Infection, Genetics and Evolution 93 (2021) 104948

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Effect of E2 and long control region polymorphisms on disease severity in human papillomavirus type 11 mediated mucosal disease: Protein modelling and functional analysis

Zsófia Nagy [a], Zoltán Pethő [b,c], Gábor Kardos [a], Tamás Major [d], Attila Szűcs [e], Krisztina Szarka [a,*]

[a] Department of Medical Microbiology, Faculty of Medicine, University of Debrecen, Nagyerdei krt.98, H-4032 Debrecen, Hungary
[b] Department of Biophysics and Cell Biology, Faculty of Medicine, University of Debrecen, Nagyerdei krt.98, H-4032 Debrecen, Hungary
[c] Institute of Physiology II, University Muenster, Robert-Koch-Str. 27B, 48147 Münster, Germany
[d] Otorhinolaryngology and Head-Neck Surgery Division, Kenézy Gyula Teaching Hospital, University of Debrecen, Bartók Béla út 2-26, H-4031 Debrecen, Hungary
[e] Otorhinolaryngology and Head and Neck Surgery Clinic, Faculty of Medicine, University of Debrecen, Nagyerdei krt. 98, H-4032 Debrecen, Hungary

ARTICLE INFO

Keywords:
Low-risk human papillomavirus
Recurrent respiratory papillomatosis
Genome polymorphisms
Long control region
Transactivating potential

ABSTRACT

Interaction of the long control region (LCR) and the E2 protein of HPV11s was studied by *in silico* modelling and *in vitro* functional analysis.

Genomes of HPV11s from fifteen (six known and nine novel) patients (two solitary papillomas, eleven respiratory papillomatoses of different severity, one condyloma acuminatum and one cervical atypia) were sequenced; E2 polymorphisms were analysed *in silico* by protein modelling. E2 and LCR variants were cloned into pcDNA3.1+ expression vector and into pALuc reporter vector, respectively, transfected to HEp2 cells alone or in different combinations and the luciferase activity was measured.

In the E2, the ubiquitous polymorphism K308R caused stronger binding between the dimers but did not alter DNA binding; E2s with this polymorphism were significantly less efficient than the reference in promoting LCR activity. The unique polymorphism Q86K changed the negative surface charge of E2 (Q86) to positive (K86). The unique polymorphisms S245F and N247T in the hinge region disrupt a probable phosphorylation site in a RXXS motif targeted by protein kinase A and B, but do not affect directly the amino acids critical to nuclear transport. Both unique patterns partly restored the LCR activating potential disrupted by K308R. A unique E2/E4 ORF with a 58-bp deletion leading to a frameshift and an early stop codon resulted in a practically nonfunctional E2, and was associated with a papillomatosis with dysplasia.

When testing existing LCR-E2 combinations, LCR with intrinsically lower enhancer capacity was only marginally activated by its E2 (R308 and the deletion mutant), and did not significantly exceed the activity of the reference LCR without E2. Combined with more potent LCRs associated with more severe disease, the activity was significantly higher, but still significantly lower than LCRs with reference E2.

In summary, LCR-E2 interaction determined by their polymorphisms may explain, at least partly, differences in disease severity.

1. Introduction

Recurrent papillomatosis may severely decrease the quality of life and may have potentially fatal complications, *i.e.* spread to the lungs inducing recurrent pneumonias or malignant transformation to squamous cell cancer (Hermann et al., 2012; Tjon Pian Gi et al., 2015). Two types are distinguished, juvenile- (JO-RRP) and adult-onset (AO-RRP) recurrent respiratory papillomatosis (Campisi et al., 2010; Larson and Derkay, 2010; Derkay and Bluher, 2019).

The pathogenesis of respiratory papillomatosis is only partially understood. Oncoproteins E6 and E7, which are crucial in the carcinogenesis induced by high-risk types, seem to play only a minor role in papilloma formation, which is attributed rather to the E5A protein activating growth factor receptors (DiMaio and Petti, 2013; Johnston et al., 1999; Vambutas et al., 1993). Efficiency of this process may be modulated by a number of viral and host factors. The transactivating

* Corresponding author.
E-mail address: szkrisz@med.unideb.hu (K. Szarka).

https://doi.org/10.1016/j.meegid.2021.104948
Received 7 April 2021; Received in revised form 20 May 2021; Accepted 30 May 2021
Available online 2 June 2021

potential of the long control region (LCR) variants as well as the regulatory role of E2 protein may directly influence E5A (as well as other oncoprotein) levels. Replication efficiency conferred by E1 and E2 proteins (Bergvall et al., 2013; McBride, 2013), speed of assembly determined by capsid proteins L1 and L2 (Buck et al., 2013; Wang and Roden, 2013) and efficacy of virus release modified by the E4 protein may affect the speed of virus production (Doorbar, 2013), thus the number of virions available for infecting new cells. Finally, interaction of the L1 proteins with host immunity may also affect the course of infection by modulating virus levels (Buck et al., 2013) and within-host spread. These assumptions are in line with the results that unique polymorphisms were found in E1, E2, E4 and L1 ORFs, but not in E6 or E7 in HPV11s from more severe papillomatoses (Gáll et al., 2013).

Though it was reported that patients with papillomata show Th2-dominant immune response to HPV; the explanation for differences in severity are largely lacking (Bonagura et al., 1999; DeVoti et al., 2004; DeVoti et al., 2008; Rosenthal et al., 2012). Earlier results implicated differences in transactivating potential of the different sequence variants of LCR as a factor possibly contributing to differences in the clinical course of the disease (Gáll et al., 2013), but polymorphisms of the coding regions scarcely received any attention.

The aim of this study was to collect more data on the role of LCR and E2 interactions in the severity of respiratory papillomatosis by investigating new samples of reported patients (Gáll et al., 2013), by enrolling new patients and examining the genome of their HPV11s and by analysis of the impact of polymorphisms in E2 and LCR on pathogenesis using *in silico* modelling and *in vitro* functional analysis of their interaction.

2. Materials and methods

2.1. Patients, clinical samples and viruses

Samples were collected between 2006 and 2018 at the Departments of Otorhinolaryngology and Head & Neck Surgery and Obstetrics and Gynecology of the University of Debrecen; all samples originated from the EU regions Northern Great Plain and Northern Hungary. Of the 38 histopathologically confirmed respiratory papillomatoses 35 were HPV positive (17 HPV6, 16 HPV11 and 2 HPV16); the whole genome could be determined in case of 13 HPV11 and 13 HPV6. Presence of multiple genotypes were never detected. Characteristics of HPV6s were reported earlier (Szinai et al., 2019).

HPV11 complete genome sequences from six patients (Patient 1–6) suffering from HPV-associated juvenile-onset recurrent respiratory papillomatosis (JO-RRP) were reported earlier (Gáll et al., 2011; Gáll et al., 2013); tissue specimens from one and two recurrences of Patient 5 and 6, respectively, as well as exfoliated epithelial cells from Patient 6 were also examined. Further four patients with HPV11-associated JO-RRP (Patients 7, 8, 9 and 10) and three patients with adult-onset disease (AO-RRP) (Patients 11, 12 and 13) were newly enrolled. For comparison, two patients with genital HPV11, Patient 14 with condyloma acuminatum (CAC) and Patient 15 with cervical atypia (CA), were also investigated. In case of Patients 5, 6, 8 and 9, multiple episodes were sampled for complete genome sequencing, while in case of Patient 7, HPV11 genomes from three different localizations (supraglottic, subglottic and site of the tracheostoma) were sequenced. Written informed consent was collected from each patient. The study received ethical committee approval (approval number: 4169–2014). Patient characteristics and clinical data are summarized in Table 1.

Table 1
Clinical characteristics of patients diagnosed with diseases associated with human papillomavirus type 11.

Patient ID	Disease	Gender	Accession Number	Age at the time of diagnosis	Localization	No. of surgical interventions	Tracheostomy	Extralaryngeal dissemination	Adjuvant therapy
Patient 5[a]	JO-RRP2	F	HE574702	3	larynx, pharynx, sinuses, nasal cavity, trachea	30<	yes	yes	IFN
Patient 6[a]	JO-RRP6	M	FR872717	1.5	larynx, soft palate	60<	yes	yes	IFV, CDV
Patient 7[b]	JO-RRP7	F	MK313764	8	larynx, pharynx, sinuses, nasal cavity, trachea	40<	yes	yes	IFN, CDV
Patient 8[c]	JO-RRP8	M	MK313765	4	larynx	4	no	no	no
Patient 9[c]	JO-RRP9	F	MK313766	8	larynx	5	no	no	no
Patient 10	JO-RRP10	F	MN788368	3	nasal cavity	3<	no	no	no
Patient 11	AO-RRP1	M	MK313761	48	nasal cavity	2	no	no	no
Patient 12	AO-RRP2	F	MK313762	29	larynx	1	no	no	no
Patient 13	AO-RRP3	M	MW404328	29	larynx	4	no	no	no
Patient 14	CAC	F	MK313767	57	vulva	1	N.A.	N.A	no
Patient 15	CA	F	MK313768	25	cervix uteri	Sampling for cytology testing and HPV detection	N.A:	N.A	no

JO-RRP juvenile-onset recurrent respiratory papillomatosis; AO-RRP adult-onset recurrent respiratory papillomatosis; CAC condyloma acuminatum; CA cervical atypia; F female; M male; IFN interferon; CDV cidofovir; N.A. not applicable.
None of the patients had malignant transformation. Mild or moderate dysplasia was detected in case of Patients 6, 7, 8, 10 and 13.

[a] Tissue samples newly collected from new recurrences were analysed and were identical with previously described genomes (Gáll et al., 2013).

[b] Samples collected from multiple localizations (supraglottic, subglottic and site of tracheostoma) were analysed; identical HPV11 genomes were found.

[c] Multiple samples from different recurrences were analysed and proved to be identical.

Patients with RRP and CAC were sampled by reserving a small piece of excised tissue for virological investigation during surgical intervention. Exfoliated cells were collected from patients with cervical atypia using a cytobrush into 1 mL phosphate buffered saline (PBS). Samples were stored at −70 °C until investigation.

DNA was isolated by InnuPrep Viral DNA/RNA Kit (Analytic Jena, Jena, Germany). For nucleic acid extraction, fresh-frozen tissue biopsies were homogenized in 400 µL PBS. First, all samples (400 µL cell homogenates and 1 mL exfoliated cell suspensions) were centrifuged (5000 *g*, 10 min, room temperature) and pellets were suspended in 200 µL PBS. From the next step, DNA extraction was performed according to the manufacturer's recommendation. MY/GP consensus nested PCR and restriction fragment length polymorphism (RFLP) analysis of MY and GP amplimers were used for detection and identification of HPV-specific sequences, as described earlier (Borbely et al., 2007; Kónya et al., 2000). Identity of HPV11 was confirmed with E7 ORF specific PCR (Evander and Wadell, 1991).

2.2. Complete genome amplification and sequencing

Sequences of primers used for complete genome amplification were published earlier (Gáll et al., 2011). PCRs were performed with Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, USA); amplimers were purified by QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) and were sequenced in duplicates from both directions using the dideoxy chain termination method (Macrogen Europe, Amsterdam, the Netherlands). The complete genome of the reference plasmid was also sequenced for comparison to the sequence deposited in the GenBank. Sequences were assembled and analysed against the HPV11 reference genome (GenBank accession number: M14119) (Dartmann et al., 1986) using CLC Main Workbench 7.9.1 (Qiagen, Aarhus, Denmark). Genomes were deposited in the GenBank under accession numbers MK313761, MK313762, MK313764-MK313768, MN788368 and MW404328 (Table 1).

2.3. Phylogenetic analysis

The 82 complete HPV11 genomes available in the GenBank (as per 21 January 2021) were collected and aligned together with our nine new (Supplementary Table 1) and six previously analysed (Gáll et al., 2011; Gáll et al., 2013) HPV11 sequences. Identical sequences (except for identical sequences of Patient 8 and Patient 14; GenBank accession numbers MK313765 and MK313767) were used only once. Dendrograms were reconstructed using the neighbour-joining method with bootstrapping 1000 times (CLC Main Workbench 7.9.1; Qiagen, Aarhus, Denmark) and were used to determine the (sub)lineage of the newly collected sequences. The HPV6A1 (formerly HPV6b) reference sequence (GenBank accession number: X00203) (Schwarz et al., 1983) was included as outgroup.

2.4. 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) (Dartmann et al., 1986), 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 (PDB ID: 1R6K, resolution 2.5 Å) was structurally aligned with the crystal structure of the dimerized transactivation domain of HPV16 (PDB ID: 1DTO, resolution 1.9 Å) in UCSF Chimera (http://www.cgl.ucsf.edu/chimera (web archive link,), Pettersen et al., 2004) and subsequently structurally optimized. For generating the homology model of the HPV11 DNA-binding domain (DBD), the SWISS-MODEL Server (https://swissmodel.expasy.org (web archive link,), Waterhouse et al., 2018) was used based on the crystal structure of the highly similar E2 DBD - DNA complex of HPV6 lineage B3 (PDB ID: 2AYG, resolution 3.1 Å). The DNA complementary to the HPV11 E2 DBD was designed with the DNA Sequence to Structure tool of SCF Bio (http://www.scfbio-iitd.res.in (web archive link,), Arnott et al., 1976). For docking the HPV11 DBD and its respective variants to the complementary DNA, the NPDock server (Tuszynska et al., 2015) was used, with the protein and DNA interface constrained to the conserved DNA-binding site described by Rogers et al. (2011) and McBride (2013). Lastly, the hinge region (amino acids 193–270) was *de novo* modelled using the I-Tasser server (zhanglab.ccmb.med.umich.edu, Yang and Zhang, 2015) 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 (Fig. 2A).

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 (Trott and Olson, 2010) and the PreDBA server (http://predba.denglab.org, Yang and Deng, 2020). For visualization, final models were depicted with UCSF Chimera X (Pettersen et al., 2021), and ligand interactions were designed with LigPlot+ (Laskowski and Swindells, 2011). Lastly, the protein model geometry was validated with SAVES Ramachandran plot (servicesn.mbi.ucla.edu/SAVES/Ramachandran/, Colovos and Yeates, 1993) and was uploaded to the Protein Model Database under PMDB ID PM0083528 (http://srv00.recas.ba.infn.it/PMDB/, Castrignano et al., 2006). Putative phosphorylation sites on the E2 protein variants were predicted using the NetPhos 3.1. server (Blom et al., 2004).

2.5. Site-directed mutagenesis

In this study, one yet unexamined nucleotide polymorphism in the LCR was identified in a moderately aggressive papillomatosis (Patient 9; JO-RRP9). The effect of T7331G polymorphism on LCR activity was examined by site-directed mutagenesis with a primer set specific to the polymorphism (Fwd 5'-GTAGTGTGTATATGTGTCTTGTATTGTGTATATG-3′ and Rev. 5'–CATATACACAATACAAGACACATATACACAC TAC–3′ at annealing temperature 61 °C; polymorphic nucleotide underlined in the forward primer) according to the protocol described earlier (Gáll et al., 2013). The construct was verified by sequencing.

2.6. Transformation, transient transfection and luciferase assay

The plasmid containing the reference HPV11 sequence (GenBank accession number: M14119) (Dartmann et al., 1986) was kindly provided by Dr. H. zur Hausen. LCR sequences of genomes from each patient as well as from the reference plasmid were cloned into luciferase reporter vector pALuc as described earlier (Gáll et al., 2013). For functional analysis of E2 variants, the reference and unique E2 ORFs were cloned into pCDNA3.1+ mammalian expression vector using HPV11_E2_*KpnI* (5′- GCCGGTACC-TTCCAAATCCATTCCCCT-3′, nucleotide position 2597–2614, KpnI restriction site is underlined) and HPV11_E2_*XbaI* (5′- GCCTCTAGA-GGCACTACCTCCATACAC -3′, nucleotide position 3867–3884, XbaI restriction site is underlined) primers. Amplification was performed with Phusion High Fidelity DNA Polymerase (Thermo Scientific, Waltham, USA). Initial denaturation (98 °C for 3 min) was followed by 40 cycles of PCR: denaturation at 98 °C for 30 s, annealing at 50 °C for 30 s and extension at 72 °C for 30 s. The final extension was 72 °C for 5 min.

The amplimers were purified with Qiaquick Gel Extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendation, digested by KpnI and XbaI restriction enzymes (Promega, Madison, WI, USA) and ligated into pre-cut pcDNA3.1+ expression vector using T4 DNA Ligase (Thermo Scientific, Waltham, USA). The

ligated constructs were transformed into *Escherichia coli* XL-1 by TransformAid Bacterial Transformation Kit (Thermo Scientific, Waltham, USA) and Zymo Pure Plasmid Mini Prep Kit (Zymo Research, Irvine, USA) was used for the plasmid minipreparations. Plasmid midipreparations were performed with PureLink HiPure Plasmid Midiprep Kit (Thermo Scientific, Waltham, USA). All kits were used according to the manufacturers' recommendations. All constructs were verified by sequencing (Macrogen, Amsterdam, the Netherlands).

To analyse the LCR activity of HPV11 sequences, HEp-2 (ATCC number CCL-23) laryngeal carcinoma cells were transfected with 2 μg of the reporter vector pALuc containing the LCR sequences and 1 μg of RSV-β-Gal plasmid as an internal control for transfection efficiency. The LCRs from Patients 1–6 of the former study (Gáll et al., 2013) were remeasured using the stored clones. 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 (10^6 cells/dish) with Lipofectamine 2000 (Invitrogen, Carlsbad, California, USA) according to the manufacturer's recommendation and were harvested 48 h posttransfection by the addition of 400 μL Reporter Lysis Buffer (Promega, Madison, WI, USA) and lysed by a freeze-thaw cycle. The luciferase activity of the cell extracts was measured by the Luciferase Assay System (Promega, Madison, WI, USA) according to the manufacturer's recommendation. Transfection was standardized using the β-galactosidase assay by adding 1 mg/mL ortho-nitrophenyl-β-galactoside chromogenic substrate and normalized to the protein concentration as measured by the Bradford method. All sequences were tested in duplicates in three independent experiments, means of

Table 2
Nucleotide changes in human papillomavirus type 11 complete genomes from different diseases associated with HPV11. The table contains only those polymorphisms which were detected in newly enrolled patients, grey shading highlights the polymorphisms newly found in this study. For polymorphisms in Patients 1–6, we refer to Gáll et al., 2013.

ORF	Nucleotide position	Reference (M14119)	Patient 7 (MK313764) JO-RRP7	Patient 8 (MK313765) JO-RRP8	Patient 9 (MK313766) JO-RRP9	Patient 10 (MN788368) JO-RRP10	Patient 11 (MK313761) AO-RRP1	Patient 12 (MK313762) AO-RRP2	Patient 13 (MW404328) AO-RRP3	Patient 14 (MK313767) CAC	Patient 15 (MK313768) CA	Amino acid change/comment
E6 (nt 102..554)	137	T	C	C	C	C	C	C	C	C	C	no
	323	T									C	no
	380	C	T	T	T	T	T	T	T	T		no
	662	G	T	T	T	T	T	T	T	T		A45S
E7 (nt 530..826)	1107	A		C		C	C	C	C	C		no
E1 (nt 832..2781)	1130	A			C							N100T
	1623	T			C							no
	1713	A						T				no
	1870	A				T						I347F
	2358	C		T		T	T	T	T	T		no
	2580	A	G	G	G	G	G	G	G	G		no
E2 (nt 2723..3826) E4 (nt 3255..3581)	2884	C	T	T	T	T	T	T	T	T		no
	2888	T	C	C	C	C	C	C	C	C		no
	3211	A	G									no
	3381						deletion					
	3391	A		G		G		G	G	G		no change in E2 Q46R in E4
	3436	G	A	A	A	A		A	A	A		no change in E2 G61E in E4
	3438											
	3456	C			T							S245F in E2 P68S in E4 (P50S in E1^E4 fusion protein)
	3462	A			C							N247T in E2 T70P in E4 (T52P in E1^E4 fusion protein)
	3487	C	T	T	T	T	T (not applicable due to the frameshift)	T	T	T		no change in E2 S78L in E4 (S60L in E1^E4 fusion protein)
	3626	A	C	C	C	C	C	C	C	C		no
	3645	A	G	G	G	G	G (not applicable due to the frameshift)	G	G	G		K308R in E2
	3727	A	C									no
Non coding sequence	3832	A	G	G	G	G	G	G	G	G		non coding
E5A (nt 3871..4146)	3952	A	T	T	T	T	T	T	T	T		I28F
	3985	G						A				D39N
	3991	G	C	C	C	C	C	C	C	C		V41L
	4142	A			C							no
Non coding sequence	4380	T			G							non coding
L2 (nt 4417..5784)	4521	A									C	E35D
	4569	T							G			no
	4602	T	A									no
	4647	C	T	T	T	T	T	T	T	T		no
	4887	C	A	A	A	A	A	A	A	A	A	no
	5437	G						T				E341*
L1 (nt 5771..7276)	6028	C	T	T	T	T	T	T	T	T		no
	6484	T		C			G			C		no
	6607	C	T									no
LCR (nt 7277..7933, 1..101)	7331	T			G							
	7413	A		C		C	C	C	C	C		
	7479	C	T	T	T	T	T	T	T	T		
	7509	T	deletion	deletion	deletion	deletion	deletion	deletion	deletion	deletion		
	7547	T	C	C	C	C	C	C	C	C		
	7584	-	C insertion	C insertion		C insertion	C insertion	C insertion	C insertion	C insertion		on a putative USF and Sp-1 binding site

ORF open reading frame; E early; L late; LCR long control region; nt nucleotide

ORF open reading frame; E early; L late; LCR long control region; nt nucleotide.

Fig. 1. Dendrogram of representative archive and presently reported HPV11 genomes constructed using the neighbour-joining method with 1000 bootstrapping rooted to the HPV6A1 (formerly HPV6b) reference sequence (GenBank accession number: X00203) used as an outgroup. Bootstrap percentages higher than 85% are shown on the branches. Our sequences are indicated with arrows. Identical sequences (except for identical sequences of Patient 8 and Patient 14; GenBank accession numbers MK313765 and MK313767) were used only once.

luciferase activities from the three experiments were compared using ANOVA with Tukey's post-tests with Bonferroni correction as appropriate (PaSt2.17c; Hammer et al., 2001). Relative luciferase activities were calculated as the ratio of the mean measured activity of a sequence variant divided by the luciferase activity of the reference.

3. Results

3.1. HPV11 complete genome sequences and genome comparison

In virus genomes from the newly enrolled patients (Patients 7–15) 19 new single nucleotide polymorphisms (SNPs) were found, of which 17 were unique (Table 2). 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 (Table 2). Seven resulted in amino acid alteration in ORFs E1, E2/E4, E5A and L2 (Table 2).

In addition, the virus genome from Patient 11 (AO-RRP1) exhibited a 58-bp deletion in the E2/E4 ORF leading to a frameshift and an early stop codon (Table 2). 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), the HPV11 genome was identical to the previously identified genome (GenBank accession number HE574702) (Gáll et al., 2013). In case of Patient 6 (JO-RRP6), 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 (GenBank accession number FR872717) (Gáll et al., 2011). Similarly, multiple follow-up samples were obtained in case of Patients 8 (JO-RRP8) and 9 (JO-RRP9) yielding identical HPV11 genomes. The genomes determined from the three localizations (supraglottis, subglottis and stoma) of Patient 7 (JO-RRP7) were also fully identical. The genomes from Patient 8 (JO-RRP8) and Patient 14 (CAC) 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 (Heinzel et al., 1995; Gáll et al., 2013).

The genome from Patient 15 (CA) clustered together with the reference genome (sublineage A1), all other genomes belonged to sublineage A2. The genome JO-RRP7 (Patient 7) clustered together with the genomes from more severe cases within sublineage A2 (including Patient 6, JO-RRP6) (Gáll et al., 2013) (Fig. 1).

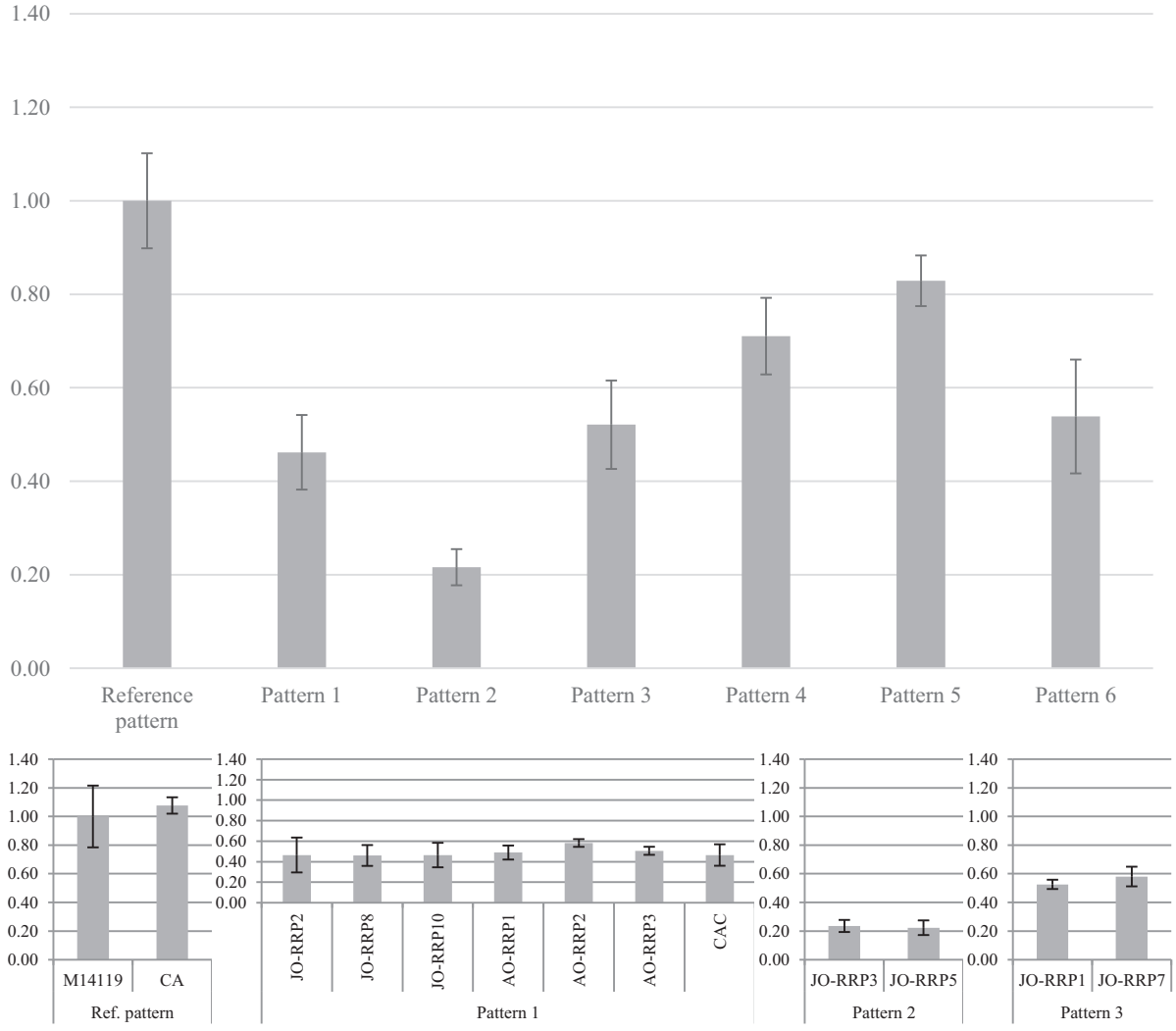

Fig. 2. Average transactivating activity of the different LCR sequence variants measured without E2. Smaller graphs in the bottom show the activities of the individual samples with the same LCR variant. Data presented represent the averages of three independent experiments, whiskers represent the standard deviation.

3.2. LCR polymorphisms, transactivating potential and site-directed mutagenesis

The difference from the reference genome deposited in the GenBank (GenBank accession number: M14119; Dartmann et al., 1986) 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 (Heinzel et al., 1995; Gáll et al., 2013). One yet unreported novel polymorphism (T7331G) was found, in the sequence from Patient 9 (JO-RRP9, Pattern 6). This polymorphism did not significantly alter the LCR activity either in the original sequence (Fig. 2) or in the mutant sequence generated by site-directed mutagenesis (data not shown).

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 (Pattern 1 in Table 3). LCR Pattern 2 is represented by two sequences from solitary papillomata (JO-RRP3 and 5) from our former study (Gáll et al., 2013). LCR Pattern 3 was found in one genome from JO-RRP of moderate severity (JO-RRP1) in our former study and one further genome from this study from a severe JO-RRP (JO-RRP7). LCR Patterns 4, 5 and 6 were unique from JO-RRPs (Patient 4, 6 and 9, respectively); the only difference between Pattern 6 and Pattern 3 was the newly found indifferent T7331G polymorphism. The sequence from CA was identical to the reference LCR (reference pattern, Table 3).

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, Fig. 2). All other LCRs excepting JO-RRP6 (Pattern 5) showed significantly lower luciferase activity ($p \leq 0.005$ in all comparisons (Fig. 2, Table 4). Lowest activities were measured in case of LCRs from patients with solitary papillomata (JO-RRP3 and JO-RRP5) representing Pattern 2 (Fig. 2); these activities were significantly lower than those produced by LCRs belonging to Pattern 1 and Patterns 3–6 ($p = 0.026$ to $p \leq 0.004$; Fig. 2, Table 4). Further statistical comparisons are shown in Table 4.

3.3. 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. This outlines five E2 variants at the amino acid level, i) the reference sequence, CA and JO-RRP6 (Reference variant); ii) the SNP K308R characterizes JO-RRP1, JO-RRP2, JO-RRP3, JO-RRP5, JO-RRP7, JO-RRP8, JO-RRP10, AO-RRP2, AO-RRP3, CAC (Variant 1); iii) AO-RRP1 with the deletion and the early stop codon (Variant 2); iv) SNPs Q86K and K308R are simultaneously present in JO-RRP4 (Variant 3); v) three SNPs, S245F, N247T and K308R are present in JO-RRP9 (Variant 4) (Table 5).

The polymorphism K308R is present in the majority of sequences (Variants 1, 3 and 4) 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 K308R-containing E2 variants are identical (17.24 kcal/mol), as measured by the PreDBA server (Yang and Deng, 2020). This is supported by the fact that the putative localization of the K308 residue does not face the DNA (Fig. 3B). 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 neighboring F311 and E312 residues of the protein chain stabilizing the DNA-binding helix (Fig. 3C and E). In case of the K308R polymorphism, the more bulky arginine side-chain faces the other protein chain in the dimer, bonding to its S322 (Fig. 3D and F). The resulting hydrogen bonds increase the total binding energy between the two E2 monomers markedly, from −17.23 kcal/mol in K308 to −22.46 kcal/mol in K308R polymorphism.

The polymorphisms S245F and N247T in E2 (Variant 4) unique to JO-RRP9 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 polymorphism does not change the electrostatic surface potential of the region, as shown in Fig. 3G and H. The polymorphism S245F disrupts a probable

Table 4
Comparison of the activity of different LCR patterns.

	Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 5	Pattern 6
Reference pattern	0.003	0.003	0.003	0.005	NS	0.003
Pattern 1		0.026	NS	0.023	0.003	NS
Pattern 2			0.004	0.003	0.003	0.003
Pattern 3				NS	0.003	NS
Pattern 4					NS	NS
Pattern 5						0.005

NS not significant.

Table 3
LCR patterns of samples.

	Ref (M14119)	Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 5	Pattern 6	Comment
7331	T						G	
7413	A	C						
7479	C	T		T		T	T	
7509	T	deletion	deletion	deletion		deletion	deletion	
7547	T	C	C	C	C	C	C	
7585	–	C insertion	C insertion	C insertion	C insertion	C insertion		on a putative USF and Sp-1 binding site
7719	G insertion	G insertion	G insertion	G insertion	G insertion	G insertion	G insertion	next to a putative Sp-1
7720	C insertion	C insertion	C insertion	C insertion	C insertion	C insertion	C insertion	binding site
7904	T					A		immediately next to an E2 binding site
Patients	CA (MK313768)	JO-RRP2 (HE574702) JO-RRP8 (MK313765) JO-RRP10 (MN788368) AO-RRP1 (MK313761) AO-RRP2 (MK313762) AO-RRP3 (MW404328) CAC (MK313767)	JO-RRP3 (HE574703) JO-RRP5 (HE574705)	JO-RRP1 (HE574701) JO-RRP7 (MK313764)	JO-RRP4 (HE574704)	JO-RRP6 (FR872717)	JO-RRP9 (MK313766)	

Table 5
Human papillomavirus 11 E2/E4 protein variants of the samples.

	Reference type		Variant 1		Variant 2	Variant 3	Variant 4	Amino acid change
	Reference sequence	Sequence type 1	Sequence type 2	Sequence type 3				
2884	C	T	T	T	T	T	T	no
2888	T	C	C	C	C	C	C	no
2978	C					A		Q86K in E2
3211	A			G				no
3381					58 bp deletion (resulted in frameshift and early stop codon)			
3391	A		G			G		Q46R in E4 (Q28R in E1^E4 fusion protein)
3436	G	A	A	A		A	A	G61E in E4 (G43E in E1^E4 fusion protein)
3438								
3456	C						T	S245F in E2 P68S in E4 (P50S in E1^E4 fusion protein)
3462	A						C	N247T in E2 T70P in E4 (T52P in E1^E4 fusion protein)
3487	C	T	T	T	T (not applicable due to the frameshift)	T	T	S78L in E4 (S60L in E1^E4 fusion protein)
3626	A	C	C	C	C	C	C	no
3645	A		G	G	G (not applicable due to the frameshift)	G	G	K308R in E2
3727	A	C		C				no
Patients	CA (MK313768)	JO-RRP6 (FR872717)	JO-RRP1 (HE574701) JO-RRP2 (HE574702) JO-RRP3 (HE574703) JO-RRP5 (HE574705) JO-RRP8 (MK313765) JO-RRP10 (MN788368) AO-RRP2 (MK313762) AO-RRP3 (MW404328) CAC (MK313767)	JO-RRP7 (MK313764)	AO-RRP1 (MK313761)	JO-RRP4 (HE574704)	JO-RRP9 (MK313766)	

phosphorylation site in a RXXS motif targeted by protein kinase A and B, as revealed by phosphorylation site prediction. (Scores of the reference sequence for being a general kinase target, for protein kinase A and B were 0.988, 0.863, 0.803, respectively). The polymorphism N247T may create a putative phosphorylation site, however, this site is predicted to be a poor target for phosphorylation (best score 0.446 for the glycogen synthase kinase 3; scores for being a general kinase target, for protein kinase A and B 0.251, 0.178, 0.051, respectively).

The unique polymorphism Q86K in E2 (Variant 3) in the TAD causes a major alteration in the surface charge of the E2. The reference E2 (Q86) possesses a negatively charged surface (Fig. 2I) and, while the surface of E2 with K86 is charged positively (Fig. 3J). Moreover, Q86 binds to K45 of the other E2 monomer chain, whereas K86 is exposed on the surface of the quaternary structure (Fig. 3K and L). 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 (−6.30 kcal/mol compared to −6.62 kcal/mol).

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**CL**KC**FR**YR**LN to **VSST**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 (data not shown). 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, JO-RRP2, JO-RRP3, JO-RRP5, JO-RRP7, JO-RRP8, JO-RRP10, AO-RRP2, AO-RRP3, CAC) 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, JO-RRP6 and AO-RRP1) 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.

The deletion in the E2/E4 ORF (AO-RRP1) 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.

In co-transfection experiments, the Reference Variant, Variant 3 and 4 of E2 increased luciferase activity of the reference LCR significantly ($p = 0.002$ in all comparisons, Table 6A); Variant 1 led to obvious but not statistically significant increase, while the truncated Variant 2 E2 did not alter LCR activity. Comparing variants, reference E2 led to the highest activity of the reference LCR ($p = 0.002$–0.003), followed by Variants 3 and 4, comparable to each other, but higher than Variants 1 and 2 ($p = 0.002$–0.003; Fig. 4A, Table 6A). Reference E2 increased the activity in case of all LCR patterns (Fig. 3B, Table 6B). Pattern 1 and 2 were characterized by only a small increase, while the reference LCR and patterns 3–6 were enhanced significantly ($p = 0.004$ in all comparisons)

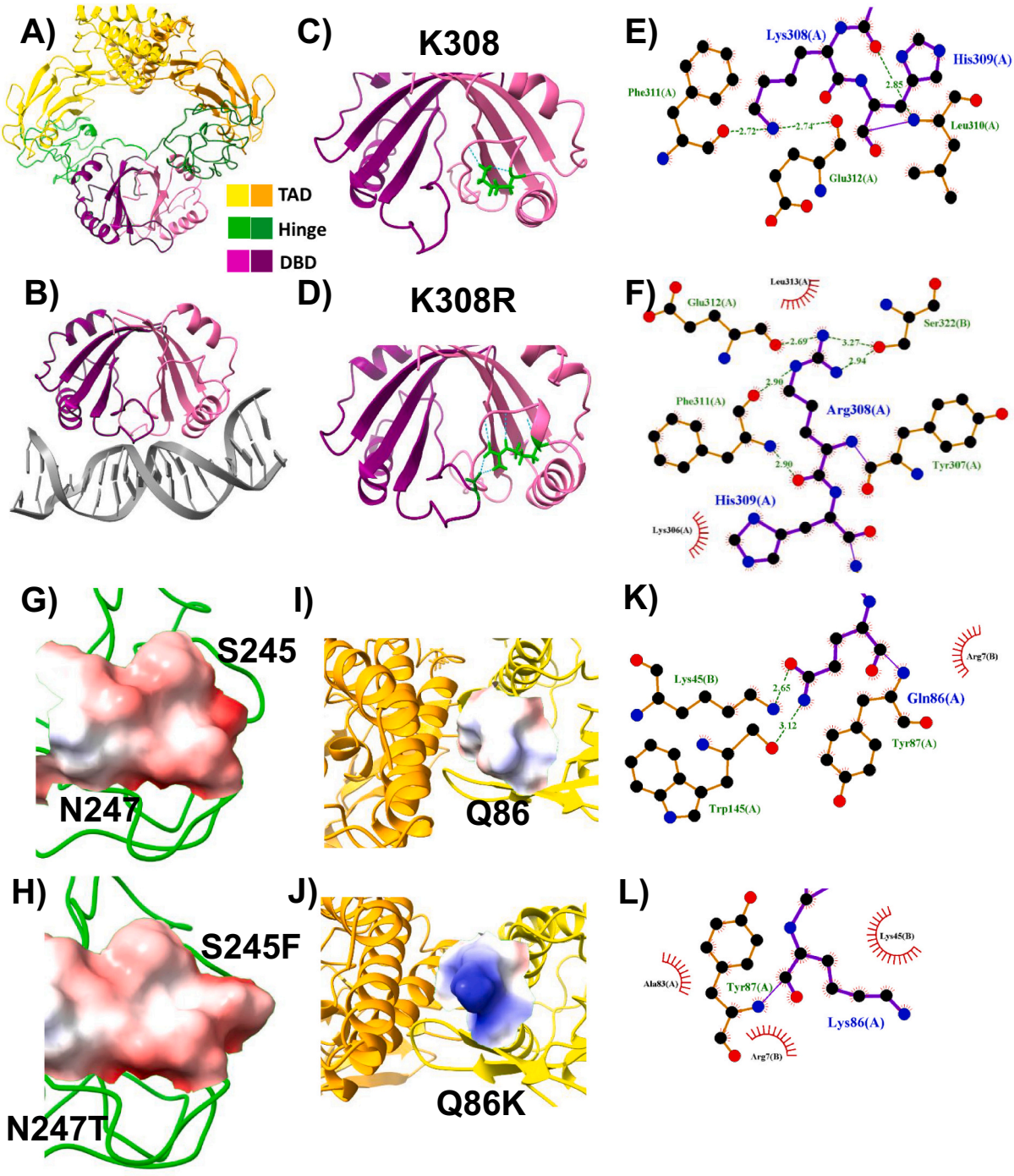

Fig. 3. Structural consequences of HPV 11 E2 polymorphisms. A) Homology model of the full-length human papillomavirus 11 E2 dimer containing the transactivation domain (TAD, yellow/orange), hinge (green/dark green) and the DNA binding domain (DBD, pink/purple). B) K308 (green) is not involved in DNA binding of the E2 DBD (DNA shown grey), as its side chain faces away from the major groove. C) and D) 3D ligand interactions of the wild-type K308 and K308R (green) with neighboring residues, with hydrogen bonds highlighted with blue dashed lines. In D), S322 is highlighted in dark green. E) and F) 2D ligand interaction plots of K308 and K308R, respectively, with chain label in brackets. Black, red and blue dots indicate carbon, oxygen and nitrogen atoms, respectively. Red radial stripes around atoms and residues mark exposure to the environment. Hydrogen bonds are indicated by green dashed lines with the distance (Å) of the H-bond. G) and H) HPV 11 E2 hinge region homology models (green) of wild-type and S245F/N247T polymorphism, respectively. Surface potential maps of residues 244–248 are shown, the range of the displayed surface potential map of residue 86 ranges from negative charge (red) to neutral (white) to positive charge (blue). I) and J) Homology models of the HPV 11 E2 transactivating domains (TAD, yellow/orange) with the reference sequence and Q86K polymorphism, respectively. K) and L) 2D ligand interaction plots of Q86 and Q86K, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6A
Comparison of the activity of different HPV11 E2 variants on the reference LCR.

	Reference pattern	Variant 1	Variant 2	Variant 3	Variant 4
Basic LCR activity	0.002	NS	NS	0.002	0.002
Reference variant		0.002	0.002	NS	0.003
Variant 1			NS	0.002	0.002
Variant 2				0.002	0.002
Variant 3					NS

NS not significant.

more by the reference E2 (Fig. 4B, Table 6B).

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 (Patterns 1 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 (Fig. 4C, Table 6C).

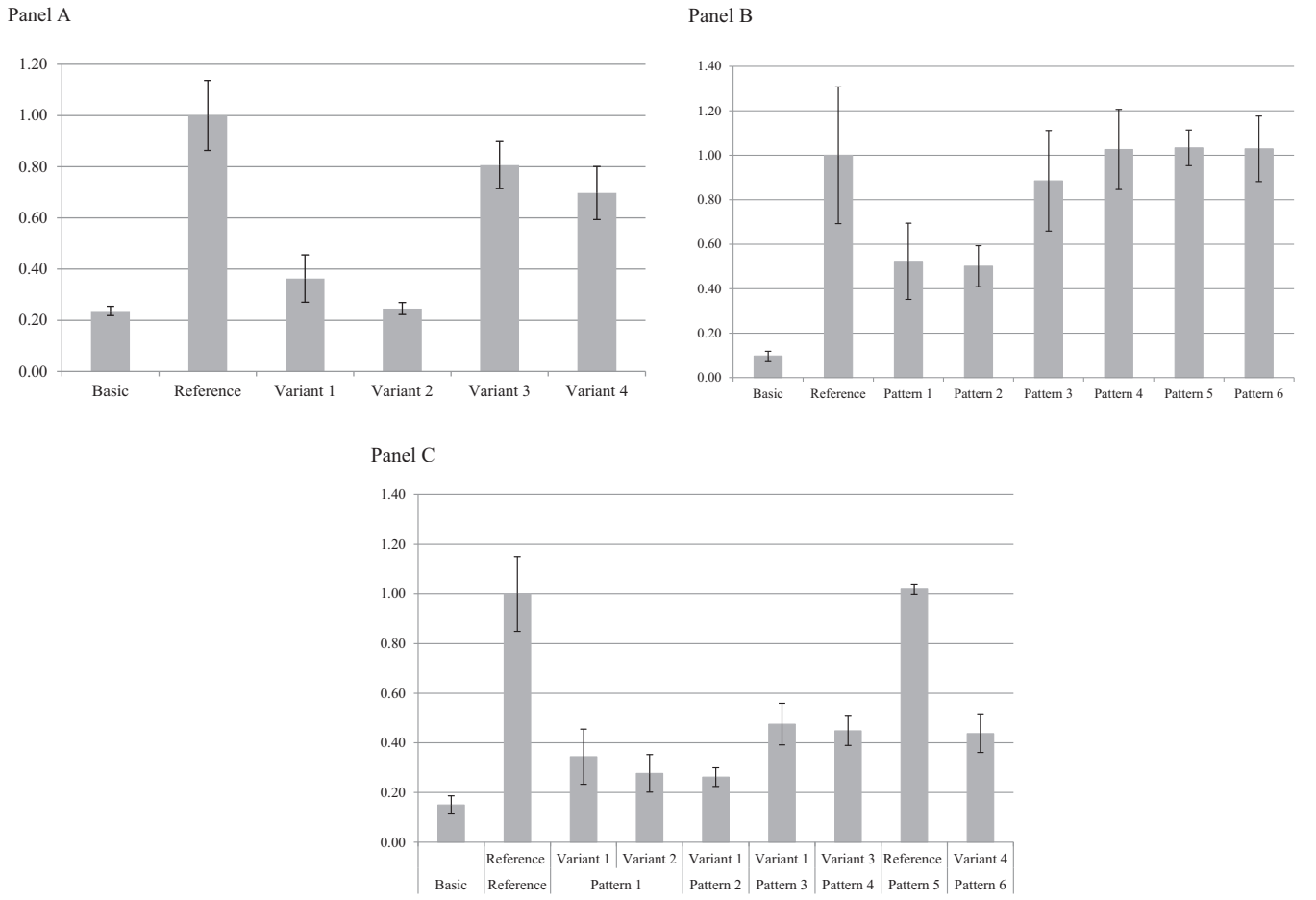

Fig. 4. A) Effect of the different E2 variants on the activity of the reference LCR. Basic activity is the luciferase activity of the reference LCR without any E2. Whiskers represent the standard deviation. B). Effect of the reference E2 on the activity of the different LCR variants. Basic activity is the luciferase activity of the reference LCR without any E2. Whiskers represent the standard deviation. C) Effect of the different E2 variants on the LCRs naturally occurring in the same genome. Basic activity is the luciferase activity of the reference LCR without any E2. Whiskers represent the standard deviation.

Table 6B
Comparison of the activity of reference HPV11E2 variant on the different LCR patterns.

	Reference	Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 5	Pattern 6
Basic LCR activity	0.004	NS	NS	0.004	0.004	0.004	0.004
Reference		NS	NS	NS	NS	NS	NS
Pattern 1			NS	NS	NS	NS	NS
Pattern 2				NS	0.097	0.084	0.092
Pattern 3					NS	NS	NS
Pattern 4						NS	NS
Pattern 5							NS

Table 6C
Comparison of the activity of the existing combinations of LCR patterns and E2 variants.

	Reference LCR –reference E2	LCR P 1 – E2 V1	LCR P1–E2 V2	LCR P2 – E2 V1	LCR P3 – E2 V1	LCR P4 – E2 V3	LCR P5 –reference E2	LCR P6 – E2 V4
Basic LCR activity	0.005	NS	NS	NS	0.006	0.011	0.005	0.017
Reference LCR –reference E2		0.005	0.005	0.005	0.005	0.005	NS	0.005
LCR P 1 – E2V1			NS	NS	NS	NS	0.005	NS
LCR P1 – E2V2				NS	NS	NS	0.005	NS
LCR P2 – E2V1					NS	NS	0.005	NS
LCR P3 – E2V1						NS	0.005	NS
LCR P4 – E2V3							0.005	NS
LCR P5 – reference E2								0.005
	M14119 CA (MK313768)	JO-RRP2 (HE574702 & MK313763) JO-RRP8 (MK313765) JO-RRP10 (MN788368) AO-RRP2 (MK313762) AO-RRP3 (MW404328) CAC (MK313767)	AO-RRP1 (MK313761)	JO-RRP3 (HE574703) JO-RRP5 (HE574705)	JO-RRP1 (HE574701) JO-RRP7 (MK313764)	JO-RRP4 (HE574704)	JO-RRP6 (FR872717)	JO-RRP9 (MK313766)

NS not significant.

4. Discussion

While disease severity and outcome varies from mild and transient disease to crippling or even fatal illness in papillomatoses caused by low-risk HPVs (Derkay and Bluher, 2019), the underlying virulence mechanisms explaining these differences are largely unknown. An earlier study reported a number of polymorphisms in the LCR with different effects on the transactivating potential with potential implications in disease severity (Gáll et al., 2013). The LCRs of the HPV 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 (Gáll et al., 2013) 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 Patient 11 (AO-RRP1) 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; (Gáll et al., 2013) 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. 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 (McBride and Warburton, 2017; Pinatti et al., 2018). Genome integration was concluded by studies of a few cases of HPV11-associated cancers, but disruption of E2 was not studied (Rady et al., 1998; Reidy et al., 2004).

The unique polymorphism Q86K in the E2 (Patient 4) may aid dimerization slightly and markedly alters the surface charge of the transactivating domain. As E2 is active as dimer (McBride, 2013), 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 (Sekhar and McBride, 2012). However, these polymorphisms did not modify the effect on the LCR; which is in line with the severity of the case (JO-RRP9) comparable to cases with a similar LCR (*e.g.* JO-RRP1). 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 (Chang et al., 2014; Sekhar and McBride, 2012). 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 to serve as a phosphorylation site. On the other hand, it is predicted to be targeted by a different kinase, 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 and CAC (GenBank accession numbers: MK313765 and MK313767, respectively) 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 et al. (2013) may be the wild type and the polymorphisms in the prototype sequence from a severe papillomatosis is in reality enhances LCR activity. This LCR sequence was found in a single novel genome from cervical atypia (CA; GenBank accession number: MK313768), 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 the reference 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) 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 (DiMaio and Petti, 2013).

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).

In addition to the ubiquitous attenuator LCR polymorphism (T7547C, Gáll et al., 2013), LCRs from JO-RRP3 and JO-RRP5 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; Pattern 5) when upregulated by the reference E2 still showed high activity and was associated with a severe disease with >30 recurrences (Gáll et al., 2013). 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.

5. Conclusion

The ubiquity of several polymorphisms in the studied genomes suggests that the reference sequence is a unique variant. Some polymorphisms in the reference as well as in the reported genomes were associated with increased severity of the disease caused, but the link is hard to prove unequivocally due to the low number of genomes with the same polymorphisms from diseases of known clinical course. Nevertheless, the transactivating potential provided by the LCR-E2 interplay seem to correspond to disease severity in a number of cases, suggesting that it is an important factor in the pathogenicity of HPV11.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgements

The help of Dr. Tamás Sápy in collecting the genital samples is gratefully acknowledged. Gábor Kardos and Krisztina Szarka were supported by János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Funding sources had no role in study design, analysis, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.104948.

References

Arnott, S., Campbell-Smith, P.J., Chandrasekaran, R., 1976. Handbook of biochemistry and molecular biology. In: Fasman, G.P. (Ed.), Nucleic Acids – Volume II, 3rd ed. CRC Press, Cleveland, pp. 411–422.

Bergvall, M., Melendy, T., Archambault, J., 2013. The E1 proteins. Virology 445, 35–56. https://doi.org/10.1016/j.virol.2013.07.020.

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., Brunak, S., 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 4, 1633–1649. https://doi.org/10.1002/pmic.200300771.

Bonagura, V.R., Hatam, L., Devoti, J., Zeng, F., Steinberg, B.M., 1999. Recurrent respiratory papillomatosis: altered CD8(+) T-cell subsets and T(H)1/T(H)2 cytokine imbalance. Clin. Immunol. 93, 302–311. https://doi.org/10.1006/clim.1999.4784.

Borbely, A.A., Murvai, M., Szarka, K., Konya, J., Gergely, L., Hernadi, Z., Veress, G., 2007. Survivin promoter polymorphism and cervical carcinogenesis. J. Clin. Pathol. 60, 303–306. https://doi.org/10.1136/jcp.2006.037804.

Buck, C.B., Day, P.M., Trus, B.L., 2013. The papillomavirus major capsid protein L1. Virology 445, 169–174. https://doi.org/10.1016/j.virol.2013.05.038.

Campisi, P., Hawkes, M., Simpson, K., Canadian Juvenile Onset Recurrent Respiratory Papillomatosis Working, G, 2010. The epidemiology of juvenile onset recurrent respiratory papillomatosis derived from a population level national database. Laryngoscope 120, 1233–1245. https://doi.org/10.1002/lary.20901.

Castrignano, T., De Meo, P.D., Cozzetto, D., Talamo, I.G., Tramontano, A., 2006. The PMDB protein model database. Nucleic Acids Res. 34, D306–D309. https://doi.org/10.1093/nar/gkj105.

Chang, S.W., Liu, W.C., Liao, K.Y., Tsao, Y.P., Hsu, P.H., Chen, S.L., 2014. Phosphorylation of HPV-16 E2 at serine 243 enables binding to brd4 and mitotic chromosomes. PLoS One 9, E110882. https://doi.org/10.1371/journal.pone.0110882.

Colovos, C., Yeates, T.O., 1993. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2, 1511–1519. https://doi.org/10.1002/pro.5560020916.

Dartmann, K., Schwarz, E., Gissmann, L., Zur Hausen, H., 1986. The nucleotide sequence and genome organization of human papilloma virus type 11. Virology 151, 124–130. https://doi.org/10.1016/0042-6822(86)90110-8.

Derkay, C.S., Bluher, A.E., 2019. Update on recurrent respiratory papillomatosis. Otolaryngol. Clin. N. Am. 52, 669–679. https://doi.org/10.1016/j.otc.2019.03.011.

DeVoti, J.A., Steinberg, B.M., Rosenthal, D.W., Hatam, L., Vambutas, A., Abramson, A.L., Shikowitz, M.J., Bonagura, V.R., 2004. Failure of gamma interferon but not interleukin-10 expression in response to human papillomavirus type 11 E6 protein in respiratory papillomatosis. Clin. Diagn. Lab. Immunol. 11, 538–547. https://doi.org/10.1128/CDLI.11.3.538-547.2004.

DeVoti, J.A., Rosenthal, D.W., Wu, R., Abramson, A.L., Steinberg, B.M., Bonagura, V.R., 2008. Immune dysregulation and tumor-associated gene changes in recurrent respiratory papillomatosis: a paired microarray analysis. Mol. Med. 14, 608–617. https://doi.org/10.2119/2008-00060.DeVoti.

DiMaio, D., Petti, L.M., 2013. The E5 proteins. Virology 445, 99–114. https://doi.org/10.1016/j.virol.2013.05.006.

Doorbar, J., 2013. The E4 protein; structure, function and patterns of expression. Virology 445, 80–98. https://doi.org/10.1016/j.virol.2013.07.

Evander, M., Wadell, G., 1991. A general primer pair for amplification and detection of genital human papillomavirus types. J. Virol. Methods 31, 239–250. https://doi.org/10.1016/0166-0934(91)90162-s.

Gáll, T., Kis, A., Feher, E., Gergely, L., Szarka, K., 2011. Virological failure of intralesional cidofovir therapy in recurrent respiratory papillomatosis is not associated with genetic or epigenetic changes of hpv11: complete genome comparison of sequential isolates. Antivir. Res. 92, 356–358. https://doi.org/10.1016/j.antiviral.2011.09.007.

Gáll, T., Kis, A., Tatar, T.Z., Kardos, G., Gergely, L., Szarka, K., 2013. Genomic differences in the background of different severity in juvenile-onset respiratory papillomatoses associated with human papillomavirus type 11. Med. Microbiol. Immunol. 202, 353–363. https://doi.org/10.1007/s00430-013-0297-y.

Hammer, Ø., Harper, D.A., Ryan, P.D., 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontol. Electron. 4, 9. http://palaeo-electronica.org/2001_1/past/issue1_01.htm.

Heinzel, P.A., Chan, S.Y., Ho, L., O'Connor, M., Balaram, P., Campo, M.S., Fujinaga, K., Kiviat, N., Kuypers, J., Pfister, H., 1995. Variation of human papillomavirus type 6 (hpv-6) and hpv-11 genomes sampled throughout the world. J. Clin. Microbiol. 33, 1746–1754.

Hermann, J.S., Pontes, P., Weckx, L.L., Fujita, R., Avelino, M., Pignatari, S.S., 2012. Laryngeal sequelae of recurrent respiratory papillomatosis surgery in children. Rev. Assoc. Med. Bras. 58, 204–208. https://doi.org/10.1590/S0104-42302012000200016.

Johnston, D., Hall, H., DiLorenzo, T.P., Steinberg, B.M., 1999. Elevation of the epidermal growth factor receptor and dependent signaling in human papillomavirus-infected laryngeal papillomas. Cancer Res. 59, 968–974.

Kónya, J., Veress, G., Juhász, A., Szarka, K., Sápy, T., Hernádi, Z., Gergely, L., 2000. Additional human papillomavirus types detected by the hybrid capture tube test

among samples from women with cytological and colposcopical atypia. J. Clin. Microbiol. 38, 408–411. https://doi.org/10.1128/JCM.38.1.408-411.2000.
Larson, D.A., Derkay, C.S., 2010. Epidemiology of recurrent respiratory papillomatosis. APMIS 118, 450–454. https://doi.org/10.1111/j.1600-0463.2010.02619.x.
Laskowski, R.A., Swindells, M.B., 2011. Ligplot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. 51, 2778–2786. https://doi.org/10.1021/ci200227u.
McBride, A.A., 2013. The papillomavirus E2 proteins. Virology 445, 57–79. https://doi.org/10.1016/j.virol.2013.06.006.
McBride, A.A., Warburton, A., 2017. The role of integration in oncogenic progression of HPV-associated cancers. PLoS Pathog. 13, e1006211 https://doi.org/10.1371/journal.ppat.1006211.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera – A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J. H., Ferrin, T.E., 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. https://doi.org/10.1002/pro.3943.
Pinatti, L.M., Walline, H.M., Carey, T.E., 2018. Human papillomavirus genome integration and head and neck cancer. J. Dent. Res. 97, 691–700. https://doi.org/10.1177/0022034517744213.
Rady, P.L., Schnadig, V.J., Weiss, R.L., Hughes, T.K., Tyring, S.K., 1998. Malignant transformation of recurrent respiratory papillomatosis associated with integrated human papillomavirus type 11 DNA and mutation of p53. Laryngoscope 108, 735–740. https://doi.org/10.1097/00005537-199805000-00021.
Reidy, P.M., Dedo, H.H., Rabah, R., Field, J.B., Mathog, R.H., Gregoire, L., Lancaster, W. D., 2004. Integration of human papillomavirus type 11 in recurrent respiratory papilloma-associated cancer. Laryngoscope 114, 1906–1909. https://doi.org/10.1097/01.mlg.0000147918.81733.49.
Rogers, A., Waltke, M., Angeletti, P.C., 2011. Evolutionary variation of papillomavirus e2 protein and e2 binding sites. Virol. J. 8, 379. https://doi.org/10.1186/1743-422X-8-379.
Rosenthal, D.W., DeVoti, J.A., Steinberg, B.M., Abramson, A.L., Bonagura, V.R., 2012. TH 2-like chemokine patterns correlate with disease severity in patients with recurrent respiratory papillomatosis. Mol. Med. 18, 1338–1345. https://doi.org/10.2119/molmed.2012.00284.
Schwarz, E., Durst, M., Demankowski, C., Lattermann, O., Zech, R., Wolfsperger, E., Suhai, S., Zur Hausen, H., 1983. DNA sequence and genome organization of genital human papillomavirus type 6b. EMBO J. 2, 2341–2348. https://doi.org/10.1002/j.1460-2075.1983.tb01744.x.
Sekhar, V., McBride, A.A., 2012. Phosphorylation regulates binding of the human papillomavirus type 8 E2 protein to host chromosomes. J. Virol. 86, 10047–10058. https://doi.org/10.1128/JVI.01140-12.
Szinai, M., Nagy, Zs, Máté, P., Kovács, D., Laczkó, L., Kardos, G., Sápy, T., Szűcs, A., Szarka, K., 2019. Comparative analysis of human papillomavirus type 6 complete genomes originated from head and neck and anogenital disorders. Infect. Genet. Evol. 71, 140–150. https://doi.org/10.1016/j.meegid.2019.03.019.
Tjon Pian Gi, R.E., San Giorgi, M.R., Slagter-Menkema, L., Van Hemel, B.M., Van Der Laan, B.F., Van Den Heuvel, E.R., Dikkers, F.G., Schuuring, E.M., 2015. Clinical course of recurrent respiratory papillomatosis: comparison between aggressiveness of human papillomavirus 6 and human papillomavirus 11. Head Neck 37, 1625–1632. https://doi.org/10.1002/hed.23808.
Trott, O., Olson, A.J., 2010. Autodock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. https://doi.org/10.1002/jcc.21334.
Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., Bujnicki, J.M., 2015. NPDock: a web server for protein-nucleic acid docking. Nucleic Acids Res. 43, W425–W430. https://doi.org/10.1093/nar/gkv493.
Vambutas, A., DiLorenzo, T.P., Steinberg, B.M., 1993. Laryngeal papilloma cells have high levels of epidermal growth factor receptor and respond to epidermal growth factor by a decrease in epithelial differentiation. Cancer Res. 53, 910–914.
Wang, J.W., Roden, R.B., 2013. L2, the minor capsid protein of papillomavirus. Virology 445, 175–186. https://doi.org/10.1016/j.virol.2013.04.017.
Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. https://doi.org/10.1093/nar/gky427.
Yang, W., Deng, L., 2020. PreDBA: a heterogeneous ensemble approach for predicting protein-dna binding affinity. Sci. Rep. 10, 1278. https://doi.org/10.1038/s41598-020-57778-1.
Yang, J., Zhang, Y., 2015. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 43, W174–W181. https://doi.org/10.1093/nar/gkv342.