SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D.)

# Epidemiology and phylogeny of torque teno virus in Eastern Hungary

by Enikő Fehér

Supervisor:

dr. Krisztina Szarka


UNIVERSITY OF DEBRECEN

DOCTORAL SCHOOL OF PHARMACEUTICAL SCIENCES

Debrecen, 2012

**EPIDEMIOLOGY AND PHYLOGENY OF TORQUE TENO VIRUS IN EASTERN HUNGARY**

By **Enikő Fehér**
biologist M.Sc. (microbiologist)

Supervisor: dr. Krisztina Szarka, Ph.D.
Doctoral School of Pharmaceutical Sciences, University of Debrecen

| | |
|---|---|
| Head of the **Examination Committee**: | dr. Árpád Tósaki, D.Sc. |
| Members of the Examination Committee: | dr. Sándor Bíró, D.Sc. |
| | dr. János Minárovits, D.Sc. |

The Examination takes place at Department of Pharmacology, Medical and Health Science Center, University of Debrecen, 11:00, 18. 05. 2012.

| | |
|---|---|
| Head of the **Defense Committee**: | dr. Árpád Tósaki, D.Sc. |
| Reviewers: | dr. Attila Bácsi, Ph.D. |
| | dr. Krisztián Bányai, Ph.D. |
| Members of the Defense Committee: | dr. Sándor Bíró, D.Sc. |
| | dr. János Minárovits, D.Sc. |

The Ph.D. Defense takes place at the Lecture Hall of the 1st Department of Medicine, Institute for Internal Medicine, Medical and Health Science Center, University of Debrecen, 13:00, 18. 05. 2012.

# INTRODUCTION

Torque teno viruses (TTVs) are single-stranded DNA viruses belonging to the *Anelloviridae* family. The first TTV was described by Japanese researchers in 1997 from the serum of a patient with posttransfusion hepatitis. Since that the virus was demonstrated from numerous tissues, but its pathogenic role has been remained to be elucidated yet.

An earlier study from our working group showed high frequency of infection with TTV, genogroup 1 TTV (ggr1TTV), and ggr1TTV-HPV coinfection in association with poorer survival in laryngeal carcinoma patients with progressing disease compared to patients without tumour progression

TTVs have been classified into genogroups and genotypes based on ORF1 sequence differences; recently TTV taxonomy was reviewed and 29 species were proposed and accepted. Genome of the virus shows considerable variability, putatively leading to differences in gene expression patterns in different species and even intraspecies clades (formerly genotypes), which may lead to substantial differences in cell or tissue preference as well as in pathogenicity. Consequently, for assessment of pathogenic potential, determination of the genogroup and genotype may be important.

As most literature data is based on the former genogroup/genotype/subtype classification, to allow for comparison of the present to former data, comparative classification both to species and to genotypes of presently determined and archive Hungarian sequences was performed.

Incited by these results, present study aims at investigation of frequency and importance of TTVs, ggr1TTV, HPV and ggr1TTV-HPV coinfection in further potentially HPV-associated oral (oral lichen planus, oral leukoplakia and oral squamous cell cancer) and cervical (cervical atypia and cervical cancer) malignant and potentially malignant lesions compared to corresponding oral and cervical control samples.

## REVIEW OF THE LITERATURE

### *Structure of TTVs*

TTVs are icosahedral, non-enveloped viruses with a diameter of 30-33 nm, possessing a cca. 3800 bases long circular single-stranded DNA genome with a negative polarity. TTVs are presently classified into the Alphatorquevirus genus of the *Anelloviridae* family, within which 29 distinct species has been described. Species and genera are defined as at least 35% and 56% difference at the nucleotide level, respectively.

Though the classificiation of the TTVs is based on species according to the presently accepted guidelines of the International Comittee on Taxonomy of Viruses (ICTV), the literature still refers to genogroups, genotypes and subtypes. The majority of the sequences available in the GenBank are partial sequences of the ORF1, the region N22 or an overlapping fragment. This N22 region is the name for the approximately 500 bases long fragment which was first identified as a TTV sequence. Though this fragment is only a small part of the complete virus genome, this region was shown to be suitable for phylogenetic investigations, at least within genogroup 1.

The former taxonomic classification listed five genogroups with more than 30 genotypes; within certain genotypes even subtypes were distinguished. Genogroups, genotypes and subtypes were defined as at least 50%, 30% and 15% difference, respectively, at the nucleotide level in region N22.

The genome of the prototype virus TA278 consists of 3852 nucleotides, includes an approximately 1.2 kilobase long untranslated region (UTR) and the approximately 2.6 kilobase long coding region. The length of the coding region may vary according to virus genotype. UTR plays a role in regulation of viral transcription and translation, includes promoter and enhancer elements, transcription factor binding sites as well as a GC-rich region with a length variable among genotypes and a conservative region. The latter two regions may form a loop, as inferred from their sequences.

The coding region of the TTV genome includes several ORFs; the number of ORFs varies among genotypes. All genotypes possess the two main ORFs, ORF1 and ORF2. ORF1 codes for a 770 amino acid long peptide, which include the capsid protein. In its N-terminal region it includes an arginine-rich region, similarly to the chicken anemia virus, as well as fragments with similarity to the Rep protein of the chicken anemia virus. The arginine-rich region aids the encapsidation of the virus due to its DNA-binding activity, while Rep proteins play an auxiliary role in replication. ORF1 sequences are characterized by three hypervariable

regions outside the N22 fragment, where sequence differences may be as high as 90%; these sequence variabilities may be related to immune evasion. The other ORF present in all TTVs, ORF2 codes for approximately 202 amino acids; its role remains unknown. Besides these two main ORFs, existence of further smaller ORFs is postulated.

Different TTV genotypes show differences in protein expression, which is attributed to nucleotide-level differences or to recombination. These may play a role in the pathogenesis of the infection or in immune evasion.

***Detection of TTVs***

The most frequently used diagnostic method is the polymerase chain reaction (PCR). Most research groups rely upon the primers targeting the N22 region (ORF1 primers). However, these primers are suitable for detection of the genotypes belonging ggr1. Primers targeting the well-conserved UTR are used to detect all TTV genotypes practically; these are suitable for survey general TTV prevalence. For detection of the viruses within the infected cells *in situ* hybridization techniques are utilized. Genotyping and determination of subtypes is performed by sequencing or by restriction fragment length polymorphism analysis (RFLP).

***Tissue distribution of TTVs***

The frequency of the virus in the general healthy population is very high; the infection is thought to be life-long. The virus and its replicative forms were demonstrated in the liver, due to liver involvement TTVs may appear in the bile and in the faeces, which allows for faecal-oral transmission. TT viruses and their replicative forms may also appear in the bone marrow, in periferal blood mononuclear cells (PBMC) and in haemopoietic stem cells; these serve as a source for blood-derived viruses. In addition, TTVs were demonstrated in the lung, spleen, pancreas, kidneys, thyroid gland, muscles, in samples from the uterine cervix, in the saliva, urine, semen and breast milk, umbilical cord blood and amnionic fluid.

***Occurrence of TTVs in different diseases***

Several authors postulated the role of TTV in idiopathic hepatitis and, as a coinfecting agent, in C hepatitis. Association between the outcome of hepatocellular carcinoma and presence as well as higher TTV load was reported, and TTVs were suggested as a prognostic factor in hepatocellular carcinoma. As TTV replicates in the bone marrow, its role in aplastic anaemia associated with hepatitis was proposed.

Among patients with idiopathic pulmonary fibrosis poorer survival was observed in the TTV-infected patient group as compared to noninfected patients. An association between presence and viral load of TTVs and lung cancer was also reported. TTVs were observed in bronchopneumonia and asthma as well; in children with bronchopneumonia dominance of genogroup 4 TTV was found. TTVs were presumed to alter the lymphocyte equilibrium leading to immunosuppressive effect, and to inhibit the function of ciliated epithelium of the airways through infecting the ciliated cells.

Hundred- to thousandfold higher virus numbers were reported from the saliva than from the serum, leading to postulating their replication in the oral cavity. Presence of TTVs in oral epithelial cells was proved in oral lichen planus and oral leukoplakia using *in situ* hybridization. Association was proposed between TTVs and periodontitis, the role of several different factors was proposed. TTVs may play a role in proinflammatory processes or the high TTV load may be the result of virus replication in infiltrating lymphocytes.

General TTV prevalence was examined in patients with lesions in the uterine cervix and in control individuals without cervical lesions using UTR PCR. TTV prevalence was significantly higher in patients than in controls; moreover, TTV was found more frequently in HPV positive samples than in HPV negatives. Exfoliated cells showed a 10-100-fold higher virus copy numbers than that found in the serum, probably indicating virus replication in cervical epithelium. In these samples the most frequently found genogroup was ggr1.

Presence of TTV was also investigated in patients with gastritis or with different autoimmune or malignant diseases.

***TTV studies in Hungary***

Previously two studies investigated the prevalence of TTV genotypes in Hungary using the ORF1 PCR. The prevalence rates in hepatitis patients were 50.4% (115/228) and 50.9% (56/110), while 18.5% (20/108) of healthy individuals proved to be TTV carriers. The majority of sequenced viruses belonged to genotype 2.

Our working group published two studies with TTVs. In the first PBMCs of kidney transplant recipients and healthy blood donors were examined. Based on the UTR PCR 100% (92/92) of transplant recipients and 95.5% (63/66) of healthy individuals, while using the ORF1 PCR 57.6% (53/92) and 19.7% (13/66) of transplant recipients and controls, respectively, were positive for TTVs. Six of the twelve sequenced viruses belonged to genotype 2.

The other study examined the tissue samples of 40 patients with respiratory papillomatosis or laryngeal cancer. UTR PCR showed 80.0%, 100% and 88.0% prevalence in respiratory papillomatosis, in respiratory papillomatosis with malignant transformation and in laryngeal cancer, respectively, while the prevalence rates of genogroup 1 TTVs were 20.0%, 100% and 44.0% as revealed by the N22 (ORF1) PCR. Coinfection with HPVs was significantly more frequent in laryngeal cancer patients with tumour progression (local recurrence or metastasis) than in patients without progression. In parallel, the survival of patients with tumour progression was significantly lower.

***Human papillomaviruses***

Human papillomaviruses belong to the Alpha-, Beta-, Gamma-, Mu- and Nupapillomavirus genera of the *Papillomaviridae* family, having icosahedral capsid symmetry of a diameter of approximately 55 nm. They are non-enveloped. The virus genome is an approximately 8 kilobase pair long double-stranded circular DNA molecule, containing early (E) and late (L) ORFs and a long control region (LCR).

Though ICTV classified HPVs into species, researchers rather use the terminology of HPV genotypes. At least 120 HPV genotypes are known, which are classified based on their oncogenic potential to groups with low and high oncogenic risk.

Transmission of HPVs may occur by direct skin-to-skin contact, by sexual contact, by microabrasions of the skin or mucosal surfaces or during delivery. The replication cycle of the virus is closely associated with the differentiation program of epithelial cells. Viruses reaching the basal cell layer of the stratified epithelium will cause the transcription of early genes and the virus replication starts, then in more differentiated layers structural proteins encoded by late genes are produced and the mature virions are released with the shed outer epithelial layer.

Proteins coded for by early ORFs can be detected in the early stage of the virus replication. These proteins are involved in virus replication, transcription (E1, E2; E4; E5), in maturation and release of virions (E4) and in transformation of host cells (E5, E6, E7). E5 is a transmembrane protein activating the host cell growth factor receptors. E6 binds the p53 tumour suppressor protein and enhances its degradation, thus inhibiting DNA repair and, in case of irreversible DNA damage, apoptosis. E7 protein binds to the pRb (retinoblastoma) protein, and allows for transactivation of different factors involved in cell proliferation by E2F. L1 ORF encodes for the main capsid protein, while L2 ORF encodes the smaller one,

which aids in incorporation of the viral nucleic acid into the capsid. LCR is the region responsible for regulation of virus replication and transcription.

The genome of the high risk genotypes may be integrated into the host cell genome; the integration site is in the region of the E1/E2 ORF. This integration leads to increased expression of the transforming E6 and E7 proteins. As these oncoproteins are crucial for restarting the cell cycle, their uncontrolled expression may lead to malignant transformation.

Cancer of the uterine cervix is one of the most frequent malignant tumour of women worldwide, in which HPVs with high oncogenic risk is considered as direct etiological factors, being present in at least 90% of cervical cancers. In cca. 70% of HPV positive cervical atypia and cervical cancer cases high risk genotypes, HPV16 and HPV18, are detected. Cervical premalignant lesions show 50-90% prevalence, while 30% of healthy women carry HPVs. Low risk genotypes are rarely found in cervical malignancies, these mostly cause anogenital warts and low grade intraepithelial lesions.

Besides the well-known oncogenic factors (alcohol consumption, smoking, hygienic, alimentary and lifestyle-related factors and, in case of oral lichen planus, autoimmunity), the role of HPVs has been suggested in squamous cell cancers of the head and neck region as well as the lesions considered potentially malignant (oral lichen planus and oral leukoplakia). HPVs were demonstrated in 20-100% of cancers and in 20-30% of potentially malignant lesions, while only in 0-10% of apparently healthy individuals depending on the target region, sample type and detection method. In 85-95% of HPV positive cases HPV16 is detected, low risk HPVs (HPV6 and HPV11) are detected with lower frequency. Low risk HPV genotypes, mainly HPV6 and HPV11 are detected with high frequency in benign oral or laryngeal papillomatoses and in laryngeal cancer cases. Neonatal infections may lead to juvenile recurrent respiratory, oral or genital papillomatoses.

***Aims***

Though the role of TTVs has been postulated in several diseases, unequivocal proof is uniformly lacking. It was suggested that, similarly to the case of HPVs, certain TTV genogroups, genotypes or subtypes may be associated with a certain tissue, anatomical region or disease. Our working group has reached such an opinion in case of ggr1TTV when examining laryngeal cancer.

Based on the abovementioned data, the following aims were pursued.

- To establish the prevalences of TTVs and ggr1TTVs in further putatively HPV-associated malignant tumour besides laryngeal cancer as well as in the associated

premalignant conditions of the anatomical regions as compared to representative control populations.

- To determine and compare the frequency of ggr1TTV-HPV coinfection in the examined samples.
- To assess the impact of the viruses and the coinfections on the survival of cancer patients.
- To determine the genotype and subtype of the detected ggr1TTVs and to compare the genotype/subtype distribution in the different study populations.
- To compare the geographical distribution of genotypes/subtypes using literature data.

## PATIENTS AND METHODS

### *Patient groups*

The study examined 904 samples of 608 individuals. Oral samples were collected at the Maxillofacial and Oral Surgery and Periodontology Departments of the Faculty of Dentistry of the University of Debrecen, gynecological samples originated from the Department of Obstetrics and Gynecology of the Faculty of Medicine of the University of Debrecen.

The oral study group consisted of 65 patients with oral squamous cell cancer (OSCC) (51 males, 14 females, mean age 54.4 years, 25-80 years of age), 44 patients with oral leukoplakia (OL) (14 males, 30 females, mean age 56.3 years, 29-91 years of age) and 119 patients with oral lichen planus (OLP) (31 males, 88 females, mean age 55.0 years, 23-79 years of age). Oral control groups included 72 healthy individuals (19 males, 53 females, mean age 52.0 years, 22-77 years of age) providing exfoliated oral epithelial cell samples. In case of OSCC samples three samples per patient were collected and processed, one tissue sample form the surgically excised tumour and two exfoliated cell sample from the apparently healthy buccal and lingual mucosa. In case of OL and OLP patients exfoliated cell samples were collected from the surface of the lesion and from the apparently healthy mucosa.

The uterine cervical study groups were comprised of 87 patients with cervical cancer (CC) (mean age 42.8 years, 27-62 years of age) providing surgically excised tumour tissue sample, 84 patients with cervical atypia (atypical cervical mucosa; ACM) (mean age 36.6 years, 20-78 years of age) and 97 healthy women (normal cervical mucosa; NCM) (mean age

38.6 years, 18-62 years of age). From the latter two groups exfoliated cervical mucosal cells were collected.

The laryngeal study group included 25 patients with laryngeal squamous cell cancer (LSCC) (25 males, 0 females, mean age 56.6 years, 43-71 years of age), ten patients with recurrent respiratory papillomatosis (RRP) (8 males, 2 females, mean age 27.7 years, 3-71 years of age,) and five patients with laryngeal papilloma with malignant transformation (MP) (4 males, 1 female, mean age 55.0 years, 43-66 years of age) providing surgically excised samples from the lesion. The laryngeal study group was examined in a previous study; during the present study the clinical follow-up data were actualized and the sequence analysis of detected ggr1TTVs was carried out. The LSCC and MP groups were collected into a composite study group and will be referred to as LSCC group in the phylogenetic analysis.

Follow-up data of cancer patients were collected based on clinical records of the wards providing the samples. These data were available for all patients in the OSCC and LSCC groups, while for 63 of 87 patients from the CC group.

***PCR and RFLP***

After DNA extraction and confirmation of the DNA quality by amplification of the human β-globin gene using the primers PCO3 and PCO4, viral DNA was demonstrated by PCR. Detection of HPVs was performed using the consensus nested PCR with the primer pairs MY09/MY11 and GP5+/ GP6+. Presence of HPVs was double-checked using a PCR assay specific for the E7 ORF of HPV type 6, 11, 16, 18, 31, 33.

For demonstration of the TTV DNA two different nested PCR assays were performed. UTR PCR (using the primer pairs NG133/NG147 and NG132/NG134 targeting the untranslated region of the TTVs) detects all human TTV genotypes, while a seminested ORF1 PCR (using the primer pairs NG059/NG063 and NG061/ NG063 targeting the N22 region of ORF1) detects ggr1TTVs, i.e. species TTV1 and TTV3 according to the presently accepted taxonomy.

Genotyping of HPVs was performed on MY PCR products (first round of HPV amplification) by RFLP using six different restriction enzymes. When the analysis of the MY PCR product did not yield results due to low number of amplimers, the analysis was performed using GP PCR (nested round amplification) amplimers.

***Statistical analysis of results***

Prevalence results were compared using Fisher exact test or khi-square test as applicable. Survival was analyzed by Kaplan-Meier test. All analyses were carried out in the SPSS 15.0 for Windows software environment with a confidence interval of 95%.

***Sequencing and phylogenetic analysis***

For sequencing of ggr1TTVs the amplimers of the inner round of the ORF1 nested PCR were used. Amplimers were directly sequenced using the amplification primers. Analysis of the resulting sequences was performed with CLC DNA Workbench 4.0 software. Dendrograms were prepared using the neighbour joining method; validity was tested with bootstrapping of 1000 repetitions. A 187 bp long fragment successfully sequenced from all amplimers was used in the phylogenetic analysis; the flanking sequence parts were truncated. All unequivocal sequences from samples containing a single TTV were compared to GenBank archive sequences. To classify sequences according to the formerly accepted genogroup/genotype/subtype classification, GenBank sequences closely similar to our sequences with a genotype (for genotypes 1 and 2 subtype also) designation overlapping the 187 bp fragment determined were downloaded and used as reference sequences.

All 44 Hungarian ggr1TTV sequences available in the GenBank overlapping the 187 bp fragment determined were also downloaded and included in the analysis in order to determine their subtype (not available in the GenBank entry) and to compare them to our sequences.

Type sequences of the 29 species of the newly accepted Alphatorquevirus genus were also compared to genotype and subtype reference sequences used, to our sequences and to archive Hungarian sequences. Apart from the 29 type sequences, other TTV sequences classified according to the new taxonomy were not found in the GenBank.

## RESULTS

***Prevalence of TTV, ggr1TTV, and ggr1TTV-HPV coinfection in oral samples***

Based on UTR PCR, general TTV prevalence was 54.2%, 54.5%, 53.8%, 785% in the control, OL, OLP, OSCC group, respectively. TTV prevalence was significantly higher in OSCC patients compared to any other oral study group ($p=0.002$, $p=0.001$ and $p=0.011$ compared to controls, OL and OLP patients, respectively). TTV prevalence in healthy mucosal samples of patients with oral lesions was comparable to that found in the oral control group. When comparing the lesion and healthy mucosal samples within the same patient

group, only OSCC showed significant difference in general TTV prevalence (49.2% and 78.5%; p=0.003).

Based on ORF1 PCR, ggr1TTV was 1.4%, 4.5%, 10.1%, 24.6% in the control, and in the lesion samples of OL, OLP and OSCC patients, respectively. The ggr1TTV was more prevalent in the lesion of OSCC and OLP patients ($p<0.001$ and $p=0.034$, respectively) as compared to the controls. Between the OL and OLP lesions ggr1TTV prevalence did not differ significantly, but was significantly lower than in OSCC ($p=0.007$ and $p=0.01$, respectively). Healthy mucosa of patient groups showed prevalence comparable to the controls and there was no significant difference between ggr1TTV prevalence of the healthy mucosa of the different patient groups. When comparing the lesion and healthy mucosal samples within the same patient group, only OSCC showed significant difference in ggr1TTV prevalence (7.7% and 24.6%; $p=0.036$).

Prevalence and genotype distribution of HPVs was established in an earlier study of our working group. HPV prevalence was significantly higher in the lesions of each patient group as compared to the controls ($p<0.001$ in all cases), while when comparing these data between the patient groups, singificant difference was only found between OLP and OSCC ($p=0.047$). When comparing the patient groups regarding HPV prevalence in the healthy mucosa, they showed a prevalence pattern similar to that seen for the lesions. When comparing the lesion to the apparently healthy mucosa of the same patient group, lesion always showed a significantly higher HPV prevalence ($p<0.05$ in all comparisons). High risk genotypes, mostly HPV16 was found dominantly in all samples.

Coinfection with ggr1TTV and HPV was 0.0%, 4.5%, 6.7%, 12.3% in the control, OL, OLP and OSCC lesion samples, respectively. Coinfection rate was significantly higher only in case of OSCC when comparing the three patient groups with oral lesions to the controls ($p<0.001$). There was no diffence between samples from the apparently healthy mucosa of patient groups and the oral controls. Lesion samples of OSCC patients showed significantly higher coinfection rates as compared to healthy mucosal samples of the same group (1.5 % and 12.3%; $p<0.001$).

***Prevalence of TTV, ggr1TTV, and ggr1TTV-HPV coinfection in cervical samples***

General TTV prevalence was 78.4%, 83.3%, 90.8% in the control, ACM and CC group, respectively. TTV prevalence was significantly higher in patients with CC compared to the cervical controls ($p=0.018$). Lower prevalence of ggr1TTV was observed in cervical controls than in ACM or CC, but the difference was not significant. The latter two groups

showed comparable prevalence of ggr1TTVs (20.6%, 29.8% and 29.9% in the control, ACM and CC group, respectively).

Prevalence of HPVs was higher than in the controls both in ACM and CC groups (27.8%, 66.7%, 95.4% in the control, ACM and CC group, respectively); CC group showed a higher prevalence when compared to the ACM group as well ($p<0.001$ in all comparisons). Coinfection with ggr1TTV and HPV was significantly more prevalent in ACM and CC patients compared to the controls (3.1%, 20.2%, 29.9% in the control, ACM, and CC groups, respectively; $p<0.001$ in both comparisons). High risk genotypes were dominant in all study groups. The most frequent genotype was HPV33 in the control group, while HPV16 in the other two groups.

***Comparison of the prevalence of TTV, ggr1TTV and ggr1TTV-HPV coinfection in LSCC, OSCC and CC***

Prevalence of TTV and HPV did not differ in LSCC and OSCC, while prevalence of ggr1TTV and ggr1TTV-HPV coinfection was significantly higher in LSCC. When comparing LSCC to CC statistically significant differences were not detected. General TTV prevalence was higher in CC than in OSCC, but ggr1TTV prevalence was comparable. Coinfection was significantly more frequent in CC than in OSCC, but this is attributable to the very high HPV prevalence of CC samples, as all ggr1TTV carriers are HPV positives as well.

***Impact of ggr1TTV and ggr1TTV-HPV coinfection on survival of cancer patients***

The previously published follow-up data in the LSCC group has been updated. Both survival rate and survival time differed significantly between ggr1TTV infected and non-infected groups; both parameters show poorer results in infected patients [survival rate 9.1% (1/11 patients) *vs*. 78.6% (11/14 patients); mean survival time 289.5 (190.9-388.2) days *vs*. 1229.2 (659.6-1798.8) days; $p=0.0052$]. Coinfection with ggr1TTV and HPV further decreased survival rates and survival time as compared to those patients where coinfection was not found [survival rate 0.0% (0/8) *vs*. 82.4% (14/17); mean survival time 240.0 (148.4-331.6) *vs*. 1154.4 (638.8-1670.0) days; $p=0.0017$].

In the OSCC group neither ggr1TTV infection alone [survival rate 75.0% (12/16) *vs*. 71.4% (35/49); mean survival time 1436.9 (1133.9-1739.9) days *vs*. 1276.0 (1093.8-1458.1) days; $p=0.59$], nor together with HPV coinfection [survival rate 62.5% (5/8) *vs* 73.7% (42/57); mean survival time 1262.2 (875.6-1648.9) days *vs*. 1312.8 (1142.5-1483.1) days; $p=0.88$] influenced survival.

In the group of patients with CC follow-up data of 63 patients were available for analysis. Significant difference was not observed between ggr1TTV positive and negative patients [survival rate 77.3% (17/22) *vs*. 75.6% (31/41); mean survival time 2480.1 (1930.6-3029.6) *vs*. 2419.9 (2037.4-2802.5) days; p=0.91]. Due to the very high HPV prevalence (95.4%), all ggr1TTV positive patients were HPV positive at the same time; consequently the same data apply to the influence of coinfection on survival.

***Results of sequencing of TTV-specific amplimers***

TTV sequence could be determined in 83 samples of 73 patients. 18 samples of 17 patients proved to be infected by more than one ggr1TTV. The 187 bp long fragments available for all samples infected with a single ggr1TTV (65 samples of 57 patients) was deposited in the GenBank (FN689730-FN689794).

Based on the sequence comparison to type sequences of all 29 TTV species known all our determined sequences belonged to the species TTV1.

Sequences belonging to genotype 2 according to the former classification were the most frequently found, mainly the subtypes 2c (found in 21 samples of 20 individuals) and 2b (found in 16 samples of 14 individuals).These were followed by genotype 1 (12 sequences from 12 individuals were of subtype 1b and ten samples of eight individuals carried a subtype 1a sequence). Three samples of three individual harboured TTV sequences showed a genetic distance from subtype 2b and 2c close to the limit of genotype-level difference (0.3). These sequences were closest to a GenBank sequence for which subtype designation was not available and our study refers to these as subtype 2'. Three samples of three individuals carried sequences belonging to genotype 3, also showing high genetic distance (0.21-0.23) to the genotype 3 reference sequence.

TTV sequences from different samples of the same patient were available for eight patients. In five patients the different samples harboured the same subtype, while three patients carried different subtypes in different samples.

The analysis and typing of ggr1TTV archive (GenBank) sequences from Hungarian samples was also performed. Marked dominance of subtype 2c was observed among the 44 serum-derived sequences, 61.4% of all sequences and 95% of genotype 2 sequences proved to be subtype 2c.

The highest proportion of subtype 2c was observed in serum samples of patient with hepatitis, subtype 2b was not at all detected in this group. Subtype 2b was infrequent in samples from healthy blood donors and kidney transplant recipients as well. The three

genotype 3 TTVs originated from serum of kidney transplant recipients. A patient with hepatits harboured a double infection with subtype 2c and genotype 6 TTVs; the latter genotype is presently classified to species TTV3.

***Genetic diversity of TTV sequences***

Genetic distances within a subtype/genotype were between 0.00 and 0.13 among the sequences determined in the present study. Within subtypes 1a and 1b of genotype 1 the genetic distances were between 0.00-0.03; within genotype 3 they were between 0.02-0.04. The three subtypes of genotype 2 showed a highly variable pattern. Genetic distances were between 0.00-0.13; the latter value was found in case of subtype 2c.

Regarding all Hungarian sequences (including archive sequences as well) the highest diversity was also found within subtype 2c, where the distance between the two most distant sequences was as high as 0.15. Within subtype 2c the 48 sequences from 42 individuals were separated into three clusters (cluster A, B and C; the genetic distance within a cluster was ≤0.10). These included 34, ten and four sequences, of which 17, one and three sequences were derived from the present study, while the remaining ones were downloaded archive sequences. All three clusters were comprised of sequences from different sample types, though cluster B consisted of nine serum-derived sequences and a single sequence from cervical atypia and most sequences from the present study (15/17) in cluster A were of tumour tissue origin. High genetic diversity was found within subtype 2' and within genotype 3 as well, but due to the low number of sequences in these types clusters could not be defined. Within genotype 3, a sequence from an OSCC tissue sample and two sequences from ACM were close, while serum-derived genotype 3 sequences were more distant.

## DISCUSSION

An earlier study by our working group showed that ggr1TTV and HPV coinfection is significantly more frequent in tumour tissue samples of laryngeal cancer with tumour progression (metastasis or local recurrence) than in cases with non-progressing disease. This coinfection influenced adversely survival of patients with tumour progression, suggesting a potential cocarcinogenic role of the viruses. This observation at the same time raised the question whether ggr1TTV or its coinfection with HPVs play a role in other HPV-associated tumours, therefore the study was extended to investigation of malignant and potentially malignant lesions in the oral cavity and in the uterine cervix.

TTVs are ubiquitous in the healthy population, though comparative analysis of literature data is difficult as the various PCR assays used for detection amplify distinct fragments and differ in sensitivity and specificity. According to studies applying the same primer pairs as the present work, prevalence of ggr1TTVs in the serum of healthy individuals is approximately 30% in Asia, 16.3% in Europe and 19.3% in Hungary. Using the UTR primers, the general TTV prevalence in serum in Asia is 97% in Japan but only 45.3% in India, while in Europe is 87.5% in Italy and 51.3% in the Czech Republic. In Hungary this prevalence is 95%.

TTV prevalence in the oral and cervical control populations is variable; general TTV prevalence is 54.2% and 78.4% in oral and cervical control samples, respectively, while ggr1TTV was found in 1.4% and 20.6%, respectively. In the head and neck region, therefore, TTVs are less frequent, while in the region of the uterine cervix the prevalence seems to be similar to that generally found in the serum samples, though data are not directly comparable to literature data due to differences in samples and consequently the differences in the tissue environment, as well as to differences in detection methods. Directly comparable data on TTV prevalence in the head and neck region are lacking. A Czech study used the primers NG133/147 and NG132/134 in a nested PCR assay to survey TTVs in samples from cervical mucosa with atypia and from healthy controls. They found 52.7% prevalence in controls and 74.7% in samples with atypia. These prevalences are lower than those found in the present study (78.4% and 83.3%, respectively).

HPV prevalence in cervical samples corresponds well to literature data. General TTV prevalence in cervical samples was significantly higher in CC samples compared to healthy controls, while ggr1TTV was comparably prevalent in all three cervical sample groups. In the head and neck region, in cancer (OSCC and LSCC) samples the prevalence of both ggr1TTV and ggr1TTV and HPV coinfection was significantly higher than in any other head and neck sample population including the healthy controls. This suggests that ggr1TTV may be associated with lesions in the head and neck but not in lesion in the uterine cervix. In contrast, other TTV genogroups/genotypes apart from ggr1TTV may be connected to cervical lesions, as suggested by significant differences in general TTV prevalence. In cervical sample groups both general TTV and ggr1TTV showed a higher prevalence compared to samples from the head and neck region. These higher prevalence data may indicate that cevical mucosa provides an environment more favourable for replication of TTVs than head and neck mucosa. It is notable in case of the CC group that due to the very high prevalence of HPVs, all samples positive for ggr1TTVs detected are at the same time HPV positive as well. Thus the

coinfection rates are primarily determined by HPV prevalence, and as prevalence of ggr1TTV is comparable in all three cervical sample groups, the importance of coinfection is improbable in this anatomical region.

In OSCC patients, general TTV prevalence, prevalence of ggr1TTV and HPV as well as that of coinfection were significantly higher than in controls or in the apparently healthy mucosa of the patients. This raises the possibility that not only HPVs but also TTV/ggr1TTVs may play a role in this group. Comparing the present results to previous data, frequency of HPVs is comparable in OSCC and LSCC patients (47.7% and 48%, respectively), but the proportion of low-risk HPV types is higher in LSCC than in OSCC. General TTV prevalence, ggr1TTV prevalence and ggr1TTV-HPV coinfection rates are lower in OSCC than in LSCC patients (in case of OSCC 78.5%, 24.6% and 12.3%, while in LSCC 88%, 44% and 32%, respectively).

Dividing LSCC patients according to outcome into groups with and without tumour progression (metastatic or recurring tumours vs. no progression), the two groups showed significant differences in virus carriage rates, which seems to be associated with poorer survival, suggesting the ggr1TTV may play a role in progression of LSCC. However, such association was not found in OSCC or CC.

Among patients with the premalignant oral lesions examined, ggr1TTV prevalence was higher in OLP with an autoimmune pathomechanism as compared to OL mainly caused by mechanical irritation. This may suggest that presence of ggr1TTV is associated with the immunological alterations and with the presence of the tumour tissue rather than with the process of carcinogenesis.

Proportion of TTVs and ggr1TTVs in the healthy mucosa of patients with oral lesions did not differ from that found in oral controls. Similarly to HPV, the presence of TTVs, therefore, is associated with the lesion, also pointing to a possible connection between ggr1TTV and oral lesions.

Presence and replication of TTVs was demonstrated in PBMC. The argument that lower TTV prevalence in controls and in patient groups sampled with exfoliated cell colllection is due to lack of PBMC in these samples is contradicted by the higher prevalence of ggr1TTV in OLP than in OL and controls as well as by the higher prevalence of TTV in exfoliated cell based cervical samples, where prevalences were comparable to the CC tissue samples. If the presence of TTV were influenced only by the presence of PBMC, higher prevalence in tumours could be expected due to infiltrating lymphocytes. As TTVs were

detected in oral epithelial cells and in the saliva, epithelial cells may serve as a source for TTVs in the head and neck region.

Several studies postulate differences in tissue preference between different TTV genotypes/subtypes. It is concievable that persistent TTV infection and active viral replication is the consequence of immunological alterations or of the underlying disease, and that the virus alone is incapable of causing disease, but may enhance disease progression. In such situations different genotypes may behave different, some may be neutral, while other genotypes may contribute to disease progression.

Such differences in tissue preference may be explained by differences in genome structure and gene expression patterns between different genotypes, which, in turn, may influence pathogenicity and interaction with the host immune system. ORF2 of TTVs was shown to inhibit inflammatory response, interferon production and interferon-induced gene expression through blocking of the NF-κB pathway. Decreased local interferon production and inhibited inflammation may, for instance, aid HPVs to establish a nonproductive persistent infection, which may lead to viral carcinogenesis. Such enhancement may explain the association of higher prevalence of ggr1TTV-HPV coinfection with OSCC, LSCC and MP as compared to benign laryngeal papillomatosis as well as with the poorer outcome of people with coinfection in LSCC.

As pathogenicity of the viruses may be associated with genomic differences and as association of different TTV genotypes with different disease was suggested several times, genotype and subtype distribution of TTVs from the abovementioned samples as well as in other available Hungarian sequences was examined by comparison of partial ORF1 sequences. According to presently valid taxonomy TTVs formerly classified as genogroup 1 belong to species TTV1 (genotypes 1-5) and to TTV3 (genotype 6), for which species intraspecific taxons have not been defined as yet. In the available literature data the new species-based taxonomy has not been used, genotype and subtype level identification was necessary to allow for comparison to previous data. If disease association of TTVs manifests at the genotype or subtype level, as suggested several times previously, intraspecific genetic diversity of the species TTV1 and incorporation of the earlier genotype/subtype classification into the present species-based taxonomic system may represent a progress in the study of the virus. Based on the data presented here, adaptation of the former system for intraspecies classification of TTV1 is directly applicable. This may aid and facilitate comparisons with earlier data and the new taxonomy. To our knowledge, the report published in 2011 serving as

a basis for the present work is the first comparing previous and presently accepted taxonomic categories.

Classifying all Hungarian ggr1TTV sequences in the database according to the genotype/subtype system and categorizing them on the basis of anatomical regions, the genotype patterns in head and neck samples and serum samples are similar, but different from that found in cervical samples. The most frequent genotype was genotype 2, out of the 44 serum-derived sequences 32 (72.7%) was genotype 2 sequences, of which 27 (61.4%) belongs to subtype 2c. Comparing the subtype distribution in different patient groups, markedly high proportion of subtype 2c in hepatitis patients is notable; 79.2% of hepatitis samples showed this subtype. Such dominance raises the possibility that cells serving as a source of TTVs in the serum may be more strongly connected with this subtype, or that this subtype may influence the development or the course of hepatitis.

Most studies on TTVs examine virus prevalence in patients and controls without subtyping, studies performing genotyping tend to work with lower sample numbers. Studies investigating genogroup distributions reported dominance of ggr1TTV. In the present study 17 of 73 ggr1TTV infected patients showed dual infection, but this proportion is expected to be higher if testing for all known genogroups and genotypes. The high prevalence rates with UTR PCR indicate that TTV not belonging to genogroup 1 are also present in high proportions.

Numerous Asian studies reported the dominance of genotype 1 over other genotypes. A study from the Czech Republic reported a genotype distribution in serum samples similar to that found by the present study; notably, this latter study used the same primers on a comparable number of samples. Dominance of subtype 2c was reported on smaller-sized sample collections from Germany, Turkey, Greece, Saudi-Arabia and Portugal. Genotype 2 without subtype data was found to be dominant in another Czech study, in Italy, Spain and in the USA. A worldwide study reported that the proportion of genotype 1 was highest in Asia, but the dominance of genotype 1 was also reported from all continents excepting Europe. Besides genotype 2, in some European countries other types are also locally prevalent, genotype 3 in Greece and genotype 4 in Spain and Portugal. Genotype 3, if not in high proportions was also detected among Hungarian samples, one OSCC, two ACM and three serum samples (of ten kidney transplant recipients) carried this genotype.

Though the overwhelming majority of the samples tested for TTVs, including the abovementioned studies, originate from the serum, comparing the data of the present study to

literature data indicate that, similarly to other viruses, geographical differences may characterize TTV genotype distribution as well.

Pairwise comparison of genetic differences revealed variable patterns. Regarding Hungarian sequences, within the subtypes of genotype 1 the genetic diversity was very low, clusters could not be identified. In contrast, genotypes 2 and 3 were more heterogenous, genetic distances sometimes are close to the limit of subtype-level difference. Within the subtype 2c, to which the most Hungarian sequences belong, three distinct clusters could be identified with a genetic distance less than 0.10 within a cluster. Cluster A was dominated by cancer-derived sequences (15/17) regarding sequences determined in the present study, while in cluster B serum-derived sequences were characteristic.

In summary, ggr1TTVs (belonging to species TTV1 and TTV3 according to the newest taxonomy) may play a role in development and/or progression of malignant and potentially malignant lesions of the head and neck region alone or in coinfection with HPVs. To correctly assess the role of TTVs may necessitate genotype/subtype determination, i.e. characterization of the TTVs to a level below species, as the role of these viruses seems to be genotype dependent. The presented data also draw attention to the probable differences in subtype distribution in different organs and tissues and to the probable geographical and/or race-based differences in subtype distribution, which may fundamentally influence the distribution in the population studied. Consequently, for assessment of the pathogenetic role of the virus, understanding of the genetic diversity within differen populations and tissues is crucial.

## SUMMARY

An earlier study of our working group demonstrated that genogroup 1 torque teno virus (ggr1TTV) and coinfection with human papillomavirus (HPV) was associated with poorer survival in laryngeal carcinoma patients with progressing disease compared to patients without tumour progression. This suggests that the virus, if not a causative agent, may influence unfavourably the disease course. Present study determined the general prevalence of TTV, ggr1TTV (presently species TTV1 and TTV3) HPV (in cervical samples) and ggr1TTV-HPV coinfection as well as genotype and subtype distribution of genogroup 1 TTV sequences in samples from the oral cavity and from the region of the uterine cervix. Tumour tissue samples (oral squamous cell cancer, cervical cancer) and exfoliated cells from precancerous lesions (oral lichen planus, oral leukoplakia and cervical atypia) were collected together with exfoliated cells from apparently healthy mucosa of patients and from the control patients.

ggr1TTV was found to be associated with tumours and oral lichen planus in the head and neck region. The immunological alterations, the presence of the tumour tissue or the process of carcinogenesis support replication of the virus and genogroup 1 TTV may contribute to initiation or progession of the diseases. TTV prevalence was significantly higher in cervical cancer patients than in control samples, while prevalence of ggr1TTV was comparable in the three cervical groups, which may suggest a role of a TTV genogroup distinct from genogroup 1. Frequency of TTV as well as ggr1 TTV was higher in cervical than in head and neck samples and HPV showed a similar distribution, which may indicate that this region is ideal environment for replication of these viruses.

Several authors postulated association between certain TTV genogroups, genotypes or subtypes and certain diseases, and these may be explained by differences in gene expression patterns of different TTV genotypes. ggr1TTV sequences were determined from our positive samples and compared to Hungarian sequences available in the GenBank as well as to other published sequences. In Hungary, similarly to European data, subtype 2c TTV is the most prevalent, typically in serum samples. According to the novel, recently accepted TTV taxonomy, all sequences determined belong to species TTV1. Intraspecific classification is presently undefined, however, as shown for different genotypes, subgroups may differ in pathogenetic potential. The earlier classification of genotypes and subtypes may be directly applicable for intraspecific classification of the TTV1 species.

Register Number: DEENKÉTK /100/2012.
Item Number:
Subject: Ph.D. List of Publications

Candidate: Enikő Fehér

Neptun ID: GQ6MU6

Doctoral School: Doctoral School of Pharmaceutical Sciences

## List of publications related to the dissertation

1. **Fehér, E.**, Kardos, G., Gáll, T., Kis, A., Gergely, L., Szarka, K.: Comparison of diversity of torque teno virus 1 in different mucosal tissues and disorders.
   *Acta Microbiol. Immunol. Hung. 58* (4), 319-337, 2011.
   DOI: http://dx.doi.org/10.1556/AMicr.58.2011.4.8
   IF:0.625 (2010)

2. **Fehér, E.**, Gáll, T., Murvai, M., Kis, A., Boda, R., Sápy, T., Tar, I., Gergely, L., Szarka, K.: Investigation of the occurrence of torque tenovirus in malignant and potentially malignant disorders associated with human papillomavirus.
   *J. Med. Virol. 81* (11), 1975-1981, 2009.
   DOI: http://dx.doi.org/10.1002/jmv.21627
   IF:2.47

## List of other publications

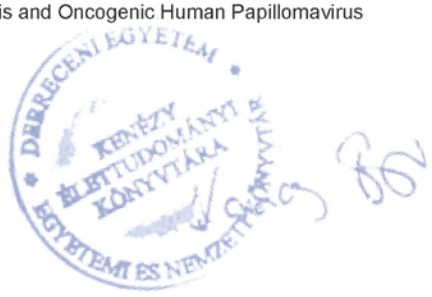
3. Gáll, T., Kis, A., **Fehér, E.**, Gergely, L., Szarka, K.: Virological failure of intralesional cidofovir therapy in recurrent respiratory papillomatosis is not associated with genetic or epigenetic changes of HPV11: Complete genome comparison of sequential isolates.
*Antiviral Res. 92* (2), 356-358, 2011.
DOI: http://dx.doi.org/10.1016/j.antiviral.2011.09.007
IF:4.439 (2010)

4. Kis, A., **Fehér, E.**, Gáll, T., Tar, I., Boda, R., D. Tóth, E., Méhes, G., Gergely, L., Szarka, K.: Epstein-Barr virus prevalence in oral squamous cell cancer and in potentially malignant oral disorders in an eastern Hungarian population.
*Eur. J. Oral. Sci. 117* (5), 536-540, 2009.
DOI: http://dx.doi.org/10.1111/j.1600-0722.2009.00660.x
IF:1.956

5. Szarka, K., Tar, I., **Fehér, E.**, Gáll, T., Kis, A., D. Tóth, E., Boda, R., Márton, I., Gergely, L.: Progressive increase of human papillomavirus carriage rates in potentially malignant and malignant oral disorders with increasing malignant potential.
*Oral Microbiol. Immunol. 24* (4), 314-318, 2009.
DOI: http://dx.doi.org/10.1111/j.1399-302X.2009.00516.x
IF:2.336

6. Szalmás, A., Bánáti, F., Koroknai, A., László, B., **Fehér, E.**, Salamon, D., Gergely, L., Minárovits, J., Kónya, J.: Lineage-specific silencing of human IL-10 gene expression by promoter methylation in cervical cancer cells.
*Eur. J. Cancer. 44* (7), 1030-1038, 2008.
DOI: http://dx.doi.org/10.1016/j.ejca.2008.02.046
IF:4.475

7. **Fehér, E.**, Szalmás, A.: Prevalence of Chlamydia trachomatis and Oncogenic Human Papillomavirus Types in Cytologic Atypia of the Uterine Cervix.
*Acta Microbiol. Immunol. Hung. 53* (4), 479-487, 2006.
DOI: http://dx.doi.org/10.1556/AMicr.53.2006.4.5



**Total IF: 16.301**

**Total IF (publications related to the dissertation): 3.095**

The Candidate's publication data submitted to the Publication Database of the University of Debrecen have been validated by Kenezy Life Sciences Library on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

29 March, 2012

## LIST OF PRESENTATIONS RELATED TO THE DISSERTATION

**Enikő Fehér**, Judit Kiss, Ildikó Tar, Krisztina Szarka. Correlation of HPV prevalence, clinical and patient data in individuals with oral precancerous lesions. 15th International Congress of the Hungarian Society for Microbiology. 2007. Budapest, Hungary.

Krisztina Szarka, Annamária Szabó, **Enikő Fehér**, Györgyi Szládek, Etelka D. Tóth, Ildikó Márton, Ildikó Tar. Etiological role of HPV and TTV infection in oral carcinogenesis. 15th International Congress of the Hungarian Society for Microbiology. 2007. Budapest, Hungary.

Krisztina Szarka, **Enikő Fehér**, Melinda Murvai, Tamás Gáll, Etelka D. Tóth, Tamás Major, Tamás Sápy, Ildikó Tar, Lajos Gergely. Possible role of genogroup 1 torque-tenovirus infection in human papillomavirus associated carcinogenesis. Annual Meeting of the Hungarian Society for Microbiology. 2008. Keszthely, Hungary.

Krisztina Szarka, **Enikő Fehér**, Tamás Gáll, Etelka D.Tóth, Róbert Boda, Ildikó Márton, Ildikó Tar. Investigations on the role of human papillomaviruses in oral carcinogenesis. Annual Meeting of the Hungarian Society for Microbiology. 2008. Keszthely, Hungary.

**Enikő Fehér**, Gábor Kardos, Attila Nochta, Csilla Péter, Lajos Gergely, Krisztina Szarka. Invesigation of the subtype distribution of genogroup 1 torque teno virus in mucosal samples collected from the uterine cervix and the head and neck region. Annual Meeting of the Hungarian Society for Microbiology. 2010. Keszthely, Hungary