Closing report to the project 128533 entitled

„Human polyomaviruses: functional analyzis of genomes and prevalence study"

The aims of the project were (I) to study the seroprevalence of novel, human pathogenic members of the *Polyomaviridae* family; and (II) to study *in vitro* the regulatory function of different non-coding control region (NCCR) on bi-directional gene expression of polyomavirus in different cell types. Seroprevalence studies help to answer many questions regarding the pathogenesis and the clinical importance of the viruses: to determine the time of the primary infection, the proportion of the infected individuals within population, and the seroreactivity level; and to reveal the susceptibility and possible risk factors for infections within the population. Even if other research teams publish some data, it will have high importance to collect more data, especially because the seropositivity rates varied in a big range for some viruses. It suggests geographical differences in infection rates and/or geographical genetic variability of the viruses. *In vitro* viral replication experiments are essential to reveal the pathogenesis of the viruses. Due to the lack of the viral isolates, to find a possibly permissive cell for virus isolation, and for *in vitro* experiments is possible by studying bi-directional gene expression regulated by the viral non-coding control region.

The seroprevalence study of the work plan has been completed with 10 human pathogenic polyomaviruses, two scientific papers have already been published, and two manuscripts are in progress. High number of serum samples — 619-1038 samples in duplicates per viral species — were analysed from different patient and age groups enabling a comprehensive, detailed analysis of the results (detailed below). Beside the scientific results and data, we developed suitable methods to continue the research topic, and these may establish the development of diagnostic methods if there will be a clinical requirement.

Several *in vitro* experiments have been also completed (some are in progress now), and the results will be published (a manuscript is being prepared, as detailed below).

Although the pandemic and its consequences caused a significant problem in the field of my research, and for me personally (as detailed below), due to the diagnostic work and the sample, it was also possible to study DNA prevalence of polyomaviruses in respiratory samples. This ensures valuable data for the functional genome analysis study, as well.

Unfortunately, the research progressed much slower than planned, and although the extension of the project by one year meant a lot, it is very difficult to make up for the extremely difficult 2-year period of work. In March 2020, my working conditions changed significantly. Medical Microbiology, University of Debrecen became a laboratory for SARS-CoV-2 diagnostics designated by the Minister of the Ministry of Human Resources, and its supply area was several counties with about 1.7 million inhabitants. It was a mandatory diagnostics activity due to my qualification as a clinical microbiologist, and then because of my compulsory healthcare status (we could not decide about it, our previous academic status was changed to healthcare status). Continuous seven-day working weeks and overtime significantly exceeding normal working hours (for long term double or more) severely limited

the time available for research. Laboratories and equipment were used for diagnostics, limiting research opportunities, resulting in wasting previously purchased reagents (of course, it limited the opportunities later). As a clinical microbiologist, one of the most experienced specialists in PCRs and other tasks, my bosses assigned many tasks to me: developing diagnostic methods, purchasing reagents, equipment and instruments, coordinating work processes and employees, providing data to superior bodies, beside which I performed manual daily work, PCRs, validated and reported the results. I belong to the minority of clinical microbiologists who, due to university obligations, also teach a significant number of hours, and I carry out public activities as secretary of a doctoral committee of 10 doctoral schools. These were also compulsory activities, which are normally well compatible with research work. Of course, as for everyone, purchasing reagents, equipment, cells were (and are now) complicated, the shipment takes extremely long period. Despite all of this, however, I tried to carry out research work, since I work in a research field where long periods cannot be missed. During the last 1.5 half year we completed most of the experiments planned, but a part of the result will be published in the near future (as detailed below).

I. Seroprevalence study supplemented with DNA prevalence*:*

1. We studied the seroprevalence of KIPyV and WUPyV by enzyme-linked immunosorbent assay (ELISA) among children and adults. The major capsid proteins (VP1) of the polyomaviruses were produced as an antigen for ELISA. VP1 genes were inserted into the pTriEx™-4 Neo vector, protein expression was carried out in Origami™ B(DE3)pLacI and Rosetta-gami™ B(DE3)pLacI competent cells. Protino Ni-TED Packed Columns was used to purify the viral proteins from the inclusion bodies. Following dialysis and concentration, the viral proteins were analysed qualitatively and quantitatively. An indirect ELISA was developed, including determination of the optimal coating concentration, blocking buffer and concentration, dilutions of primary and secondary antibodies, incubation time, temperature, washing buffer and cycles. The WUPyV ELISA was performed by using 705 serum samples (373 from children and 332 from adults), while the KIPyV antibody was screened in 692 samples (325 children and 367 adult sera). The number of the serum samples analysed exceeded the numbers planned. The overall seropositivity in this study was 82.1% for KIPyV and 79.1% for WUPyV. At the same time, the adult seropositivity was 93.7% for KIPyV and 89.2% for WUPyV. The sex distribution of the patients seropositive for KIPyV and WUPyV was not different from the total group of the patients. Moreover, there was no difference between the children and adult groups studied. The age of the seropositive patients was significantly higher than the age of the total population studied for both KIPyV and WUPyV. There was a statistically significant age difference between seropositive and seronegative patients within the cohort, and also within the group of children for both viruses, but not within the group of adults. In accordance with data published by others, our seroprevalence results suggest that primary infection by both KIPyV and WUPyV occurs during childhood: we detected 55% and 45.5% seropositivity for patients < 2 years of age for KIPyV and WUPyV, respectively. The seropositivity rate increased with age, reaching the maximum level > 60 years.

We also determined the prevalence of KIPyV and WUPyV DNA in 1030 nasopharyngeal samples from patients who had tested positive or negative for SARS-CoV-2 RNA. KIPyV DNA was detected in 0.49% samples of which 0.29% were SARS-CoV-2-positive samples and 0.86% were SARS-CoV-2-negative samples. The WUPyV DNA prevalence was 0.19%: positivity was 0.15% for SARS-CoV-2-positive samples and 0.29% for SARS-CoV-2-negative samples. Although our prevalence data are in accordance with publications and also with our previous data, but these are markedly lower KIPyV and WUPyV positivity in SARS-CoV-2-positive patients than published by an Italian team. Since we did not find geographically different seroprevalence of these viruses, the explanation for the difference in DNA prevalence might be that our samples were collected after a long-term lockdown due to the COVID-19 pandemic, while the Italian team collected and analysed samples at the beginning of the pandemic. This hypothesis is strengthened by the publications which proved that long-term lockdown significantly influenced the spread of respiratory transmitted viruses. **These results are already published:** Katona Melinda, Jeles Krisztina, Kovács Renátó, Csoma Eszter: KI and WU Polyomaviruses: Seroprevalence Study and DNA Prevalence in SARS-CoV-2 RNA Positive and Negative Respiratory Samples, MICROORGANISMS 10: (4) 752, 2022

2. The seroprevalence of human polyomaviruses (HPyV) linked to skin diseases, namely Merkel cell polyomavirus (MCPyV), HPyV6, HPyV7 and Trichodysplasia spinulosa–associated polyomavirus (TSPyV) were studied by our in-house ELISA to detect IgG antibodies. Because the published seroepidemiological data vary widely for the viruses studied, we performed a comprehensive analysis with data available from the literature. We compared published age-specific seropositivities with our data. 552-552 serum samples were analysed for each virus, the study population ranged from 0.8 to 85 years. The overall seropositivity was 63.9%, 79.2%, 72.5% and 78.4% for MCPyV, HPyV6, HPyV7 and TSPyV, respectively, while the adulthood (>18 years) seroprevalence was 69.3%, 87.7%, 83.8% and 85% for MCPyV, HPyV6, HpyV7 and TSPyV, respectively. The seropositivity for all four polyomaviruses increased significantly with age. Our data strengthens the findings that primary infection occurs in early childhood: the vast majority is possible in children and young adults, but it may happen throughout life. There was a significant increase in seroprevalence in different age groups for the studied polyomaviruses, which is in agreement with the results of seroreactivity analysis: the age groups with significantly higher seropositivity showed significantly higher OD values for HPyV6, HPyV7 and TSPyV. After reaching the adulthood level of seroprevalence in the 14–21 years age group (for MCPyV and TSPyV) or the 21–40 years age group (for HPyV6 and HPyV7), the antibody levels against the viruses remained relatively stable. There was a significant increase in seroprevalence in the oldest age group (>60 years) for TSPyV. This phenomenon might be the consequence of reactivation of latent infection. The available seroprevalence data cover a wide range for each virus and differences can be observed not only in the overall data, but also in the trends regarding how seropositivity changes with age. Although the cohorts, methods and even the antigens vary among the studies, there are similar results to our data proving the suitability of our methods. For MCPyV, geographically distinct genotypes might exist that might underlie differences in reported seroprevalence. **These results are already published:** Jeles Krisztina,

Katona Melinda, Csoma Eszter: Seroprevalence of Four Polyomaviruses Linked to Dermatological Diseases: New Findings and a Comprehensive Analysis, VIRUSES 14: (10) 2282, 2022

3. Seroprevalence of HPyV9, Malawi, Saint Louis and New Jersey polyomaviruses (MWPyV, STLPyV and NJPyV) were studied using the previously detailed ELISA methods with antigen expressed and purified. Based on our results from the analysis of >2500 sample data, HPyV9 and NJPyV are the least common viruses, the adulthood seroprevalence is < 50%. Our results are in agreement with the international studies (however the number of the available scientific papers in the field are very low). Unfortunately, we detected high rate of cross-reactivity with other polyomaviruses using the VP1 antigen of NJPyV.

Altogether 1038 samples were testes for the presence of anti-HPyV9 antibody. Significant differences in seroreactivity were observed between the under 3 years and 3-6 years age groups and between the 14-21 and 21-40 years age group. Seropositivity increases with age, a significant difference was detected between the age group 10-14 and 14-21 years, reaching a maximum 37 % seropositivity in adults. Within the >80 years group a slightly decreased seropositivity was observed. **These results will be published together with DNA prevalence data, and the results from in vitro, functional genome analysis (detailed below).**

For MWPyV and STLPyV seroprevalence study 619 samples from children and adults (0.8-90 years) were analysed. The rates of patients with past MWPyV infection increased with age reaching 55% seropositivity among adults. We observed significant elevation in seropositivity and seroreactivity between the age groups <6 and 6-10 years, and between the age groups 10-14 and 14-21 years. In contrast, STLPyV primary infection might occur mostly in the early ages, since the anti-viral antibodies were detected also in high rate (59%) among children < 6 years. A significant elevation of the seropositivity and seroreactivity was observed in age group 14-21 years. The rate of seropositivity also increases with age among adult reaching the highest 83 % among adults >60 years. These high rates of infections suggest a transmission via the respiratory route, so we supplemented the seroprevalence study with DNA prevalence examination: nasopharyngeal and pharyngeal swab samples, middle ear discharge samples and secondary lymphoid tissue samples were analysed for the presence of MWPyV and STLPyV DNA. Both viruses were detected in respiratory swab samples and middle ear discharge samples (the positivity rates were 3-10 % depending on the ages), and STLPyV was detected in an adenoid tissue sample (1%). The viral non-coding control region was sequenced revealing some new point mutation which might have effect on the rate of viral replication. This result establishes further *in vitro* experiments. **A manuscript is being prepared** to submit into a Q1 journal (Journal of Clinical Virology is planned).

II. Functional genome analysis:

To determine the biological importance of the differences in non-coding, control regions (NCCRs) of polyomaviruses, KIPyV, WUPyV and HPyV9, *in vitro* experiments were planned.

After synthesis, NCCRs were cloned into bi-directional expression vector, amplified in bacterium, purified, and then sequenced. Some point mutations and combined mutations were created by site direct mutagenesis using specific primers designed for it. Site direct mutagenesis was successful; all the vectors contain the adequate NCCR checked by sequencing. These reporter systems enable the expression of two luciferases regulated by the NCCR inserted: the firefly luciferase encoded on one strand of the vector and the renilla luciferase on the other strand in opposite directions. Previously we planned to study fluorescent gene expression using flow cytometer, but this luciferase reporter vector is also useful to study bi-directional control of expression. It was easier to obtain (which had high importance during the period when purchasing anything was complicated and took extremely long time) and we have not only the instrument required for the measurements, but also experience with this type of reporter vector system.

Up to now, we have no information about the exact route of transmission, the cell tropism, and the site for possible latent infection for these viruses. Transient transfection of different cells enables to study early and late promoter strength, which may help to find the possibly permissive cells for the viruses to be able to isolate and propagate them for further experiments. To determine the biological importance of the differences in NCCRs, early and late promoter strength was studied in different human cell lines (kidney epithelial, lung epithelial and fibroblast, and endothelial cells) and in primary small airway cells. The effect of LTAg on the early and late promoter activities were also examined, for which LT-antigen expressing vector was also produced successfully. The expression of the tagged protein was proven by Western-blot analysis.

Transient transfections, co-transfections and dual luciferase reporter assays were optimized successfully for all cell types. Experiments with HPyV9 have been completed and a manuscript is being prepared (as mentioned above). Experiments with WUPyV and KIPyV are in progress. Based on our results, early and late promoter activities of those NCCRs that have point mutations compared to reference sequences were stronger in each cell types; and we observed stronger effect of LTAg on these NCCRs. Gene expression regulated by polyomavirus NCCRs were the highest level in kidney and lung epithelial cell lines. LTAg resulted in significantly stronger late promoter activity, which together with the suppression of the early promoter activity may be the key for productive viral replication.

Despite the unexpected, both personally and globally particularly difficult period, the research project mostly completed successfully, some experiments are in progress, but will be finished in the near future. Although the time point of the final report we published only two scientific articles, but two additional manuscripts are being prepared, and one more is planned. The results were presented on international and national conferences. The results, the methods and *in vitro* experiments developed during the project serve a good basis to continue the research in order to reveal further the clinical importance and the pathogenesis of human pathogenic viruses described as novel viral genomes.