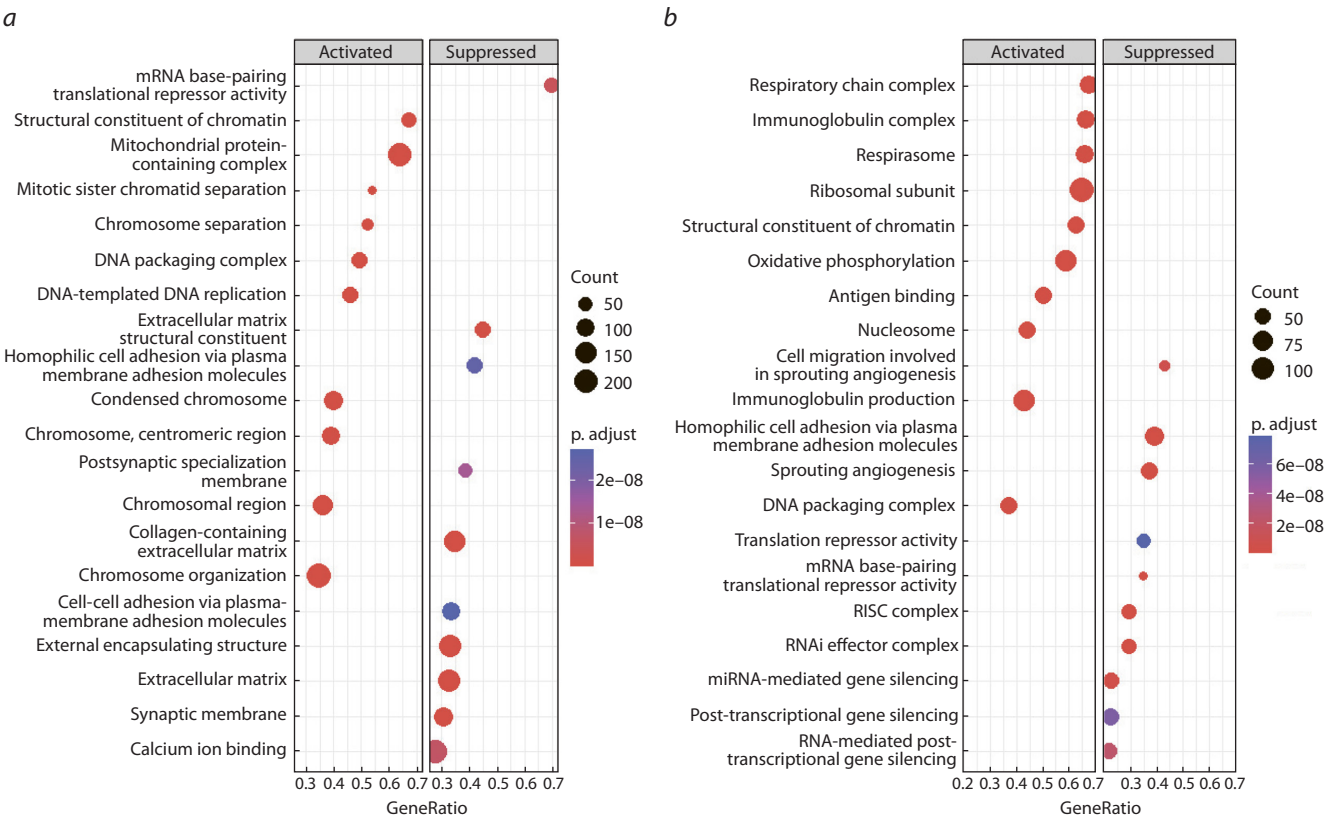

Fig. 4. Results of gene set enrichment analysis for the "cancer *vs.* control" (*a*) and "cancer *vs.* neoplasia" (*b*) comparison groups.

Table 3. Results of differentially expressed lncRNAs subcellular localization analysis using Locate-R and iLoc-LncRNA 2.0

lncRNA	Locate-R		iLoc-LncRNA 2.0	
	Localization	Score	Localization	Score
AP003548.1	Cytoplasm	1	Cytoplasm	0.84
MGC15885		1		0.73
LOC101928882		0.99		0.85
KCNIP4-IT1		1		0.83
lnc-LAMC1-1		0.94		0.86
LINC01085		1		0.69
lnc-ZDHHC7-3		0.92		0.87
ADAMTS19-AS1	Nucleus	0.97	Exosome	0.66

Table 4. Comparative analysis of the DEGs in datasets for study and validation

Comparison	Studied dataset (SRP048735+ SRA189004 + GSE149763, RNA-seq)	Validation dataset (GSE63514, microarray)	Overlap total	Overlap within the expression signature
Neoplasia *vs.* control	2	450	1	*CDKN2A*
Cancer *vs.* control	1,572	961	171	*CDKN2A*, *SLC27A6*
Cancer *vs.* neoplasia	1,260	215	15	–

Analysis of the subcellular localization of differentially expressed lncRNAs using two different web resources (Table 3) showed that subcellular localization is identified ambiguously for ADAMTS19-AS1 (nucleus and exosome), which may be related to differences in the computational approaches by which Locate-R (Local Deep SVM approach) and iLoc-LncRNA 2.0 (SVM approach) are implemented, despite the fact that both models are based on an analysis of the RNALocate lncRNA localization database. The results clearly indicated cytoplasmic localization for most of the transcripts (Table 3).

Validation on an independent dataset

Differential gene expression analysis on an independent dataset (Table 4) demonstrated that the number of DEGs observed in the "neoplasia *vs.* control" comparison was much higher than in the same comparison in our study, whereas for the other two comparisons, the number of DEGs was lower in the independent dataset than in our study. In particular, the *CDKN2A* and *SLC27A6* genes confirmed their expression change in the same pattern after validation.

Discussion

The division of HPV-infected cervical epithelial cells leads to neoplastic tissue changes or cervical intraepithelial neoplasia (CIN). The changes detected at the levels of cell morphology and tissue structure are the consequence of alterations at the molecular level. Neoplastic changes of epithelial cells in HPV infection are characterized by an increased level of transcription of the *CCL20* and *CDKN2A* genes, which lasts through the progression of the malignant process.

Tumor development assumes long-term persistence of HPV and the formation of a high viral load. Moreover, the virus can use the replicative apparatus of human cells and avoid the action of immune system factors. The main mechanisms of evasion from the immune system include modulation of antigen presentation, inhibition of cytokines and chemoattractants, modulation of cell adhesion molecule synthesis and inhibition of antigen-presenting cell migration.

The *CCL20* gene, differentially expressed in neoplastic changes of infected cells, may be a direct participant in these processes. *CCL20* belongs to the subfamily of small cytokine CC genes, it is located on chromosome 2q and contains 4 exons and 3 introns. This gene encodes macrophage inflammatory protein (MIP)-3α, predominantly expressed in liver, colon, prostate, cervix, and skin. It has been reported that endothelial cells, neutrophils, T helper 17 (Th17) cells, B cells, natural killer cells, dendritic cells (DC) and macrophages secrete CCL20 (Yamazaki et al., 2008; Nandi et al., 2014). CCL20 as a chemoattractant is involved in recruiting lymphocytes and dendritic cells to epithelial cells. It is believed that CCL20 may play an important role in the regulation of Langerhans cells, which are the main antigen-presenting cells for HPV presentation, causing an immune response.

From this perspective, it is reasonable to assume that active expression of CCL20 would be triggered in response to the appearance of human papillomavirus in the body. However, many studies indicate that HPV oncoproteins E6 and E7 may reduce the production of the chemokine CCL20 in keratinocytes by inhibiting its transcription. And thus, HPV, in an attempt to avoid an immune response, may negatively modulate the expression of this chemokine in the epithelium, thereby blocking the migration of inflammatory cells, such as Langerhans cells, to the lesion site (Guess, McCance, 2005; Wang et al., 2010; Jiang, Xue, 2015; Fernandes et al., 2021).

However, in the later stages of cervical carcinogenesis, the landscape changes and CIN3 lesions often contain myeloid cells such as macrophages and dendritic cells (Mazibrada et

al., 2008) and, as a number of studies have shown, CCL20 levels in cervical cancer tissues are significantly higher than in non-tumor and normal control tissues (Yu et al., 2015). Cervical cancer cells have been found to instruct cervical fibroblasts to produce CCL20 (Walch-Ruckheim et al., 2015). The rationale is that although normal immune cells attack and suppress tumor cells, some immune cells that infiltrate cancer tissue lose their anti-tumor function and play a role in promoting tumor progression (Beatty, Gladney, 2015; Binnewies et al., 2018).

Alteration of the cell cycle of infected epitheliocytes is possible due to changes in the transcription level of the *CDKN2A* gene. *CDKN2A* is a cyclin-dependent kinase 2a inhibitor gene that, through the use of alternative reading frames, produces two major proteins: p16 (INK4), an inhibitor of cyclin-dependent kinase 2 which arrests the G1-S transition in the cell cycle, and p14 (ARF), which binds the p53-stabilizing protein MDM2 (Robertson, Jones, 1999). It is important to note that *CDKN2A* is overexpressed in various cancers, and often its expression level correlates with the number of mutations, microsatellite instability in the tumor genome, and immune infiltration in the tumor microenvironment (Chen Z. et al., 2021). However, a study of *CDKN2A* expression in cervical cancer cell lines performed by real-time PCR and western blotting showed that it was reduced; moreover, the authors concluded that CDKN2A inhibits cell proliferation and invasion in cervical cancer through the AKT-mTOR lactate dehydrogenase mediated pathway (Luan et al., 2021). Several bioinformatics studies analyzing RNA sequencing data of cervical cancer samples have found that *CDKN2A* is a kind of "nodal gene" of tumorigenesis through interactions with various transcription factors, signaling molecules and microRNAs (e. g. miR-424-5p and miR-9-5p) and is overexpressed in cervical carcinoma in the TCGA project (Zhao et al., 2018; Chen Z. et al., 2021). In our study, *CDKN2A* expression was upregulated in patients with both HPV-associated neoplasia and HPV-associated cervical cancer, which draws attention to the importance of a more thorough study of the expression pattern of this gene and the features of the above pathologic conditions.

At the same time, the pattern of *CDKN2A* methylation in cervical cancer is relatively well known; several meta-analyses have shown that *CDKN2A* hypermethylation (relative to control samples) can be an indicator of early disease progression (Li J. et al., 2016). *CDKN2A* methylation was found to gradually increase with disease progression from stage 1 neoplasia to cervical cancer (Wijetunga et al., 2016), which can also be used as a comparative marker of disease severity. We would like to emphasize the need for a study linking the expression and methylation status of *CDKN2A* in HPV-associated neoplasia and cervical cancer to expand the understanding of the functional role of *CDKN2A* regulation in these conditions.

The most significant reduction of expression level in cancer cells relative to both control and neoplasia was found for five transcripts: *SLC27A6*, *EPHA5*, *TRH*, *CTCFL*, and *piR-55219.*

The *SLC27A6* gene encodes a fatty acid transfer protein through the cell membrane. Long-chain fatty acids are essential for various physiological processes. The function of SLC27A6 in cervical cancer has not, to our knowledge, been clarified. However, it is reported that *SLC27A6* expression was decreased in esophageal squamous cell carcinoma and breast cancer cells as well as nasopharyngeal carcinoma cells compared to normal cells (Xu C.Q. et al., 2015; Yen et al., 2019). It was also observed that the methylation ratio of the *SLC27A6* promoter was higher in nasopharyngeal carcinoma than in nonmalignant tissues (Xu C.Q. et al., 2015). On the contrary, *SLC27A6* gene expression was increased in papillary thyroid carcinoma (Dai et al., 2020).

The function of the ephrin A5 receptor encoded by the *EPHA5* gene in cervical cancer is also unclear. However, suppression of *EPHA5* expression by methylation has been established for breast cancer (Fu et al., 2010), prostate cancer (Li S. et al., 2015), and colorectal cancer (Kober et al., 2011). The loss of *EPHA5* expression was associated with the degree of serous ovarian carcinoma – the expression of this gene in cancer was reduced by 45 % in relation to neoplasia (Chen X. et al., 2016).

The *TRH* gene encodes a member of the thyrotropin-releasing hormone family involved in the hypothalamus–pituitary–thyroid axis which exhibits feedback of thyroid hormone, thereby regulating metabolic and immunological homeostasis. *TRH* has been well investigated in the type of cancer such as acute myeloid leukemia, and a correlation between risk groups and *TRH* expression was found, and it was discovered that patients with higher *TRH* expression were more sensitive to chemotherapy (Gao et al., 2022). Regarding CIN and cervical cancer, site-specific assessment of *TRH* gene methylation (cg01009664) was investigated for the detection of CIN2+ and demonstrated high sensitivity and specificity with clinician-collected samples, but not with the self-collected ones (Chaiwongkot et al., 2023). A similar analysis was also performed using screening of *TRH* cg01009664 methylation for prediction of oral squamous cell carcinoma and oropharyngeal squamous cell carcinoma (Puttipanyalears et al., 2018).

Conversely, the *CTCFL* and *piR-55219* genes are significantly upregulated in their expression in both "cancer *vs.* neoplasia" and "cancer *vs.* control" comparisons.

The *CTCFL* gene, which is sometimes also called *BORIS*, is a paralog of the widely known CTCF transcription factor and is normally expressed in pre-meiotic male germ cells together with ubiquitously expressed *CTCF* being involved in the regulation of the testis-specific genes (Soltanian, Dehghani, 2018; Debaugny, Skok, 2020). Unlike *CTCF*, *CTCFL* is more frequently amplified or transcriptionally activated, rather than mutated in cancers, and in cervical cancer the aberrant expression of *CTCFL* is linked with the re-initiating promoter hypomethylation of this gene (Debaugny, Skok, 2020). Moreover, a study performed on the cervical cancer stem-like cells (CSCs)/cancer-initiating cells (CICs) claimed that *BORIS* sf6 (isoform from subfamily 6) is specifically expressed in cervical CSCs/CICs and has a role in the maintenance of CSCs/CICs and proposed a peptide isoform BORIS C34_24(9) as a promising candidate for cervical CSC/CIC-targeting immunotherapy (Asano et al., 2016). Clinically, in cases of epithelial ovarian cancer and cervical cancer, high levels of *BORIS* expression were associated with poorer prognosis/less median survival times of patients and advanced cancer stages (Soltanian, Dehghani, 2018).

Way less is known about piwi-interacting RNA *piR-55219*. In general, piwi-interacting RNAs (piRNAs), which are

25–31 nucleotides in length, have been found to cluster at transposon loci in the genome and are thought to be critical for silencing these mobile genetic elements, via DNA methylation, to maintain genomic integrity in germline stem cells. Although they have only recently been identified in cancers, it is possible that the piRNAs that mediate transposon silencing during normal germline differentiation are hijacked in cancer cells to silence other parts of the genome, resulting in a tumorigenic state (Siddiqi, Matushansky, 2012; Suzuki et al., 2012). Unfortunately, no specific information on the involvement of *piR-55219* in cancer processes has been shown, which emphasizes the need for a more detailed investigation to establish a functional relationship between piRNAs and other cancer-specific genes.

Conclusion

We identified a predominantly cytoplasmic localization for the majority of differentially expressed lncRNAs. These lncRNAs can be involved in post-transcriptional regulation through their influence on the stability of mRNAs, act as translation regulators while forming mRNA-lncRNA complexes and can release miRNAs from their target genes as miRNA "sponges" (Xu Y. et al., 2023). All these processes may be impaired in cervical cancer, so it is important to further investigate the molecular mechanisms of function of lncRNAs selected in this study.

The results of our study differ significantly between the discovery and validation cohorts, which may be related to sample preparation protocols (FF+FFPE *vs.* cryosectioning) and expression assessment method (RNA-seq *vs.* microarray), which once again confirms the need to generate large protocol-uniformed datasets for studying neoplasia and cervical cancer at the same time.

Therefore, the analysis of differential gene expression in HPV-infected neoplasia and cervical cancer revealed a pattern of 7 genes with altered transcription levels. The increased transcription level of the *CCL20* and *CDKN2A* genes occurs at the stage of neoplastic epithelial changes and persists in cervical cancer. The *CDKN2A* and *SLC27A6* genes confirmed their expression change in the same patterns after validation on the independent microarray dataset.

References

Ahmad A., Lin H., Shatabda S. Locate-R: subcellular localization of long non-coding RNAs using nucleotide compositions. *Genomics.* 2020;112(3):2583-2589. DOI 10.1016/j.ygeno.2020.02.011

Asano T., Hirohashi Y., Torigoe T., Mariya T., Horibe R., Kuroda T., Tabuchi Y., Saijo H., Yasuda K., Mizuuchi M., Takahashi A., Asanuma H., Hasegawa T., Saito T., Sato N. Brother of the regulator of the imprinted site (BORIS) variant subfamily 6 is involved in cervical cancer stemness and can be a target of immunotherapy. *Oncotarget.* 2016;7(10):11223-11237. DOI 10.18632/oncotarget.7165

Bao Y., Wang L., Shi L., Yun F., Liu X., Chen Y., Chen C., Ren Y., Jia Y. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cell. Mol. Biol. Lett*. 2019;24:38. DOI 10.1186/s11658-019-0162-0

Beatty G.L., Gladney W.L. Immune escape mechanisms as a guide for cancer immunotherapy. *Clin. Cancer Res*. 2015;21(4):687-692. DOI 10.1158/1078-0432.CCR-14-1860

Binnewies M., Roberts E.W., Kersten K., Chan V., Fearon D., Merad M., Coussens L., Gabrilovich D., Ostrand-Rosenberg S., Hedrick C., Vonderheide R., Pittet M., Jain R., Zou W., Howcroft T., Woodhouse E., Weinberg R., Krummel M. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med*. 2018;24(5):541-550. DOI 10.1038/s41591-018-0014-x

Chaiwongkot A., Buranapraditkun S., Oranratanaphan S., Chuen-Im T., Kitkumthorn N. Efficiency of CIN2+ detection by thyrotropin-releasing hormone (TRH) site-specific methylation. *Viruse.* 2023; 15(9):1802. DOI 10.3390/v15091802

Chen X., Wang X., Wei X., Wang J. EphA5 protein, a potential marker for distinguishing histological grade and prognosis in ovarian serous carcinoma. *J. Ovarian Res*. 2016;9(1):83. DOI 10.1186/s13048-016-0292-1

Chen Z., Guo Y., Zhao D., Zou Q., Yu F., Zhang L., Xu L. Comprehensive analysis revealed that *CDKN2A* is a biomarker for immune infiltrates in multiple cancers. *Front. Cell Dev. Biol*. 2021;9:808208. DOI 10.3389/fcell.2021.808208

Dai J., Yu X., Han Y., Chai L., Liao Y., Zhong P., Xie R., Sun X., Huang Q., Wang J., Yin Z., Zhang Y., Lv Z., Jia C. TMT-labeling proteomics of papillary thyroid carcinoma reveal invasive biomarkers. *J. Cancer*. 2020;11(20):6122-6132. DOI 10.7150/jca.47290

Debaugny R., Skok J. CTCF and CTCFL in cancer. *Curr. Opin. Genet. Dev*. 2020;61:44-52. DOI 10.1016/j.gde.2020.02.021

Eklund C., Lagheden C., Robertsson K.D., Forslund O., Dillner J. Technical Report on the Global HPV LabNet DNA Genotyping Proficiency Panel 2019. International Human Papillomavirus (HPV) Reference Center, 2020

Fernandes A.T., Carvalho M., Avvad-Portari E., Rocha N., Russomano F., Roma E.H., Bonecini-Almeida M. A prognostic value of $CD45RA^+$, $CD45RO^+$, $CCL20^+$ and $CCR6^+$ expressing cells as 'immunoscore' to predict cervical cancer induced by HPV. *Sci. Rep*. 2021;11(1):8782. DOI 10.1038/s41598-021-88248-x

Fu D.Y., Wang Z.M., Wang B.L., Chen L., Yang W.T., Shen Z.Z., Huang W., Shao Z.M. Frequent epigenetic inactivation of the receptor tyrosine kinase *EphA5* by promoter methylation in human breast cancer. *Hum. Pathol*. 2010;41(1):48-58. DOI 10.1016/j.humpath.2009.06.007

Gao Y., Zhou J., Mao J., Jiang L., Li X.-P. Identification of the Thyrotropin-Releasing Hormone (TRH) as a novel biomarker in the prognosis for acute myeloid leukemia. *Biomolecules*. 2022;12(10): 1359. DOI 10.3390/biom12101359

Gebrie A. Disease progression role as well as the diagnostic and prognostic value of microRNA-21 in patients with cervical cancer: a systematic review and meta-analysis. *PLoS One*. 2022;17(7):e0268480. DOI 10.1371/journal.pone.0268480

Guess J.C., McCance D.J. Decreased migration of Langerhans precursor-like cells in response to human keratinocytes expressing human Papillomavirus type 16 E6/E7 is related to reduced macrophage inflammatory protein-3α production. *J. Virol.* 2005;79(23):14852-14862. DOI 10.1128/JVI.79.23.14852-14862.2005

Hu Z., Zhu D., Wang W., Li W., Jia W., Zeng X., Ding W., Yu L., Wang X., Wang L., Shen H., Zhang C., Liu H., Liu X., Zhao Y., Fang X., Li S., Chen W., Tang T., Fu A., Wang Z., Chen G., Gao Q., Li S., Xi L., Wang C., Liao S., Ma X., Wu P., Li K., Wang S., Zhou J., Wang J., Xu X., Wang H., Ma D. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat. Genet*. 2015;47(2):158-163. DOI 10.1038/ng.3178

Jiang B., Xue M. Correlation of E6 and E7 levels in high-risk HPV16 type cervical lesions with CCL20 and Langerhans cells. *Genet. Mol. Res.* 2015;14(3):10473-10481. DOI 10.4238/2015.September.8.8

Karimzadeh M., Arlidge C., Rostami A., Lupien M., Bratman S., Hoffman M. Human papillomavirus integration transforms chromatin to drive oncogenesis. *Genome Biol.* 2023;24(1):142. DOI 10.1186/s13059-023-02926-9

Kober P., Bujko M., Olędzki J., Tysarowski A., Siedlecki J.A. Methyl-CpG binding column-based identification of nine genes hypermethylated in colorectal cancer. *Mol. Carcinog.* 2011;50(11):846-856. DOI 10.1002/mc.20763

Li J., Zhou C., Zhou H., Bao T., Gao T., Jiang X., Ye M. The association between methylated *CDKN2A* and cervical carcinogenesis, and its

diagnostic value in cervical cancer: a meta-analysis. *Ther. Clin. Risk Manag*. 2016;12:1249-1260. DOI 10.2147/TCRM.S108094

Li S., Zhu Y., Ma C., Qiu Z., Zhang X., Kang Z., Wu Z., Wang H., Xu X., Zhang H., Ren G., Tang J., Li X., Guan M. Downregulation of EphA5 by promoter methylation in human prostate cancer. *BMC Cancer*. 2015;15:18. DOI 10.1186/s12885-015-1025-3

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

Martin J.A., Wang Z. Next-generation transcriptome assembly. *Nat. Rev. Genet*. 2011;12(10):671-682. DOI 10.1038/nrg3068

Mazibrada J., Rittà M., Mondini M., De Andrea M., Azzimonti B., Borgogna C., Ciotti M., Orlando A., Surico N., Chiusa L., Landolfo S., Gariglio M. Interaction between inflammation and angiogenesis during different stages of cervical carcinogenesis. *Gynecol. Oncol*. 2008;108(1):112-120. DOI 10.1016/j.ygyno.2007.08.095

Nandi B., Pai C., Huang Q., Prabhala R., Munshi N., Gold J. CCR6, the sole receptor for the chemokine CCL20, promotes spontaneous intestinal tumorigenesis. *PLoS One*. 2014;9(5):e97566. DOI 10.1371/journal.pone.0097566

Okunade K.S. Human papillomavirus and cervical cancer. *J. Obstet. Gynaecol*. 2020;40(5):602-608. DOI 10.1080/01443615.2019.1634030

Puttipanyalears C., Arayataweegool A., Chalertpet K., Rattanachayoto P., Mahattanasakul P., Tangjaturonsasme N., Kerekhanjanarong V., Mutirangura A., Kitkumthorn N. TRH site-specific methylation in oral and oropharyngeal squamous cell carcinoma. *BMC Cancer.* 2018;18(1):786. DOI 10.1186/s12885-018-4706-x

Qi D., Li H., Wang S., Wang S., Zheng R., Liu N., Han B., Liu L. Construction of ceRNA network and key gene screening in cervical squamous intraepithelial lesions. *Medicine* (*Baltimore*). 2022;101(48):e31928. DOI 10.1097/MD.0000000000031928

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

Royse K., Zhi D., Conner M., Clodfelder-Miller B., Srinivasasainagendra V., Vaughan L., Skibola C., Crossman D., Levy S., Shrestha S. Differential gene expression landscape of co-existing cervical pre-cancer lesions using RNA-seq. *Front. Oncol.* 2014;4:339. DOI 10.3389/fonc.2014.00339

Schubert M., Bauerschlag D., Muallem M., Maass N., Alkatout I. Challenges in the diagnosis and individualized treatment of cervical cancer. *Medicina* (*Kaunas*). 2023;59(5):925. DOI 10.3390/medicina59050925

Siddiqi S., Matushansky I. Piwis and piwi-interacting RNAs in the epigenetics of cancer. *J. Cell. Biochem*. 2012;113(2):373-380. DOI 10.1002/jcb.23363

Soltanian S., Dehghani H. BORIS: a key regulator of cancer stemness. *Cancer Cell Int.* 2018;18:154. DOI 10.1186/s12935-018-0650-8

Su Z.D., Huang Y., Zhang Z.Y., Zhao Y.W., Wang D., Chen W., Chou K.C., Lin H. iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics.* 2018;34(24):4196-4204. DOI 10.1093/bioinformatics/bty508

Sung H., Ferlay J., Siegel R.L., 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):209-249. DOI 10.3322/caac.21660

Suzuki R., Honda S., Kirino Y. PIWI expression and function in cancer. *Front. Gene*. 2012;3:204. DOI 10.3389/fgene.2012.00204

Walch-Ruckheim B., Mavrova R., Henning M., Vicinus B., Kim Y.J., Bohle R., Juhasz-Boss I., Solomayer E.F., Smola S. Stromal fibroblasts induce CCL20 through IL6/C/EBPβ to support the recruitment of Th17 cells during cervical cancer progression. *Cancer Res*. 2015;75(24):5248-5259. DOI 10.1158/0008-5472.CAN-15-0732

Wang X., Gao X.H., Hong Y., Li X., Chen H.D. Local hyperthermia decreases the expression of CCL-20 in condyloma acuminatum. *Virol. J.* 2010;7:301. DOI 10.1186/1743-422X-7-301

Wijetunga N.A., Belbin T., Burk R., Whitney K., Abadi M., Greally J., Einstein M., Schlecht N. Novel epigenetic changes in *CDKN2A* are associated with progression of cervical intraepithelial neoplasia. *Gynecol. Oncol*. 2016;142(3):566-573. DOI 10.1016/j.ygyno.2016.07.006

Xu C.Q., Zhu S.T., Wang M., Guo S.L., Sun X.J., Cheng R., Xing J., Wang W.H., Shao L.L., Zhang S.T. Pathway analysis of differentially expressed genes in human esophageal squamous cell carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* 2015;19(9):1652-1661

Xu Y., Sun Y., Song X., Ren J. The mechanisms and diagnostic potential of lncRNAs, miRNAs, and their related signaling pathways in cervical cancer. *Front. Cell Dev. Biol.* 2023;11:1170059. DOI 10.3389/fcell.2023.1170059

Yamazaki T., Yang X., Chung Y., Fukunaga A., Nurieva R., Pappu B., Martin-Orozco N., Kang H.S., Ma L., Panopoulos A., Craig S., Watowich S., Jetten A., Tian Q., Dong C. CCR6 regulates the migration of inflammatory and regulatory T cells. *J. Immunol*. 2008;181(12):8391-8401. DOI 10.4049/jimmunol.181.12.8391

Yan X., Hu Z., Feng Y., Hu X., Yuan J., Zhao S., Zhang Y., Yang L., Shan W., He Q., Fan L., Kandalaft L., Tanyi J., Li C., Yuan C.X., Zhang D., Yuan H., Hua K., Lu Y., Katsaros D., Huang O., Montone K., Fan Y., Coukos G., Boyd J., Sood A., Rebbeck T., Mills G., Dang C., Zhang L. Comprehensive genomic characterization of long non-coding RNAs across human cancers. *Cancer Cell*. 2015;28(4):529-540. DOI 10.1016/j.ccell.2015.09.006

Yen M.C., Chou S.K., Kan J.Y., Kuo P.L., Hou M.F., Hsu Y.L. New insight on solute carrier family 27 member 6 (SLC27A6) in tumoral and non-tumoral breast cells. *Int. J. Med. Sci.* 2019;16(3):366-375. DOI 10.7150/ijms.29946

Yu Q., Lou X.M., He Y. Preferential recruitment of Th17 cells to cervical cancer via CCR6-CCL20 pathway. *PLoS One*. 2015;10(3):e0120855. DOI 10.1371/journal.pone.0120855

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

Conflict of interest. The authors declare no conflict of interest.

Received July 21, 2023. Revised January 18, 2024. Accepted January 19, 2024.