**Complete genome analysis of the human papillomavirus type 16 (HPV16) isolates from cervical carcinomas in Latvia**

Nikita Zrelovs1\*, Juris Jansons 2,3, Arta Spridzane3,4, Liba Sokolovska3,5, Karina Biserova3,5, Daira Krisane5, Svetlana Gebrila3,5, Beatrise Orlova6, Marta Petrovska6, Valery Ilinsky7, Anna Ilinskaya7, Jurijs Nazarovs5, Androniks Mitildzans3,4, Maria Isaguliants3.

1Scientific Laboratory of Molecular Genetics, Riga Stradins University, Riga, Latvia

2 Biomedical Research and Study Centers, Riga, Latvia

3 Institute of Microbiology and Virology, Riga Stradins University, Riga, Latvia

4 Pathology Centre, Riga East Clinical University Hospital, Riga, Latvia

5 Pathology Institute, Pauls Stradins Clinical University Hospital, Riga, Latvia;

6 Centrala Laboratorija Ltd, Riga, Latvia;

7 Eligens, Human Genetics, Digital Health, Riga, Latvia

\*Presenting author

**Introduction:** Cervical cancer ranks as the fourth most prevalent cancer among women. Most cases of cervical squamous cell carcinomas (CSCC) are linked to high-risk human papillomaviruses (hrHPV). HPV has over 150 genotypes with varying characteristics, twelve to fourteen - HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68 - have been identified as high-risk for cervix cancer, with HPV16 the most prevalent. Each type has distinct genetic sequences that lead to differences in how they interact with host cells and cause the disease. Despite a high degree of conservation, hrHPVs can still mutate with mutations influencing pathogenicity and anti-viral immune response through immune escape. Such variability would shape the course of HPV-associated lesions and cancer and may affect the efficacy of anti-HPV vaccines. The aim of this study was to update our knowledge on the variability of HPV16 circulating in the Baltic countries, specifically Latvia.

**Methods:** Women (n=86) aged 42.2±10.1 years, recruited into the prospective study (Ethical Committee of Riga Stradins University (RSU) N2-PĒK-4/415/2022 dated 26/09/ 2022), visited a gynecologist during up-to 2 years with 3-12 month intervals. Cervical biopsies were collected from women with pathologies, tissues were formalin-fixed, and paraffin-embedded (FFPE-blocks). Part of FFPE-blocks were biopsies of the same participants dated 2012-2021. FFPE-blocks were sectioned, 5 mm-sections were H&E-stained and subjected to histopathological screening to grade the lesion. Part of sections was used to extract DNA (QIAamp DNA FFPE Advanced Kit). DNA was subjected to PCR detecting 14 hrHPV genotypes (Allplex HPV14, Seegene, South Korea). Samples (n=24) with high HPV16 load (Ct<25) were selected for whole genome sequencing (WGS) (Eligens, Latvia; CeGaT, Germany) running WGS protocol on the Illumina platform. Reads were mapped onto the extended human genome reference, additionally extended by the genomes of HPV types 16, 33, 39, and 56 added as additional contigs using Burrows-Wheeler Aligner (BWA-MEM algorithm). Analysis of nucleotide and amino acid (aa) substitutions was performed using the iVar tool using HPV16 NC\_001526.4 and its coding sequence annotations as a reference.

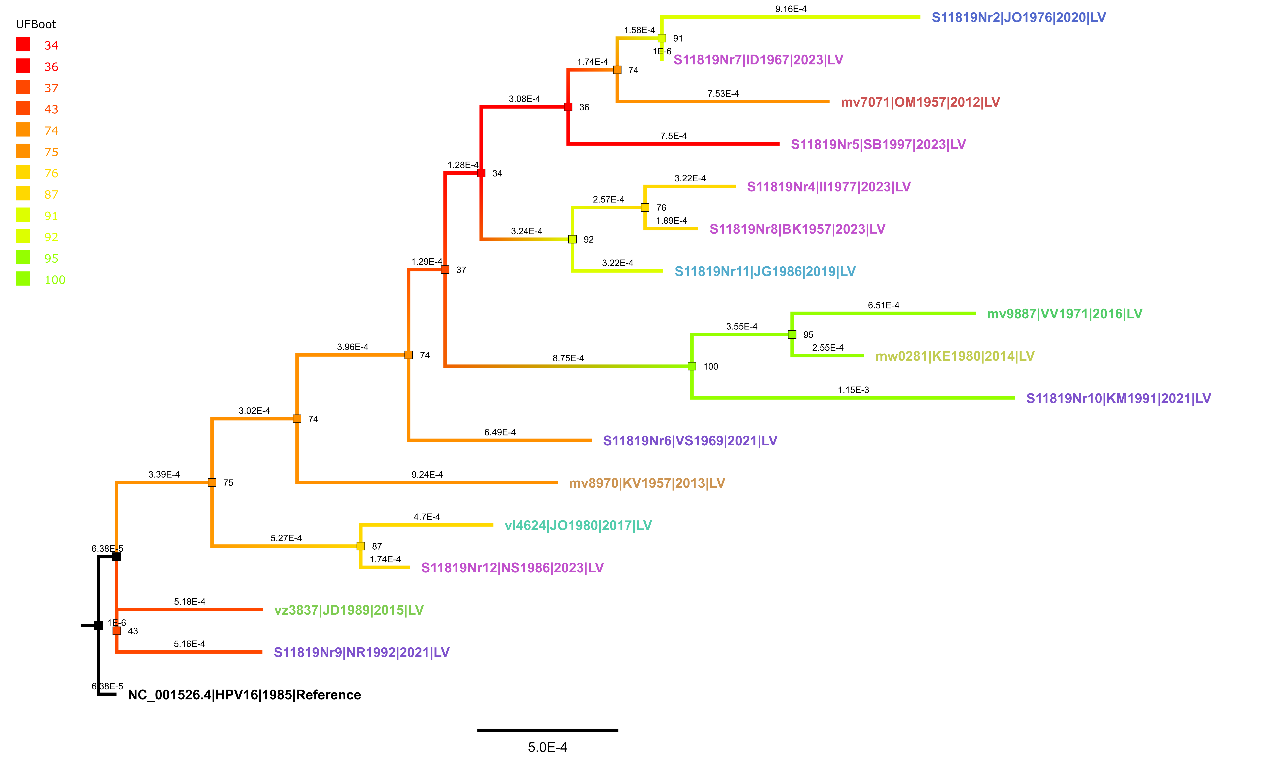
**Results:** DNA samples with a high HPV16 load, 20 samples with HPV16 mono-infection and 4 coinfected with HPV 33, 39, 56, were subjected to WGS. For 16, the whole 7906 bp long genome of the HPV16, and for one, also of HPV33 were assembled. Other attempts were unsuccessful due to DNA fragmentation. The phylogenetic analysis (Figure 1) showed that HPV16 genomic DNA was relatively highly conserved. The most divergent local sequences (≤ 24 substitutions) demonstrated less than 0,36% differences with the reference genome, as described earlier [1]. A total of 93 non-redundant variants were observed across the local genomes, with 25/93 not previously documented in other publicly available complete HPV16 isolate genomes. Most variants localized to the non-coding regions, protein-coding regions with the most substitutions were capsid proteins L1 and L2 (Table 1). Amino acid (aa) substitutions in L1 and L2 were specific to the geographic area [2] A number of non-synonymous aa substitutions were earlier found to be associated to an increased HPV16 pathogenicity or increased risks of developing CC: S220T (31% of isolates) in E1 DNA helicase/NTPase [3]; P219S (62%) and T310K (19%) in E2 transcriptional transregulator [4]; a combination of I44L and I65V (56%) in E5 oncoprotein [5]; L90V (44%) in E6 transforming protein [6]. E7 protein was identical among all analyzed local isolates (Table 1).

**Conclusions:** We have obtained 16 genomic sequences of HPV16 circulating in squamous cell carcinomas in Riga, Latvia in 2012-2023. The variability of HPV16 genomes was low. Aa substitutions occur at the sites determining functions of HPV16 proteins and appear to be specific to the East-Europe/Baltic region. Updated information on HPV16 variability is important for epidemiology, the development of HPV vaccines, and personalized treatment of HPV16-associated dysplasia and cancer.

**Acknowledgments** Latvian Science Council grant LZP-2021/1-0484.

**References** (1) Liu Y et al. *Sci Rep.* 2017**; (2)** Tsakogiannis D et al. *Viruses.* 2022; (3) Yao Y, et al. *Int J Med Sci.* 2019**; (4)** Kahla S, et al. *Biomed Res Int.* 2014; (5) da Silva-Júnior AHP, et al. *Trop Med Infect Dis.* 2024; (6) Kottaridi C, et al. *Viruses.* 2022

**Keywords:** cervical squamous cell carcinoma,high-risk human papillomavirus (HPV), whole genome sequencing, amino acid substitutions.



**Figure 1.** Reference-rooted maximum likelihood tree of the complete HPV16 genomes from LV. The tree is drawn to scale, and branch lengths (indicated above the branches) correspond to the number of nucleotide substitutions per site. Branches also have their UFBoot support percentage (out of 1000 replicates) indicated next to their distal nodes and are colored in gradient according to it. Tip labels of the Latvian strains reconstructed from WGS are in the form of "Accession|Strain|Year|Origin" and are colored arbitrarily based on the sample isolation year. Multiple sequence alignment used as input for the generation of the tree was performed using MAFFT in L-INS-i mode and had 17 sequences, 7911 columns, 97 distinct patterns, 30 parsimony-informative, 51 singleton sites, and 7830 constant sites Maximum-likelihood tree was generated using IQ-TREE with the HKY+F+I chosen as the best-fit model according to BIC after the ModelFinder analysis.

**Table 1**. Amino acid substitutions were found in the analyzed complete genome sequences of HPV16 isolates from Latvia. Variants allegedly unique to Latvia (i.e., not found in the publicly available complete HPV16 isolate genomes from elsewhere) are annotated in green.

