**CHARACTERISTICS OF E6 AND E7 ONCOPROTEINS OF HPV16 ISOLATES CIRCULATING IN EUROPE WITH FOCUS ON THE BALTIC REGION**

Juris Jansons 1,2\*, Nikita Zrelovs3, Liba Sokolovska1,4, Arta Spridzane1,5, Karina Biserova1,4, Daria Avdoshina, Daira Krisane4, Svetlana Gebrila1,4, Beatrise Orlova6, Marta Petrovska6, Ilze Lindenberga7, Dmitrijs Perminovs7, Valery Ilinsky8, Anna Ilinskaya8, Androniks Mitildzans1,4, Jurijs Nazarovs5, Stefan Petkov9, Maria Isaguliants1,9.

*1 Institute of Microbiology and Virology, Riga Stradins University, Riga, Latvia*

*2 Biomedical Research and Study Centers, Riga, Latvia*

*3 Scientific Laboratory of Molecular Genetics, Riga Stradins University, Riga, Latvia*

*4 Pathology Institute, Pauls Stradins Clinical University Hospital, Riga, Latvia;*

*5 Pathology Centre, Riga East Clinical University Hospital, Riga, Latvia*

*6 Centrala Laboratorija Ltd, Riga, Latvia, Riga, Latvia;*

*7 E. Gulbis Laboratory, Riga, Latvia;*

*8 Eligens, Human Genetics, Digital Health, Riga, Latvia;*

*9 Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden.*

*\*Presenting author*

**Introduction** E6 and E7 oncoproteins of the high risk human papillomaviruses (hrHPVs) play a key role in the oncogenesis associated with hrHPV infection. Data on the variability of these proteins are limited, and the factors affecting their variability are still poorly understood. The aim of this study was to characterize E6 and E7 oncoproteins derived from HPV16 strains circulating in Europe, specifically the North-East Europe/Baltic region, with respect to their variability/conservation and evolution.

**Materials and Methods** Formalin-fixed paraffin embedded (FFPE) cervical biopsies and conization materials from the archive of Paul Stradins University Hospital (n=32), Academic Laboratory (n=21) or patient donates (n=6) were sectioned (5 mm). H&E-stained sections were assessed by histopathologists to grade lesions. DNA, extracted from 10 sections per sample using QIAamp DNA FFPE Advanced Kit, was screened for the presence of 14 hrHPV genotypes by commercial PCR DNA (Anyplex or Allplex HPV14, Seegene, South Korea). Samples with medium or high HPV16 load by Anyplex (n=68) were subjected to Sanger sequencing of E6-E7 coding region, using pairs of primers E6-53/E6-603, and E6-531/E7-923 [1]. Samples generating Ct<25 in quantitative hrHPV PCR (n=24) were subjected to the whole-genome sequencing (WGC) run on the Illumina platform (Eligens, Latvia; CeGaT, Germany). Reads were mapped onto the extended human genome reference, additionally extended by HPV16 genome using Burrows-Wheeler Aligner (BWA-MEM algorithm). Analysis of nucleotide and amino acid (aa) substitutions was performed using the iVar tool with coding sequence annotations of HPV16 NC\_001526.4 as a reference. Further, sequences encoding E6 and E7 of HPV16 containing respective open reading frames with known collection date from the European countries (n=296) were downloaded from GenBank. Multiple sequence alignments of E6 and E7 sequences were subjected to covariation analysis/correlated mutation analysis measuring dependence between each two positions [2]. Significance of the difference of dependence from the noise was assessed by series of permutation tests (n=300) in which the analysis was done on the shuffled sequences generating random scores. Dependences were considered significant if there were higher than the random “noise” assessed using derivative integration (DI,) mutual information (MI) and observed minus expected squared (OMES) tests.

**Results** Sanger sequencing generated 32 (GenBank JA19850413-12012021) and WGS, 16 sequences of E6E7-coding region. The phylogenetic analysis of the European E6 sequences revealed no separation into country- or region-specific clades. E6 contained multiple aa substitutions, while E7 was conserved. For E6 and E7 from the North-East Europe including the Latvian sequences, the most common were substitutions of aa R17 (11,4%), Q21 (3,7%), D32 (1,9%), H85 (4,6%) and L90 (48,7%) in E6 and N29 in E7 (4,8%) compared to HPV16 reference (Table 1). Some were in the domains involved in E6/p53-binding, but only one - R17G/I/T - localized within structural/functional motive of E6 (NLS1; Fig. 1). Covariation analysis of multiple sequence alignments for E6 done by OMES demonstrated significant position association scores for variable aa positions 17-21-32-71-85-90 of E6 (p<0.05); DI revealed multiple associations including those found by OMES, while MI was insensitive. Importantly, most of the covariant positions (all except aa 71 by OMES) were among the ones subjected to frequent aa substitutions. No covariance networks were found for E7.

**Conclusions** We have analyzed aa sequences of E6 and E7 oncoproteins of HPV16 circulating in squamous cell carcinomas in Riga, Latvia in 2012-2023, and compared them to E6 from other HPV16 strains circulating in the Europe. We found isolates of HPV16 circulating in Europe to be genetically homogeneous. Pattern of aa substitutions was that characteristic to European, but not American, African or Asian HPV16 isolates [3]. Homogeneity of the European HPV16 isolates may reflect spread of infection with free migration within Europe where by 2022 HPV-vaccination was not yet accepted in 25 of 53 countries (WHO news, 2022). While E7 was highly conserved, E6 was variable. However, substitutions were not affecting the main structural/functional domains of the oncoprotein. Aa positions affected by substitutions formed covariance networks indicating viral evolution under selective pressure, well known for RNA viruses as HCV and SARS-CoV-2 [4, 5], but yet understudied for hrHPVs.

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

References

1. <https://pubmed.ncbi.nlm.nih.gov/10644829/> ;

2. <https://doi.org/10.1134/S0026893324700213>;

3. <https://doi.org/10.7554/eLife.08932.001>;

4. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3302335/>;

5. <https://elifesciences.org/reviewed-preprints/92063>

**Keywords:** HPV16, oncoproteins E6 and E7, phylogenetic analysis, amino acid substitutions, covariation analysis of multiple sequence alignments, viral evolution.

Table 1 Frequency of occurrence of aa substitutions in E6 and E7 oncoproteins of HPV16 isolates from the countries of the North-West Europe including Baltic region (nn E6/nn E7 sequences; Sweden n=2/2, Germany n=1/1, Netherlands n=162/166, Poland n=3/3, European part of Russia n=12/12, Latvia n=48/48)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Protein | AA position | AA substitutions | Number of AA substitutions | Frequency of occurrence of aa substitutions, % |
| E6 (n=212+16) | 6 | T6N | 1 | 0. 43% |
| 10 | Q10E | 1 | 0. 43% |
| 14 | E14K | 1 | 0. 43% |
| 17 | R17G  R17I  R17T | 16+1  8  1 | 17/228=7,5%  3.5%  0, 43% |
| 21 | Q21H  Q21D | 2  8 | 0.9%  3.5% |
| 24 | T24I | 1 | 0. 43% |
| 31 | H31Q | 1 | 0. 43% |
| 32 | D32N  D32E | 1  3 | 0. 43%  1.3% |
| 34 | I34L | 1 | 0. 43% |
| 35 | L35V | 1 | 0. 43% |
| 36 | E36Q  E36K | 1+1  1 | 2/228=0,9%  0.43% |
| 84 | R84S | 1 | 0. 43% |
| 85 | H85Y | 10 | 4.4% |
| 90 | L90V | 112+1 | 113/228=49,6% |
| 101 | K101T | 1 | 0.43% |
| 103 | L103V | 1 | 0.43% |
| 121 | E121K | 1 | 0.43% |
| 131 | R131T | 1 | 0.43% |
| E7 (N=216+16 WGS) | 4 | D4V | 1 | 1/232=0,43% |
| 29 | N29S | 11 | 11/232=4,7% |