SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Investigation of Prognostic Factors for Recurrent Respiratory Papillomatosis

by Zsófia Nagy

Supervisor: Dr Krisztina Szarka

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF PHARMACEUTICAL SCIENCES

DEBRECEN, 2022

INVESTIGATION OF PROGNOSTIC FACTORS FOR RECURRENT RESPIRATORY PAPILLOMATOSIS

By Zsófia Nagy, Molecular Biologist (MSc degree)

Supervisor: Dr Krisztina Szarka

Doctoral School of Pharmaceutical Sciences, University of Debrecen

Head of the **Defence Committee**:	Dr. Árpád Tósaki, DSc
Reviewers:	Dr. Enikő Fehér, PhD
	Dr. István Varga, PhD
Members of the Defence Committee:	Dr. Erika Tóth, PhD
	Dr. Bence Kozma, PhD

The PhD Defence takes place at the Augusta 315 seminar room, University of Debrecen, 11:00, 08. 11. 2022.

Introduction

Although recurrent respiratory papillomatosis (RRP) is a benign disease, the recurrences, the extensive spread of the lesions throughout the respiratory tract and the risk of malignization may cause considerable deterioration of the quality of life of some patients. The etiological role of low-risk human papillomavirus (HPV) 6 and 11 in the development of these lesions has been demonstrated, and further clinical studies have shown that the HPV11 genotype is more common in papillomatosis cases with severe outcome. The primary therapeutic option is surgical removal of the lesions, however, in 20% of cases, some form of adjuvant treatment may be required. Our earlier results showed that the presence of nucleotide polymorphisms identified in the long control region (LCR) of HPV11 may contribute to the severity of the RRP, however, the polymorphisms found in the late (L1) or in E1 and E2 early proteins may also play a role in the severe outcome of the disease caused.

Individual variation of patients has been shown to influence the development, progression and expected therapeutic response of many acute and chronic diseases. In case of infectious agents, their genetic variability may also affect the progression and outcome of the disease caused. For HPVs, this association has been extensively studied in case of high-risk genotypes, especially HPV16, however, the number of similar studies is limited in association with low-risk HPVs. Our studies provide additional data to explore the relationship between the intratypic variation of the low-risk HPV6 and HPV11, the pathogenicity of the variants, and the course of the related diseases, which may facilitate the assessment of the prognosis or even assist inthe development of personalized therapy.

Review of literature

Previously, papillomaviruses (PV) together with polyomaviruses have been classified into the *Papovaviridae* family; this was primarily due to the similar morphological characteristics of the two viruses. However, after recognizing the differences of viral genomes and physiology, the *Papovaviridae* family was divided into two families, *Papillomaviridae* and *Polyomaviridae,* by the International Committee on Taxonomy of Viruses (ICTV).

The classification of papillomaviruses is based on the sequence of the most conserved open reading frame (ORF) of the virus genome, which encodes L1 major capsid protein. If the L1 ORF sequence of a new putative PV is different from the closest PV genome at least 10%, it should be enrolled to a new PV genotype. If the difference is 2-10% between the reference and the new sequences, the genome is a new subtype, while in case of difference lower than 2%, the new genome is determined as an intratypic variant.

Papillomavirus types causing infection in humans belong to five genera (*Alpha-*, *Beta-*, *Gamma-*,

Mu-, and *Nupapillomvirus*) based on the genome similarity; the genotypes belonging to different genera have different life-cycle characteristics and disease association.

PVs have nonenveloped, icosahedral capsid and an approximately eight kilobases long circular double-stranded DNA genome. Their genome consists of eight open reading frames (ORFs). The three oncogenes, E5, E6, and E7, modulate the transformation process, E1 and E2 regulatory proteins, are required for the regulation of transcription and replication, and E4 protein plays a crucial role in the virus release. PVs have two structural proteins; the L1 is the major and L2 is the minor capsid protein. The long control region (LCR) of the PV genome – which contains several cis-responsive regulatory elements - does not encode viral proteins, but it is involved in the regulation of the viral life cycle. The LCR is localized between the L1 and E6 ORFs.

Several differences can be observed in the genome organization of the low- and high-risk HPV genotypes. While the high-risk HPVs have only one E5 protein, the low-risk genotypes encode two E5 proteins (E5A and E5B or E5γ and E5δ). The E5A shows weak transforming activity similarly to the E5 protein of the high-risk HPVs.

The other significant difference between the low- and high-risk genotypes is the regulation of the expression of the E6 and E7 oncoproteins. The E6/E7 of the high-risk HPV genotypes has one promoter sequence (the p97 in case of HPV16), while two separate promoters regulate the expression of the oncoproteins of the low-risk HPVs. In case of HPV11 the promoter of E6 protein is located within the LCR at nucleotide position 90, while the promoter region which regulates the E7 protein expression starts at nucleotide position 264.

The life cycle of HPV is strictly linked to the differentiation program of the host keratinocytes, whereby the assembly of mature virions is restricted to terminally differentiated suprabasal cells because of the differentiation-dependence of the late promoters. Due to a skin injury, the primary barrier function of the skin is disrupted, this gives possibility for infectious HPV virion to reach the cells of basal membrane layer of the stratified epithelium; the replication of the infected basal cells results in the virus genome partitioning into the daughter cells. After the mitotic cell division, one daughter cell stays in the basal membrane, while the other one starts to migrate toward to the suprabasal layers, where it differentiates terminally: during the cell migration and differentiation the infection crosses through the layers of the epithelium. During the replication cycle, first, the L1 major capsid protein binds to the heparan sulphate proteoglycans in the extracellular matrix inducing conformational changes in the viral capsid. The conformational change in the capsid exposes a conserved site on the amino terminus of minor capsid protein L2, which is then cleaved by the extracellular furin. The cleavage induces further conformational changes in the viral capsid which allows the virus binding to secondary receptors, such as α6β4 integrin and annexin A2. This is

followed by virus uptake and internalization via endosomal vesicles into the target cells. Once HPV has reached the target cells, L2 protein helps to travel of the PV DNA towards the nucleus.

Human papillomaviruses (HPV) are the most common sexually transmitted infectious agents worldwide. Two major groups of HPV-associated infections can be distinguished: the HPV-associated skin infections and mucosal diseases. Skin infections include different types of warts (for example flat, pigmented, plantar, common, filiform), the *epidermodysplasia verruciformis* (which most caused by HPV5 and HPV8), and other malignant skin tumours. HPV-related mucosal infections can result in benign lesions in different anatomical sites, such as *condyloma acuminata* and recurrent respiratory papillomatosis caused by low-risk HPV genotypes, while high-risk types can cause precancerous lesions and malignant diseases in the anogenital tract and in the head and neck region.

The anogenital HPV infections are the most common sexually transmitted diseases. The most frequent manifestation of genital papillomas is the *condyloma acuminata* appearing in the cervix, vulva or other anogenital sites in women, and the penis and anus in men.

The *leucoplakia penis* is a premalignant disorder, which can be caused by HPV infection, poor hygiene habits, but smoking can also be a risk factor.

In case of atypical squamous cells of undetermined significance (ASCUS) and low-grade squamous intraepithelial lesions (LSILs) – in contrast to high-grade squamous intraepithelial lesions (HSILs) – the affected cells can regress to their normal status more frequently, and HSIL develops only in few cases. ASCUS and LSIL are identified in 5% of the population. According to retrospective follow-up studies, 5,2% to 8,7% of ASCUS/LSIL cases can progress to HSIL.

The recurrent respiratory papillomatosis (RRP) is a rare, benign disease, which is caused by low-risk HPV6 and HPV11 genotypes. Their risk of malignant transformation is lower than 1%.

RRP is categorised into juvenile-onset (JO-RRP) and adult-onset (AO-RRP) disease based on diagnosis before or after 12 years of age, respectively. Multiple studies have shown that infection with HPV 11 is associated with a more aggressive course requiring more surgical interventions and more frequently manifests in childhood.

Numerous molecular biological methods can be used for the detection of HPV infections. The most important techniques for the HPV detection and genotyping are the target, signal, and probe amplification methods.

For the adequate diagnosis of RRP, clinicians should know the etiological aspects, the clinical features, and results of bronchoscopy and imaging examinations (for example CT). For the diagnosis of RRP, the best method is the visualization of the larynx with a flexible fibreoptic nasopharyngoscopy. This provides information about the mobility of the vocal cords and the presence

of obstructions in the airways which helps determine the need for and urgency of surgical intervention.

Currently, there is no effective treatment for eradication of the RRP. The therapy primarily focuses on the maintaining airways and quality of the voice. The current standard treatment is surgical excision with adjuvant therapies as needed. In clinical trials, the effects of interferon-alpha, cidofovir, bevacizumab, celecoxib, acyclovir, ribavirin, and indole-3-carbinol have been investigated as therapeutic options for RRP.

Three licensed prophylactic HPV vaccines are available; the bivalent Cervarix vaccine contains the virus like particles (VLPs) of the most common high-risk HPV genotypes (HPV16 and HPV18) responsible for the development of cervical cancer. In addition to the two most common high oncogenic genotypes, the tetravalent Gardasil includes the VLPs of the two most frequently identified low oncogenic HPV genotypes, HPV6 and HPV11, in RRP. The nonavalent Gardasil 9 contains VLPs of nine HPV genotypes (HPV6, HPV11, HPV16, HPV18, HPV31, HPV33, HPV45, HPV52, and HPV58).

The presence of intratypic variants was also examined for HPV6 and 11 genotypes. Our research group investigated first the presence of intratypic variation of HPV11 and the association between papillomatosis with different severity and virological status by sequence, variant, and phylogenetic analysis of HPV11 genomes. On the bases of the results, the variability of the HPV11 LCR and the region encoding the capsid proteins may be responsible for the development of diseases with different severity. Phylogenetic analysis of HPV11 GenBank sequences formerly identified two major phylogenetic lineages (A1 and A2).

In case of HPV6, two main lineages (clusters A and B) and further five sublineages (B1, B2, B3, B4, and B5) within cluster B were identified during phylogenetic analysis of the complete genome

In 2014 and 2016 Jelen and her research group investigated the intratypic variation in HPV6 and HPV11, and their results showed that there was no relationship between the intratypic diversity and the global migration pattern, in contrast to the high-risk HPV16 and HPV18 genotypes.

Aims

1. Complete genome and phylogenetic analysis of HPV11 genome sequences of the new and previously examined patients together with HPV11 genomes deposited in the GenBank. We extended our studies to the HPV6 genotype.
2. Investigation of the effect of nucleotide polymorphisms on amino acid sequence and protein structure. In case of HPV 11, we used protein modelling to investigate the relationship between amino acid changes and the function of the E2 viral protein.

3. In case of HPV11, investigation of the effect of newly identified unique polymorphisms in the long control region (LCR) by site-directed mutagenesis for HPV11.
4. Functional analysis of the long control regions (LCRs) and investigation of the interaction between HPV11 E2 variants and LCRs.

Materials and methods

Patients

Our research group examined samples from patients with recurrent respiratory papillomatosis (RRP) (tissue samples from papillomas and exfoliated cell samples from a healthy mucosa) as well as genital tissue and exfoliated cell samples. Nine of the RRP patients were HPV6 positive and 13 of them were HPV11 positive (six of which were presented by our team in a previous study). In two previously examined HPV11 positive patients (Patients 5 and 6), we could examine tissue samples from new relapses as well as exfoliated cells. Tissue samples from multiple recurrences and from multiple localizations (supraglottis, subglottis, and stoma) were also examined in case of some newly enrolled patients. The samples were collected from diseases with different severity (solitary, moderate, and aggressive papillomatosis) and 10 and 12 patients were diagnosed with adult and juvenile onset RRP, respectively.

Beside head and neck specimens, we also examined samples originating HPV6 and HPV11 positive anogenital diseases (condyloma acuminatum, ASCUS, LSIL, Cervical atypia, leucoplakia penis). The study received ethical committee approval (approval number: 4169-2014).

Nucleic acid isolation and detection of viruses

DNA was isolated from tissue and exfoliated cell samples with InnuPrep Viral DNA/RNA Kit (Analytic Jena, Jena, Germany) according to the manufacturer's recommendation. The quality of DNA was evaluated with β-globin gene-specific polymerase chain reaction (PCR). HPV-specific sequences were detected and identified by MY/GP consensus nested PCR and restriction fragment length polymorphism (RFLP) analysis of MY amplimers.

Complete genome amplification and sequencing

After determination of HPV6 and 11 genotypes, complete genome amplification was performed with primers designed by the study group. Following PCR, the amplimers were electrophorized in a 1% low melting point preparative agarose gel and the appropriate products were purified. Purified samples were checked by agarose gel electrophoresis and sequenced from both directions (Macrogen, Amsterdam, the Netherlands). Complete genome assembly, alignment with reference genomes, and analysis of sequence data were performed using CLC Main Workbench 7.9.1 software.

Phylogenetic analysis

Phylogenetic analysis was performed using a total of 197 HPV6 and 88 HPV11 complete genome sequences available in the GenBank (including the six sequences previously examined by the research group) as well as our newly identified sequences. Identical sequences were used only once.

Dendrograms were reconstructed by the CLC Main Workbench 7.9.1 software using neighbour-joining method with bootstrapping 1000 times. As an outgroup, the HPV11 reference sequence was used for phylogenetic analysis of HPV6 sequences, and the HPV6A1 (formerly HPV6b) reference sequence was used for HPV11.

Investigation of the transcription binding-sites in case of HPV6

Putative transcription factor binding sites were predicted by the PROMO tool on the ALGGEN server using a maximum dissimilarity of 10%. Loss or gain of putative transcription factor binding sites in the LCRs were determined by the comparison to the reference sequence of the (sub)lineages.

In silico modelling of the structure of the different E2 variants

Based on the amino acid sequences deduced from prototype sequence (GenBank accession number: M14119), the full-length quaternary structure of the reference E2 protein was modelled by homology modelling and loop modelling. To create the HPV11 transactivation domain (TAD) dimer, the crystal structure of the HPV11 TAD, was structurally aligned with the crystal structure of the dimerized transactivation domain of HPV16 in UCSF Chimera and subsequently structurally optimized. For generating the homology model of the HPV11 DNA-binding domain (DBD), the SWISS-MODEL Server was used based on the crystal structure of the highly similar E2 DBD - DNA complex of HPV6 lineage B3. The DNA complementary to the HPV11 E2 DBD was designed with the DNA Sequence to Structure tool of SCF Bio. For docking the HPV11 DBD and its respective variants to the complementary DNA, the NPDock server was used, with the protein and DNA interface constrained to the conserved DNA-binding site described. Lastly, the hinge region (amino acids 193–270) was de novo modelled using the I-Tasser server and the model with the highest C-score was subsequently joined with the TAD and DBD domains in UCSF Chimera to illustrate the full-length HPV11 E2 protein.

After model building, three approaches were used to analyse the difference between the prototype sequence and the sequence variants identified. Surface charge changes were calculated and compared; loss or gain of putative phosphorylation sites were predicted; and amino acid interactions within the protein chain, between chains or, when applicable, with the natural ligand were calculated and the free enthalpy differences were estimated and compared with Autodock Vina and the PreDBA server. For visualization, final models were depicted with UCSF Chimera X, and ligand interactions were

designed with LigPlot+. Lastly, the protein model geometry was validated with SAVES Ramachandran plot and was uploaded to the Protein Model Database. Putative phosphorylation sites on the E2 protein variants were predicted using the NetPhos 3.1. server.

Transformation

The HPV6 and HPV11 LCRs of the different strains were cloned into the BamHI and KpnI sites of the luciferase reporter vector pALuc. For functional analysis of E2 variants, the reference and unique E2 ORFs were cloned into pCDNA3.1+ mammalian expression vector using HPV11_E2_KpnI and HPV11_E2_XbaI primers.

The amplimers were purified from preparative agarose gel and digested by BamHI/KpnI and XbaI/KpnI restriction enzymes and ligated into pre-cut pALuc luciferase reporter vector or pcDNA3.1+ expression vector. The ligated constructs were transformed into *Escherichia coli* XL-1 the constructs were purified and checked by restriction digestion; then larger amounts of the corresponding clones were purified. All constructs were verified by sequencing (Macrogen, Amsterdam, the Netherlands).

Site-directed mutagenesis

In this study, one new, yet unexamined nucleotide polymorphism in the LCR was identified in a moderately severe papillomatosis (Patient 9, JO-RRP9/11). The effect of T7331G polymorphism on LCR activity was examined by site-directed mutagenesis. The mutagenized LCR plasmid was transformed into *Escherichia coli* XL1. The steps from transformation to transfection are described in the previous section. The construct was verified by sequencing (Macrogen, Amsterdam, the Netherlands).

Transient transfection and luciferase assay

To analyse the LCR activity of HPV6 and HPV11 sequences, HEp-2 (ATCC number CCL-23) laryngeal carcinoma cells were transfected with 2 μg of the reporter vector pALuc containing the unique LCR sequences and 1 μg of RSV-β-Gal plasmid as an internal control for transfection efficiency. For evaluation of the effect of E2 variability on the LCR activity, HEp-2 cells were co-transfected with 2 μg of reporter vector pALuc containing the reference or unique LCR variants and 2 μg of expression vector pCDNA3.1+ with reference or variant E2 sequences; 1 μg of RSV-β-Gal plasmid was used as an internal control for transfection efficiency. Cells were transfected in 6 cm diameter dishes with Lipofectamine 2000 according to the manufacturer's recommendation and were harvested 48 h posttransfection. The luciferase activity of the cell extracts was measured by the Luciferase Assay System. Transfection was standardized using the β-galactosidase assay by adding 1 mg/mL ortho-nitrophenyl- β-galactosidase chromogenic substrate and normalized to the protein

concentration as measured by the Bradford method. All sequences were tested in duplicates in three independent experiments.

Results

Complete genome sequences and phylogenetic analysis

Results of the investigation of the HPV6 complete genome sequences

Sequences from different samples of the same patients (AO-RRP1/6, AO-RRP3/6, AO- RRP5/6, JO-RRP1/6 and JO-RRP2/6) were always identical. Among the thirteen complete genomes, three (AO-RRP2/6, AO-RRP3/6, AO-RRP4/6) clustered together with the HPV6 lineage A reference, nine (JO-RRP1/6, JO-RRP2/6, AO-RRP1/6, AO-RRP5/6, AO-RRP6/6, AO-RRP7/6, CAC1/6, LSIL1/6, LP1/6) with the sublineage B1 reference and one (ASCUS1/6) sequence clustered with sequences of sublineage B2. Sublineage B3 (HPV6a), B4 and B5 did not occur.

In case of HPV6 lineage A and sublineage B1, size of all genomes was identical to the size of the reference genomes (7902 and 8012 bps, respectively); the sequence from sublineage B2 showed a 19 bp deletion absent in the 19 available sublineage B2 genomes. ORFs E2, E6 and E7 were highly conserved, while ORFs E5a, L1, L2 and the LCR showed higher variability.

In case of HPV6 lineage A, thirteen nucleotide positions were different between the reference sequence and the consensus; in addition, a 94-bp insertion was present in the LCR of the consensus sequence (between nucleotide positions 7250 and 7251), which is absent from the reference sequence or the sequences of this study. In case of sublineage B1 in contrast, the reference and the consensus differed only at a single nucleotide position. In case of sublineage B2 six nucleotide positions were polymorphic between the reference sequence and the consensus.

In case of HPV6 lineage A, 25 nucleotide polymorphisms were detected, the SNPs were found in ORFs E1, E5a, E6, L1 and L2 as well as in the noncoding region between E5b and L2 and the LCR. At the deduced protein level, nevertheless, these did not cause an amino acid alteration, except for two polymorphisms found in a single lineage A sequence in the E5a ORF.

At DNA level, 26 and eight SNPs were identified in case of HPV6 sublineage B1 and B2, respectively, compared to the reference genomes. In HPV6 sublineage B1, the variability of genomes was higher in case of sequences from juvenile-onset RRPs and from condyloma than in case of adult-onset RRPs (0.87–1.00 vs. 0.25–0.62 SNPs per 1000 bps). SNPs were found in the LCR as well as in all ORFs except for E7 in case of sublineage B1; in contrast to lineage A, four, one and three SNPs causing amino acid changes were detected in the E2/E4, L2 and L1 ORFs, respectively. All of these were unique except for Y219D in the L1 ORF, which was present in all sublineage B1 sequences from this study.

Thus, at the deduced protein level only one of the three lineage A sequences was different from the reference, while all sublineage B1 sequence variants exhibited at least one SNP leading to amino acid alteration compared to the reference genome. The sublineage B2 sequence differed from the B2 reference sequence in the E1, E5a, L1 and L2ORFs as well as in the LCR; of which only a single SNP in the L2 (H157N) resulted in amino acid change compared to the reference genome.

Results of the investigation of the HPV11 complete genome sequences

In virus genomes from the newly enrolled patients (Patients 7–15) 19 new single nucleotide polymorphisms (SNPs) were found, of which 17 were unique. Of the 19 SNPs, one was in the non-coding region between the ORFs E5B and L2, one was in the LCR and 17 in different ORFs. Seven resulted in amino acid alteration in ORFs E1, E2/ E4, E5A and L2. In addition, the virus genome from patient 11 (AO-RRP1/11) exhibited a 58-bp deletion in the E2/E4 ORF leading to a frameshift and an early stop codon. The deduced E2 protein is truncated (266 amino acid long), resulting in shorter DBD in which multiple amino acids are replaced as well. This deletion also affects the E4 ORF, leading to loss of a part of the proline-rich region and the complete positively charged region of the E1^E4 fusion protein.

In case of Patient 5 (JO-RRP5/11), the HPV11 genome was identical to the previously identified genome. In case of Patient 6 (JO-RRP6/11), HPV11 genomes originating from the papillomata and from exfoliated cells of the healthy oral mucosa collected in 2015 and 2016 were identical to the previously sequenced genomes. Similarly, multiple follow-up samples were obtained in case of Patients 8 (JO-RRP8/11) and 9 (JO-RRP9/11) yielding identical HPV11 genomes. The genomes determined from the three localizations (supraglottis, subglottis and stoma) of Patient 7 (JO-RRP7/11) were also fully identical. The genomes from Patient 8 (JO-RRP8/11) and Patient 14 (CAC/11) were identical. All other genomes exhibited a few SNP differences. The two already identified sequencing errors in the reference sequence (1783–1784 and 7719–7720) were consistently detected in all genomes.

The genome from CA/11 clustered together with the reference genome (sublineage A1), all other genomes belonged to sublineage A2. The genome JO-RRP7/11 clustered together with the genomes from more severe cases within sublineage A2 (including Patient 6, JO-RRP6/11).

In the E1 ORF three SNPs were determined, which causing amino acid alteration. Two unique polymorphisms A72E (JO-RRP4/11) and N100T (JO-RRP9/11), are located in the protein's N-terminal *in vivo* regulatory region, which is a poorly structured protein region, thus modelling was not possible. These polymorphisms do not overlap any putative nuclear localization signal. The sequencing error found in the E1 ORF corresponds to presence of alanine (position 318) instead of arginine translated in the incorrect sequence. Hexamer structures prepared with arginine 318 are

characterized with strong secondary bonds formed by this arginine with Trp322, Phe323 and Ile327 of the neighbouring monomer. In the models with the true amino acid sequence, the alanine present does not interact with any other amino acids in the E1 hexamer.

In the E5a ORF three amino acid alteration were determined, two of which (I28F, V41L) have been previously described by our research group and found in all Hungarian sequences expect CA/11. The probable localization of I28F is the N-terminal transmembrane region, while V41L is located between the first and second transmembrane regions. The D39N amino acid alteration was only found int AO-RRP1/11 and it seems it is located between the first and second transmembrane regions too.

No nucleotide polymorphism resulting in amino acid changes in the E6 ORF was identified. A single amino acid substitution (A45S) was detected in the E7 ORF region, which is present in all HPV11 genomes except for the HPV11 sequence from the CA / 11 sample. It is located in the C-terminal part of the protein, close to the second conserved region of the N-terminus (CR2) and involved in the formation of a putative phosphorylation site. The amino acid differences identified in the E5a ORF and E7 ORF, respectively, could not be modelled due to the lack of a known crystal structure.

New polymorphisms leading to amino acid alterations were not found int the L1 ORF, but three polymorphisms were detected at DNA level. The C6028T is present in all Hungarian samples except CA/11. Nucleotide polymorphisms were identified at position 6484 in three samples; thymine is replaced by cytosine at this position in two sequences (JO-RRP8/11 and CAC/11) and thymine is substituted by guanine in one genome (AO-RRP1/11). The nucleotide polymorphism that did not result in the last amino acid change was C6607T, which was detected in a sample from the JO-RRP7/11 patient.

Our research group described earlier two polymorphisms leading to amino acid alterations in the L1 (A476V and S486F); these were unique in the sequence from JO-RRP2/11. However, HPV11 sequence identified in the exfoliated cell sample from the healthy oral mucosa of the same patient, did not carry these polymorphisms causing amino acid changes in this L1 ORF.

The amino acid alteration E35D identified in the L2 ORF is localized in the actin-binding domain in the N-terminal region and is found only in the CA/11. The G5437T, detected only in a sample of AO-RRP2/11 patients, results in the formation of an early stop codon in the C-terminal region and the synthesis of a truncated protein (L1 binding domain is between 396 and 439 amino acids in HPV11 L2).

Functional analysis of LCRs

Results of the investigation of the HPV6s LCR functional analysis

In HPV6 lineage A sequences, LCRs exhibited six SNPs compared to the reference sequence, all

of which except for G7815C were present in all three genomes. Genome polymorphisms in the HPV6 lineage A LCRs influence the putative transcription factor binding sites. Transcription activation potential of all three sequences was statistically comparable to each other and to the reference LCR.

In case of HPV6 sublineage B1 genomes, sequences of seven of the nine LCRs were identical to the reference LCR, the remaining two exhibited SNPs; in JO-RRP2/6 an A7332C and in CAC1/6 A7342G and T7909G SNPs were found. These SNPs do not alter any putative transcription factor binding sites. Luciferase activity of three LCRs identical to the reference sequence was tested to determine the reference LCR activity, because reference HPV6 sublineage B1 genome containing plasmid was not available. The mean luciferase activities of the three tested LCRs were closely similar; the reference luciferase activity was defined as the mean of the luciferase activities measured for these LCRs. Activity of the JO-RRP2/6 was comparable to the reference activity. Activity of CAC1/6, in contrast, was significantly higher than either the referenceactivity or the activity of JO-RRP2/6.

The LCR of the single sublineage B2 genome carried a 19-bps deletion (between nucleotid positions 7361 and 7379) and two SNPs G7623T and T7645C compared to the B2 reference sequence. Curiously, with this 19-bps deletion, the size of its LCR is identical to that of sublineage B1 (compared to sublineage B1 reference four SNPs were found in the LCR, i.e. C7613G, T7626C, C7669G, C7900A). The deletion and the T7645C SNP led to loss of four and two putative transcription factor binding sites, respectively. LCR activity was comparable to that of sublineage B1 sequences, except for CAC1/6.

Comparing the LCR activity of HPV6 intratypic variants, lineage A LCRs showed significantly higher activity than sublineage B1 or B2 LCRs when comparing all individual sequence variants as well as when comparing means of HPV6 lineage A to means of HPV6 lineage B1 sequences.

Results of the investigation of the HPV11s LCR functional analysis

The difference from the reference genome deposited in the GenBank representing the sequencing error (GC insertion at positions 7719 and 7720), as confirmed by means of resequencing the reference plasmid, was found in all LCR sequences. One yet unreported novel polymorphism (T7331G) was found in the LCR of the sequence from JO-RRP9/11 (Pattern 6). This polymorphism did not alter significantly the LCR activity either in the original sequence or in the mutant sequence generated by site-directed mutagenesis

Seven out of the 15 LCRs were identical; five of these were derived from JO-RRPs and AO- RRPs of moderate severity, one from a severe JO-RRP and one from CAC, these sequences belong the LCR Pattern 1. LCR Pattern 2 is represented by two sequences from solitary papillomata (JO- RRP3/11 and 5/11) from our former study. LCR Pattern 3 was found in one genome from JO-RRP of moderate severity (JO-RRP1/11) in our former study and one further genome from this study from a severe JO-

RRP (JO-RRP7/11). LCR Patterns 4, 5 and 6 were unique from JO-RRPs (Patient 4/11, 6/11 and 9/11, respectively); the only difference between Pattern 6 and Pattern 3 was the newly found indifferent T7331G polymorphism. The sequence from CA/11 was identical to the reference LCR (reference pattern).

The highest luciferase activities were measured in case of the reference plasmid and in case of the identical LCR from the CA, which were statistically comparable (reference pattern). All other LCRs excepting JO-RRP6/11 (Pattern 5) showed significantly lower luciferase activity in all comparisons. Lowest activities were measured in case of LCRs from patients with solitary papillomata (JO- RRP3/11 and JO-RRP5/11) representing Pattern 2; these activities were significantly lower than those produced by LCRs belonging to Pattern 1 and Patterns 3–6.

E2 variants and interaction with the LCR

In the E2/E4 ORF, twelve SNPs were found, of which two, three and two caused amino acid change in the protein E2, E4 and both, respectively.

The unique polymorphism Q86K in E2 only found in the JO-RRP4/11, is located in the TAD causes a major alteration in the surface charge of the E2. The reference E2 (Q86) possesses a negatively charged surface and, while the surface of E2 with K86 is charged positively. Moreover, Q86 binds to K45 of the other E2 monomer chain, whereas K86 is exposed on the surface of the quaternary structure. Thus, in case of the Q86K polymorphism, the total binding energies between the two monomer TADs were slightly lower in the reference than in the mutant E2.

The polymorphisms S245F and N247T in E2 unique to JO-RRP9/11 are located in the hinge region. This region also plays a role in nuclear localization, but the polymorphisms do not affect directly the amino acids critical to nuclear transport. The polymorphisms do not change the electrostatic surface potential of the region. The polymorphism S245F disrupts a probable phosphorylation site in a RXXS motif targeted by protein kinase A and B, as revealed by phosphorylation site prediction. The polymorphism N247T may create a putativephosphorylation site, however, this site is predicted to be a poor target for phosphorylation.

The polymorphism K308R is present in the majority of sequences and is localised in the DBD. The K308 polymorphism does not alter DNA binding of the DBD directly, as the protein-DNA binding energies of the reference K308- and the R308-containing E2 variants are identical. However, the dimerization of the protein may be substantially altered by the K308 polymorphism; the side chain of K308 in the reference sequence faces the neighbouring F311 and E312 residues of the protein chain stabilizing the DNA-binding helix. In case of the K308R polymorphism, the bulkier arginine side chain faces the other protein chain in the dimer, bonding to its S322. The resulting hydrogen bonds increase the total binding energy between the two E2 monomers markedly.

The deletion causing frameshift drastically alters the C-terminal region of the E2, disturbing the stabilizing beta strands and changing the DNA binding motif from **N**CLK**C**FR**YR**LN to **V**SS**T**VR**EV** (putative DNA-binding residues highlighted as bold). This leads to loss of positively charged amino acids and changes the net surface charge to slightly negative. Taken together, these alterations are expected to lead to loss of DNA binding as well as of the capacity to form dimers.

Out of the seven polymorphisms in the E2/E4 ORF, five as well as the deletion also affects the E1^E4 fusion protein. The polymorphism Q46R (Q28R in E1^E4 fusion protein) (JO-RRP1/11, JO-RRP2/11, JO-RRP3/11, JO-RRP5/11, JO-RRP7/11, JO-RRP8/11, JO-RRP10/11, AO-RRP2/11, AO-RRP3/11, CAC/11) affects the proline-rich region; the polymorphism G61E (G43E in the fusion protein) and S78L (S60L in the fusion protein) (present in all sequences excepting CA/11, JO-RRP6/11 and AO-RRP1/11) are localised in the loop and in the negatively charged proline-rich region, respectively. The polymorphisms P68S and T70P (P50S and T52P in the fusion protein, respectively), both in the negatively charged proline-rich region, are unique to JO-RRP9/11.

The deletion in the E2/E4 ORF (AO-RRP1/11) leads to loss of amino acids 25 to 43 of the E1- E4 fusion protein affecting the proline-rich, positively charged regions and the N-terminal part of the loop, which probably leads to loss of E4 folding mediated by interactions between the proline-rich and the negatively charged proline-rich regions.

For the co-transfection experiments we determined the variant groups of the E2 considering the amino acid alterations. This outlines five E2 variants at amino acid level, i) the reference sequence, CA/11 and JO-RRP6/11 (Reference variant); ii) the SNP K308R characterizes JO- RRP1/11, JO-RRP2/11, JO-RRP3/11, JO-RRP5/11, JO-RRP7/11, JO-RRP8/11, JO-RRP10/11, AO-RRP2/11, AO-RRP3/11, CAC/11 (Variant 1); iii) AO-RRP1/11 with the deletion and the early stop codon (Variant 2); iv) SNPs Q86K and K308R are simultaneously present in JO-RRP4/11 (Variant 3); v) three SNPs, S245F, N247T and K308R are present in JO-RRP9/11 (Variant 4).

In co-transfection experiments, the Reference Variant, Variant 3 and 4 of E2 increased luciferase activity of the reference LCR significantly; Variant 1 led to obvious but not statistically significant increase, while the truncated Variant 2 E2 did not alter LCR activity. Comparingvariants, reference E2 led to the highest activity of the reference LCR, followed by Variants 3 and 4, comparable to each other, but higher than Variants 1 and 2.

Reference E2 increased the activity of all LCR patterns. Patterns 1 and 2 were characterized by only a minor increase, while the reference LCR and patterns 3–6 were enhanced significantly more by the reference E2.

Examining the existing combinations of the polymorphisms of LCRs and E2s found in the genomes, E2s which increased the activity of the reference LCR (the reference variant and Variants

3 and 4), showed a similar effect when tested with their corresponding LCRs. As expected, the truncated E2 (Variant 2) did not affect the activity of its LCR. However, E2 Variant 1, which was the most common among the examined sequences, behaved differently with the different LCRs tested. Variant 1 when combined with an LCR with intrinsically lower enhancer capacity (Patterns1 and 2) results in marginally increased LCR activity, which does not significantly exceed the basic activity of the reference LCR. When combined with a more potent LCR (Pattern 3), the activity is significantly higher than the LCR activity alone, but still significantly lower than LCRs potentiated by the reference E2.

Discussion

Although diseases caused by HPVs with a low oncogenic risk are usually mild, in some cases, the course of the disease can significantly impair quality of life or sometimes a life-threatening condition may develop. While differences in intratypic variance and course of the disease in case of high-risk HPV genotypes become known increasingly, the virulence mechanisms to explain differences in disease course in low-risk genotypes are largely unknown. Previous studies by our study group on HPV11 have shown that nucleotide polymorphisms in the LCR may affect its transcriptional regulatory effect, so their occurrence may be related to the severity of the disease caused by the given variant.

The results of our phylogenetic analysis of HPV6 complete genome sequences - similarly to the results obtained by Jelen et al. (2014) during the analysis of European genomes – showed the dominance of the phylogenetic group B1 (formerly HPV6vc) and the significantly lower incidence of the B3 sublineage (formerly HPV6a) among HPV6 complete genome sequences. In a South American study, the cumulative incidence of sublineage B1 was also described and the sublineage B3 was identified as the second most common sublineage; these results are also consistent with the distribution data for South America obtained by our study and by Jelen et al. (2014). A further similarity between our results and the South American results is that a significant proportion of the HPV6 sequences were identified from adult-onset papillomatosis. In a Japanese study, the sublineage B1 was also found as the most common phylogenetic group. In the present study, due to the low number of cases, no association can be established between the incidence of phylogenetic groups and the gender of the patients, but male dominance can be observed in case of adult-onset disease.

HPV6 genome was stable in consecutive samples of the patients; similar findings were also reported unequivocally by earlier studies. Low variability in a given host was also observed in case of HPV11, but the opposite was reported for HPV16 by some authors. Thus, low-risk types seem to be markedly conserved between temporally distant virus sequences, while in high-risk types within-

host variability was observed. This genome stability, at the same time, confirms that the symptomatic episodes are linked to reactivation of the virus in the patient rather than reinfections.

Major genome rearrangements are rarely reported in low-risk HPVs and are related to unfavourable outcome, malignant transformation or extreme high number of recurrences. Notably, such genome rearrangements are not needed for progression of the malignancies caused by high-risk types, where tumour progression was repeatedly found to be associated with accumulation of point mutation in the viral genome, especially in E6 ORF. In contrast, SNPs causing amino acid change were never found in E6 in the sequences reported in this study. This is in line with the findings in Hungarian HPV11 genomes as well. These suggest that E6 may be more conserved in low-risk types than in high-risk HPVs. The conserved nature of E7, in contrast, seems to be a shared characteristic between the two groups of HPVs.

Other early and late ORFs were more variable among the HPV6 sequences reported in this study; moreover, the variability of these ORFs and the LCR showed a notable difference between (sub)lineages. In our three HPV6 lineage A genomes almost all SNPs were synonymous, and the sequences were highly similar. Differences between the reference sequence and the consensus, especially in the LCR, suggests that the lineage A reference sequence may be a unique variant and does not represent the wild type similarly to the HPV11 reference sequence. In contrast, our HPV6 sublineage B1 sequences were diverse and were positioned at various distances to the reference sequence, which was highly like the consensus.Most SNPs in sublineage B1 sequences were unique, many of which result in amino acid changes. Sequences with these unique SNPs originated from RRPs with several episodes of recurrence (5–8) and/or mild to moderate dysplasia in the papilloma/condyloma tissue, while the two sequences with only the ubiquitous L1 polymorphism showed only a single recurrence after removal of the initial papilloma.

Only a single lineage A sequence (AO-RRP4/6) showed SNPs leading to amino acid change, both in the E5a. The SNP E39D (notably, the amino acid in this position of HPV11 is D) corresponded to the loop between the putative first and the second transmembrane domains of HPV16 E5, which is localized probably on the extracellular part of the protein, while the other (P78S) created a potential phosphorylation site mapping to the putative third transmembrane domain of HPV16 E5. The probably extracellularly localized E39D may influence growth factor binding, thus may contribute to the moderate dysplasia found in the papilloma of this patient.

ORFs showing the highest diversity were E2/E4, L1 and L2, suggesting that individual viruses may exhibit differences in the function of these proteins, which, in turn, may explain, at least partly, the higher number of recurrences; similarly to findings reported for HPV11.

In HPV6 sublineage B1 sequence variants, E2/E4, L1 and L2 ORF showed notable variability

at the deduced protein level. Two polymorphisms in the E2/E4 ORFs affected the transactivation domain (T116N in AO-RRP1/6 and S144T in JO-RRP2/6) of the E2 regulatory protein; one polymorphism was in the hinge region (S246A in JORRP1/6), eliminating a potential phosphorylation site in this phosphorylation hot-spot of the protein; and one (E340D in AO-RRP5/6) was in the DNA-binding and dimerization domain. As inferred from their position, these polymorphisms may influence E1 binding and consequently initiation of DNA replication as well as transcriptional activation; intracellular localization, chromatin binding and self-regulation; or L1 binding and regulating oncoprotein expression through LCR binding of the E2, respectively. The polymorphism S246A also affected the E1/E4 fusion protein as S50R (corresponding to S68R when directly translating the E4 ORF alone) localized in the negatively charged proline rich region.

The unique polymorphisms F441L in patient CAC1/6 and K449E in patient JO-RRP1/6 are localized in the C-terminus invading arms linking L1 capsomers, thus in the region involved in capsid assembly and encapsidation as well as in heparan sulphate binding, which also contains several epitopes. Remarkably, heparan sulphate binding is linked to lysine residues in HPV16, thus the loss of lysine (K449E) may impair this binding.

The localization and the inferred functional characteristics of these polymorphisms suggest that variations may exist between individual viruses in regulation of viral transcription and/or replication (E2), virus release (E4), capsid assembly (L1, L2) or interaction with the host (E5a, L1, L2).

Sequence variation of the LCR across the HPV6 lineages showed differences markedly similar to those found in the variability of ORFs. While two HPV6 sublineage B1 LCRs contained altogether three SNPs, in case of lineage A the differences are shared by all three sequences. In line with this, transactivating potential of our lineage A LCRs was highly similar to each other as well as to the reference. In case of sublineage B1, LCRs of seven of nine viruses were identical, a LCR with the unique SNPs A7342G and T7909G (Patient CAC1/6 with six episodes) showed a significantly increased activity compared to other sublineage B1 LCRs. This pattern reflects what was reported by Measso do Bonfim et al. (2015) concerned with the link between LCR sequence differences and transactivating potential. Existence of HPV6 and HPV11 variants with partially duplicated LCRs and consequent higher transactivating potential, which caused more aggressively spreading disease lend further support to this hypothesis. However, Grassmann and his research group did not report differences in activities of LCR sequence variants, as also seen with the polymorphic LCR in JO-RRP2/6 with an activity comparable to the reference activity. Thus, as also reported in HPV11, polymorphic sites may or may not cause significant alterations in LCR activity.

Measso do Bonfim and his research group (2015) also reported a major difference between the LCR activity of HPV6 sublineages; HPV6 sublineage B3 (HPV6a) was found to exhibit lower activity

than HPV6 sublineage B1. Notably, transactivating potential of HPV6 lineage A in our study was significantly higher than that of sublineage B1 sequence variants and of the lineage B2 sequence, except for the mutant LCR of Patient CAC1/6, which has statistically similar, though somewhat lower, transactivating potential.

This, together with our data suggests that the LCR activity shows a gradient from the highest HPV6 lineage A through sublineage B1 to sublineage B3 showing the lowest activity.

These taken together suggests that there are differences in the ways to achieve cell proliferation between the HPV6 lineage A and sublineage B1, possibly between the two main HPV6 lineages A (represented by HPV6b) and B (represented by HPV6 sublineage B1 and the single B2 sequence in this study). HPV6 lineage A seems to rely more on transactivating potential than sublineage B1; the possible greater importance of LCR activity is also supported by the frequent alteration of transcription binding sites by the SNPs in lineage A but not in sublineage B1 LCR. Similarly, such differences between clusters were also suggested for HPV11.

HPV6 sublineage B1 also shares some similarities with HPV11 generally accepted to be more aggressive. Both JO-RRPs were caused by HPV6 sublineage B1, while HPV6 lineage A genomes were found exclusively in AO-RRPs. E2/E4, L2 and L1 ORFs contained unique SNPs. Some genomes had unique LCRs, which showed increased transactivating potential as compared to that of the reference LCR. Nevertheless, differences in the disease course or severity among the patients infected with either detected subtypes of HPV6 were much less prominent than those reported among HPV11s in a similarly small patient population.

The LCRs of the HPV11 genomes from the new patients enrolled in this study exhibited mainly these known polymorphisms; only one novel polymorphism was detected, which did not affect LCR activity. Similarly, most polymorphisms found in the coding region were silent polymorphisms identified earlier, being ubiquitous in the sequences from this and from our former study as well as in many other genomes in the GenBank. However, some newly identified polymorphisms causing amino acid alteration were unique with presumable effects on virus physiology.

A remarkable example of the role of unique variants is the HPV11 genome from AO-RRP1/11 with the truncated and frameshifted E2 ORF. As expected, this severely altered E2 showed inability to increase the LCR activity substantially. The transactivating potential was comparable to that measured for the interaction of the Pattern 2 LCR (two simultaneous attenuator polymorphisms; and E2 Variant 1 (the most common E2 variant with K308R). The dysplasia found in this papilloma may be explained by losing the regulatory ability of E2 on expression of oncoprotein ORFs. Genome integration was concluded by studies of a few cases of HPV11-associated cancers, but disruption of E2 was not studied. The deletion may have been caused by integration of the HPV11 genome into the

host cell genome; it is generally accepted in case of high-risk genotypes that the virus genome opens up in the E1-E2 region.

The unique polymorphism Q86K in the E2 (JO-RRP4/11) may aid dimerization slightly and markedly alters the surface charge of the transactivating domain. As E2 is active as dimer, the moderate increase in the efficiency to enhance the activity of the LCR as compared to the E2 variants without this polymorphism may be linked to the easier dimerization. However, this is compensated by the LCR of lower intrinsic activity; the net effect is a moderately severe papillomatosis.

Unique variants may involve loss and/or gain of phosphorylation sites as by the polymorphisms S245F and N247T, the former disrupting a highly probable protein kinase A/B phosphorylation site, the latter creating a less probable site for glycogen synthase kinase 3. This obviously would alter the ability of E2 to interact with cellular regulatory pathways. However, these polymorphisms did not modify the effect on the LCR; which is in line with the severity of the case (JO-RRP9/11) comparable to cases with a similar LCR (e.g., JO-RRP1/11). Loss or alteration of a phosphorylation site may act through a different mechanism, as E2 phosphorylation may be mediated by various protein kinases and phosphorylation modifies interactions between cellular proteins and E2. The site S245F corresponds to phosphorylation site S253 of HPV8 and S243 of HPV16, which were shown to increase half-life of E2 and to be involved in host chromatin binding when phosphorylated. The mutant S245F cannot be phosphorylated by the protein kinases A and B phosphorylating wild-type E2, leading probably to impaired chromatin binding and shorter half-life of E2, which may reduce the capacity of the virus to be transmitted to progeny host cells thus limiting within-host spread. Though the polymorphism N247T creates a putative phosphorylation site very close to S245F, this is partially hidden in the inner part of the helix, making it less likely toserve as a phosphorylation site. On the other hand, it is predicted to be targeted by a differentkinase, thus may be linked to different regulatory pathways, therefore, it cannot be considered as a site equivalent to the lost S245.

The pattern of non-unique polymorphisms found supports the previously raised assumption that the reference sequence represents a particularly aggressive unique variant, and the wild type HPV11 may be closer to the sequence variant group represented by JO-RRP8/11 and CAC/11 in this study. Along this line, the prototype sequence possesses multiple unique polymorphisms. E.g., the attenuator polymorphism (T7547C) in the LCR reported by Gáll and his research group. may be thewild type and the polymorphisms in the prototype sequence from a severe papillomatosis is inreality enhances LCR activity. This LCR sequence was found in a single novel genome from cervical atypia (CA/11) characterized by a similarly high basic LCR activity. This alone may explain the increased severity of the disease caused.

Similarly, the ubiquitous A45S polymorphism in the E7 protein is rather a unique polymorphism

of the prototype which leads to loss of a putative phosphorylation site in thereference sequence, while the presumed wild-type HPV11 as well as the closely related HPV6 contains serine in this amino acid position. Curiously, the corresponding amino acid is also alanine in HPV16.

Two ubiquitous polymorphisms in the E5A protein (I28F and V41L) also suggest that the reference sequence (along with CA/11) represents an uncommon variant. It is tempting to assume that these may also contribute to the increased severity of the disease caused, since E5A is accepted widely as the oncoprotein with the most prominent role in pathogenesis of diseases caused by low-risk HPVs.

A similarly ubiquitous polymorphism in the E2 is the K308R, which is present in all except one sequence, suggesting that the wild type is the arginine, which creates a strong link between the chains stabilizing the E2 dimer, while the lysine in the reference sequence rather aids in stronger DNA (LCR) binding. This presumable stronger binding is reflected by the marked increase in transactivation caused by this E2 variant on practically all LCR sequence patterns. Accordingly, this may contribute to the increased severity of the diseases caused by the HPV11s containing this E2 sequence variant (reference sequence, CA, JO-RRP6/11).

In addition to the ubiquitous attenuator LCR polymorphism (T7547C), LCRs from JO- RRP3/11 and JO-RRP5/11 contain an additional attenuator polymorphism (T7509Δ), leading to severely impaired LCR activity. Their E2 variant was also associated with low capacity for LCR upregulation leading to the lowest LCR activity measured among sequences with fully functional E2 and LCR. These papillomata remained solitary until now.

When a less potent LCR was enhanced by a moderately effective E2, the resulting LCR activity was comparable to that of a typical LCR with a defective E2, both resulting in a papillomatosis of low severity. This confirms that the LCR activity is the net effect of the interplay between the intrinsic transactivating potential of the LCR and the capacity of E2 to upregulate LCR activity. The highest LCR activities were exhibited by the most potent LCR pattern (reference pattern, from a highly aggressively spreading JO-RRP) and E2 variant (reference E2). A slightly less potent LCR (JO-RRP6/11; Pattern 5) when upregulated by the reference E2 still showed high activity and was associated with a severe disease with >60 recurrences. Many cases of moderate severity (2–10 recurrences) yielded HPV11 genomes containing the most frequent LCR pattern of moderate intrinsic transactivating potential combined with the most frequently found E2; this interplay results in a moderate transactivating potential.

Summarizing the results of our phylogenetic studies - in line with the results of other research groups – in case of HPV6 and HPV11 sequences the B1 and A2 sublineages are dominant among our sequences, respectively, and these sublineages show the highest frequency worldwide.

Examination of our HPV6 complete genome sequences revealed several similarities and differences in the characteristics of sublineages: i.) the number of nucleotide polymorphisms identified are comparable in our lineage A and sublineage B1 sequences; ii.) however, in the case of sublineage B1 mostly unique polymorphisms were identified, in case of lineage A the major part of the polymorphisms were present in a significant proportion of the examined samples, iii.) in contrast to lineage A in which usually synonymous polymorphisms were detected, in case of sublineage B1 several nucleotide polymorphisms cause amino acid changes compared to the references; iv.) more nucleotide polymorphisms were identified in the LCR of lineage A than in sublineage B1; v.) the HPV6 sequences in lineage A were detected only in a sample of AO-RRP patients, while the sequences in sublineage B1 were derived from a more heterogeneous group of samples (AO-RRP, JO-RRP, CAC, LP, CA).

In conclusion, it seems that differences in the phylogenetic (sub)groups of HPV6 – similarly to high-risk HPV genotypes - may influence the occurrence of variants in different diseases/patient groups, although further studies with a larger study population are required to confirm this hypothesis. Sequence analysis of the HPV6 and 11 variants shows that the E6 and E7 oncoproteins are highly conserved in HPV11 and HPV6, respectively, suggesting that the role of the oncoproteins is regulated very strictly and able to tolerate only minimal structural changes. This is in line with the fact that there is a significant difference in the regulation of replication and transcription of low- and high-risk HPV genotypes. The E6 and E7 oncoproteins of high-risk HPVs play a key role in cell transformation, and the variability of these proteins determines the oncogenic risk of genotypes and subtypes. In contrast, it seems that in the background of the different pathogenicity/virulence of the low-risk HPV genotypes/variants rather the variability of the E1, E2/E4, and E5a, as well as capsid proteins and LCR may be identified than the variability of the main oncoproteins.

Summary

The etiological role of low-risk HPV6 and 11 has been demonstrated in the development of many benign diseases, including recurrent respiratory papillomatosis, but the association between disease severity and genetic characteristics of the causative agents is even less well known. However, it is also increasingly known that HPV genotypes with high and low oncogenic potential show basic differences in the course of the diseasecaused and in the underlying genetic variability.

The aim of our work was to perform the complete genome and phylogenetic analysis of HPV6 and HPV11 from newly enrolled patients as well as new recurrences from already known patients. The potential effect of amino acid changes on the structure and function of viral proteins was also investigated by *in silico* protein modelling. In transient transfection experiments, we performed

functional analysis of different LCR patterns and examined the effect of E2 protein variants on LCR activity.

Based on the phylogenetic analysis of HPV6 genomes, our sequences belonged to the lineage A and to the sublineages B1 and B2; sequences belonging to the sublineage B1 were the most common. The results of the sequence analysis showed that nucleotide substitutions in HPV6 sublineage B1 sequences resulted in amino acid changes more frequently than those of HPV6 genomes classified as lineage A. In contrast, in case of lineage A sequences the LCRs showed higher variability. In the functional analysis of the HPV6 LCRs, a gradient-like decrease in the LCR activity of the variants was observed; the highest activity was related to the HPV6 lineage A variants, followed by the LCR activity of the phylogenetic sublineages B1 and finally B3. These results support the hypothesis that there is a difference among the effect of each HPV6 (sub)lineages in the modulation of the cell proliferation. In contrast to sublineage B1, the lineage A LCR has more efficient transcription activating effect. The greater variability of the transcription factor binding sites identified in the HPV6A lineage may also support this.

In case of HPV11 genotype, both the ORFs and the LCR contain many polymorphisms in comparison to the reference sequence; greater variability was observed for E1, E2/E4, and E5a as well as regions encoding capsid proteins and the LCR. Based on the phylogenetic analysis, the dominance of the sublineage A2 was detected among our HPV11 sequences.

The K308R amino acid change identified in the E2 ORF is capable of forming strong interaction between the stabilizing chains in E2 dimers but does not influence the DNA binding, resulting in this E2 variant with less efficient enhancer activity than the reference E2. The unique amino acid alteration Q86K changed the negative surface charge of E2 (Q86) to positive (K86). The unique alterations S245F and N247T in the hinge region of the protein cause changes in the phosphorylation pattern. The 58 bp long deletion identified in the E2/E4 ORF results in a frameshift and the formation of an early stop codon, which lead to the formation of truncated and non-functional E2 protein; the loss of E2 function may have contributed to the development of papillomatosis with dysplasia.

Co-transfection experiments with different LCR and E2 combinations confirm that the basic transcription activating effect of the LCR and the enhancer activity of the E2 protein together determine the gross transactivating effect of the LCR.

In conclusion, based on our results, it appears that in case of intratypic variants of low-risk HPV genotypes - similarly to high-risk ones – such genetic changes can be identified which can cause differences in the pathogenicity/virulence of the virus variants. In the background of these differences rather the variability of the E1, E2/E4, and E5a, as well as capsid proteins and LCR maybe identified than the variation of the main oncoproteins.

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/85/2022.PL
Subject: PhD Publication List

Candidate: Zsófia Nagy
Doctoral School: Doctoral School of Pharmacy
MTMT ID: 10053201

List of publications related to the dissertation

1. **Nagy, Z.**, Pethő, Z., Kardos, G., Major, T., Szűcs, A., Szarka, K.: Effect of E2 and long control region polymorphisms on disease severity in human papillomavirus type 11 mediated mucosal disease: protein modelling and functional analysis.
   Infect. Genet. Evol. 93, 1-13, 2021.
   DOI: http://dx.doi.org/10.1016/j.meegid.2021.104948
   IF: 3.342 (2020)
2. Szinai, M., **Nagy, Z.**, Máté, P., Kovács, D., Laczkó, L., Kardos, G., Sápy, T., Szűcs, A., Szarka, K.: Comparative analysis of human papillomavirus type 6 complete genomes originated from head and neck and anogenital disorders.
   Infect. Genet. Evol. 71, 140-150, 2019.
   DOI: http://dx.doi.org/10.1016/j.meegid.2019.03.019
   IF: 2.773

DEBRECENI EGYETEM * Egyetemi és Nemzeti Könyvtár *

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

List of other publications

3. Major, T., Szarka, K., **Nagy, Z.**, Kovács, I., Balog, C., Karosi, T.: Gyulladás vagy daganat?: egy hirtelen jelentkező lateralis cysticus nyaki terime tanulságai.
   Orv. hetil. 162 (15), 595-600, 2021.
   DOI: http://dx.doi.org/10.1556/650.2021.32087
   IF: 0.54 (2020)
4. Király, R., Thangaraju, K., **Nagy, Z.**, Collighan, R., Nemes, Z., Griffin, M., Fésüs, L.: Isopeptidase activity of human transglutaminase 2: disconnection from transamidation and characterization by kinetic parameters.
   Amino Acids. 48 (1), 31-40, 2016.
   DOI: http://dx.doi.org/10.1007/s00726-015-2063-5
   IF: 3.173

Total IF of journals (all publications): 9,828
Total IF of journals (publications related to the dissertation): 6,115

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

17 February, 2022

DEBRECENI EGYETEM Egyetemi és Nemzeti Könyvtár