Transcriptome analysis of Pseudorabies virus reveals selective regulation of different kinetic classes of genes

Nándor Póka

PhD Thesis

University of Szeged

Faculty of Medicine

Department of Medical Biology

Supervisor: Prof. Zsolt Boldogkői

Multidisciplinary Sciences PhD School

of Health and Welfare Doctoral Programme

2017

Szeged

List of publications used in this thesis

I. Takács, I.F. et al., 2013. The ICP22 protein selectively modifies the transcription of different kinetic classes of pseudorabies virus genes. *BMC molecular biology*, 14, p.2.

II. Oláh, P. et al., 2015. Characterization of pseudorabies virus transcriptome by Illumina sequencing. *BMC Microbiology*, 15(1), p.130.

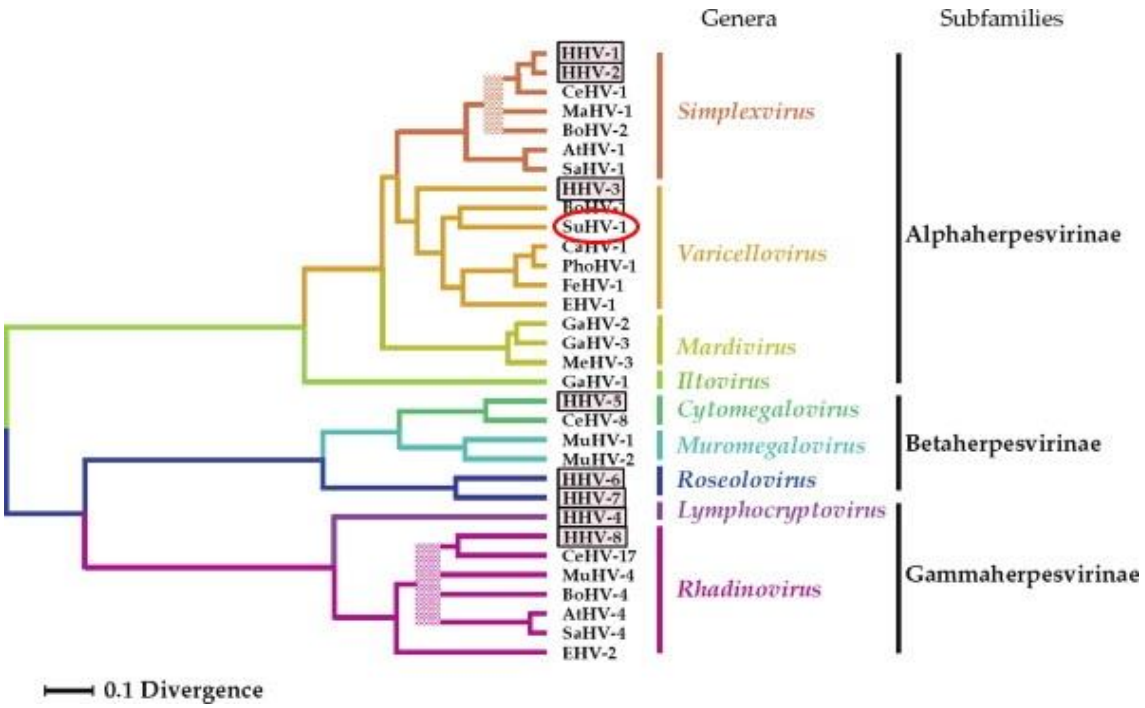
III. Póka, N. et al., 2017. Deletion of the us7 and us8 genes of pseudorabies virus exerts a differential effect on the expression of early and late viral genes. *Virus Genes*, pp.1–10.

Table of contents

Introduction

History and importance

Pseudorabies virus (or by its taxonomic name *Suid herpes virus 1* – SuHV-1) is a member of the Alphaherpesvirinae subfamily within the family Herpesviridae, (see Figure 1 for the phylogenetic tree of herpes viruses) that was first documented by a Hungarian veterinary Aladár Ajueszky in 1902 (Ajueszky 1902). It is a pathogen of swine that causes severe infections and economic losses around the world – in late 1990's, early 2000's, the estimated annual cost of eradication / prevention was $30 Million (approx. 43.8 Million at 2017 value) in the USA. Piglets younger than 3 months and sows are the most susceptible to the infection, and mortality is nearing 100% among the young animals (< 3 weeks old) (Pomeranz et al. 2005) while infected sows suffer from approx. 50% abortion rate (Woźniakowski & Samorek-Salamonowicz 2015). In nursing piglets early symptoms are fever, listlessness and stolidity towards nursing. In the following 24 hours after the first symptoms have emerged, progressive development of signs of central nervous system infection such as trembling, ataxia, seizures or



1. Figure. The phylogenetic tree of herpes viruses based on amino acid sequence alignments. The human herpes viruses are boxed for better identification. The Pseudorabies virus – abbreviated here as SuHV-1 after its taxonomic name *Suid herpes virus 1* – is outlined with red. The PRV belongs to the *Varicellovirus* genus, and is close relative of several human herpes viruses, for example the *Human herpes virus 1*. Original image: (STRAUSS et al. 2008)

excessive salivation occur. After showing signs of central nervous system infection, infected animals die within 24-36 hours. Mortality decreases with the age of the infected animals with 50% being typical for piglets around the age of 1 month, and with careful treatment typically less than 10% of the 1,5-2,5 month old infected animals die. For adult animals mortality is low – approximately 1-2% -, but complications such as pneumonia may occur.

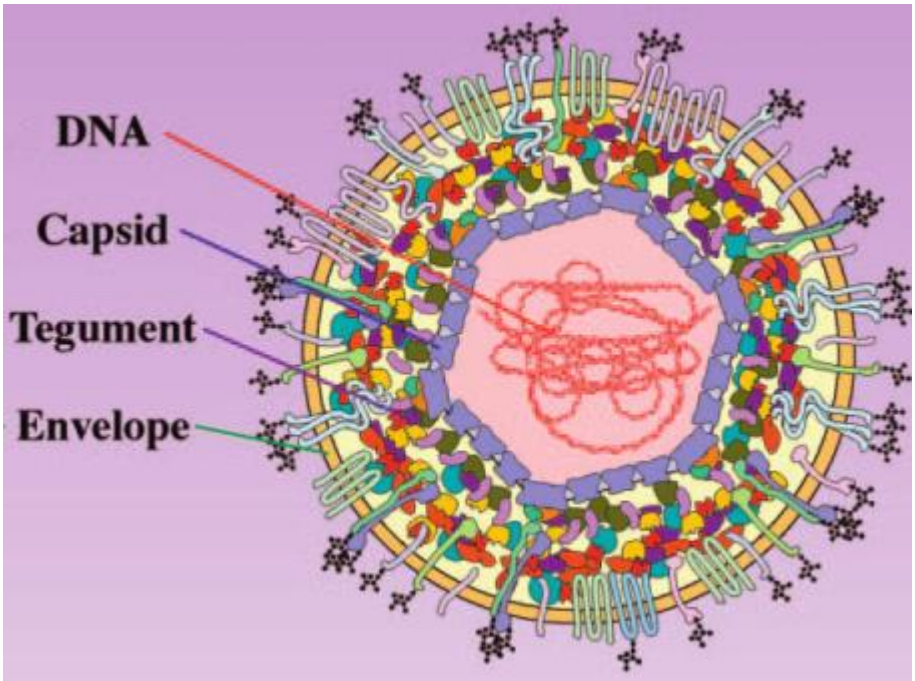
Although, PRV was officially eradicated from the United States (Anderson et al. 2008) and most European countries, there are still reports of outbreaks in China (Wang et al. 2015) (See Figure 2 for outbreaks in China in 2013), and the virus is also present in other areas world in either domestic swine or wild boars (Yamane et al. 2015; Verin et al. 2014; Verpoest et al. 2014). PRV is also able to infect a broad range of hosts outside the Suidae family including sheep, cats, dogs, rabbits, guinea pigs and more (Woźniakowski & Samorek-Salamonowicz 2015). Newly emerging strains may also decimate previously vaccinated swine populations (Yu et al. 2016). Transmission to humans is circumstantial, and disease does not manifests in immunocompetent recipients, thus it is considered safe for laboratory use. Owing to its economic importance and its relatedness to human pathogen herpes viruses, PRV is extensively researched and used as a model system for other herpes viruses. The ability of PRV to naturally infect synaptically connected neurons (Strack & Loewy 1990) (spreading in anterograde

2. Figure. Areas of China affected by the Pseudorabies virus are marked pink. Black triangles show cities with seropositive samples. Original image: (Wu et al. 2013)

direction) makes it a useful live neural tracer and activity reporter (Boldogkoi et al. 2009), further extending its research value.

Overview of PRV molecular biology

Like other members of the Alphaherpesvirinae subfamily, the PRV has double stranded DNA genome surrounded by an icosahedral capsid, forming the so called nucleocapsid. The nucleocapsid is embedded within the tegument – a matrix of various proteins – all of this enclosed in the lipid membrane called envelope. Figure 3 illustrates the structure of a PRV virion. Within the envelope are many viral glycoproteins that are responsible for viral attachment to cell surface and entry of the virion into the host cell. PRV has a compact genome of approximately 144kbp harboring 67 protein coding genes. Similar to the prototypical alpha herpesvirus genome of HSVI-1, the genome of PRV has two unique regions – long and short – termed UL and US, the latter surrounded by the inverted internal and terminal repeat sequences (IRS and TRS respectively). Gene nomenclature is based on the region (UL or US) and position where the gene resides. Due to the small genome harboring relatively high number of genes and



3. Figure The strucutre of the PRV virion. In the center is the double stranded DNA of the virus that is surrounded by the icosahedral capsid. The capsid is embedded in the tegument – a loose matrix of various proteins. The final layer is the envelope with several proteins anchored within. Original figure: (Pomeranz et al. 2005).

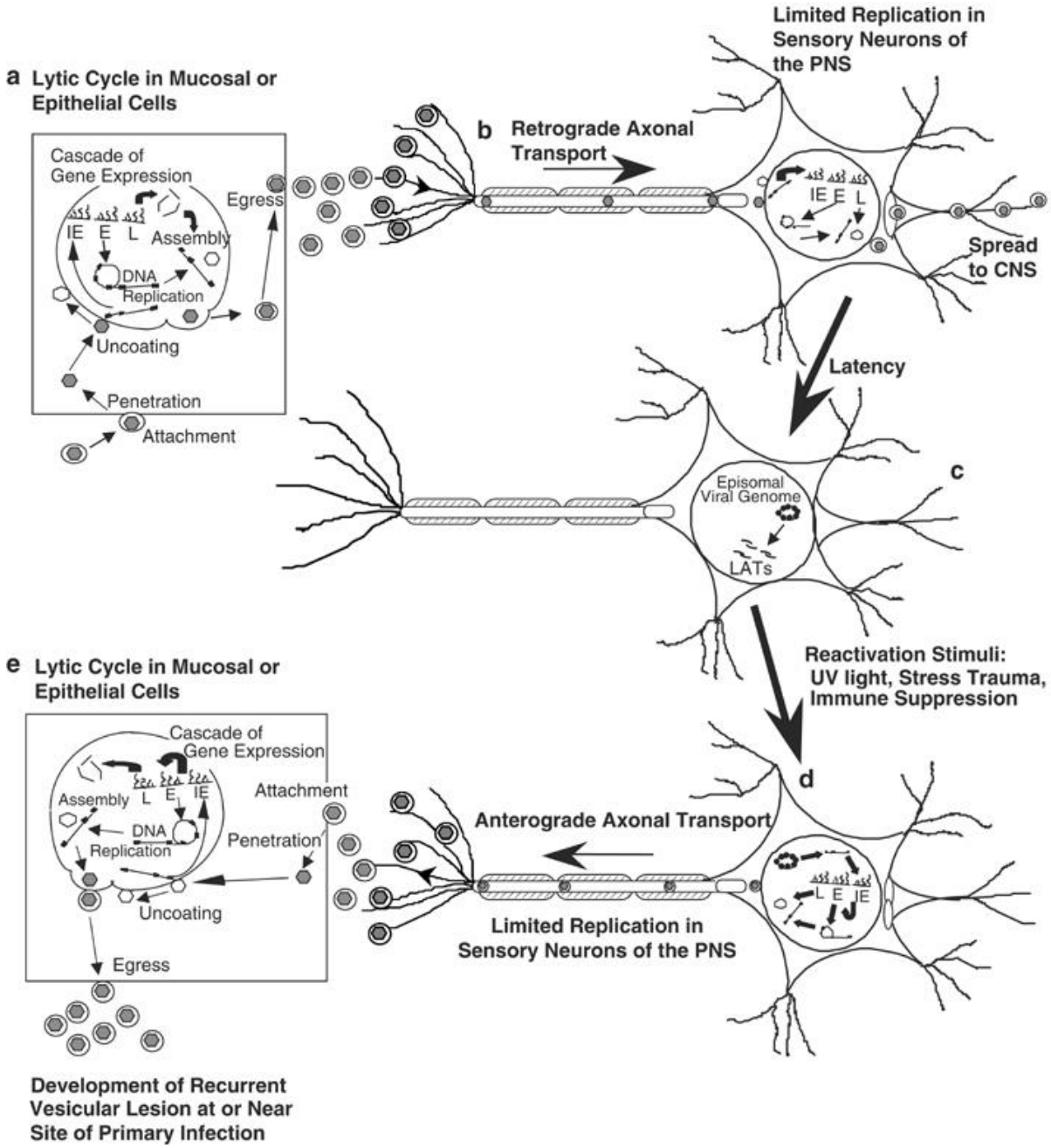
the lack of large intergenic regions, many of the PRV's genes are nested within each other, resulting in 3' co-terminal transcripts. (Oláh et al. 2015)

The 67 protein coding genes can be classified into four kinetic classes based on their peak expression after infection. Immediate early (IE) genes are transcriptional activators that activate early genes required for DNA multiplication and shutting down host defense mechanisms. The only immediate early gene in PRV, called IE180, is detectable as early as 40 minutes post infection (p.i). Early (E) and early-late (E/L) genes are responsible for the multiplication and nucleotide metabolism. Their expression peaks at approximately 3-4 hours after infection. Late (L) genes encode for the structural proteins of the virus and are responsible for viral egress and cell-to-cell spreading. Detectable amount of late gene products are found at 3 hours after infection, with amounts gradually increasing to higher levels later on. A detailed review on the molecular biology of PRV please see reference (Pomeranz et al. 2005).

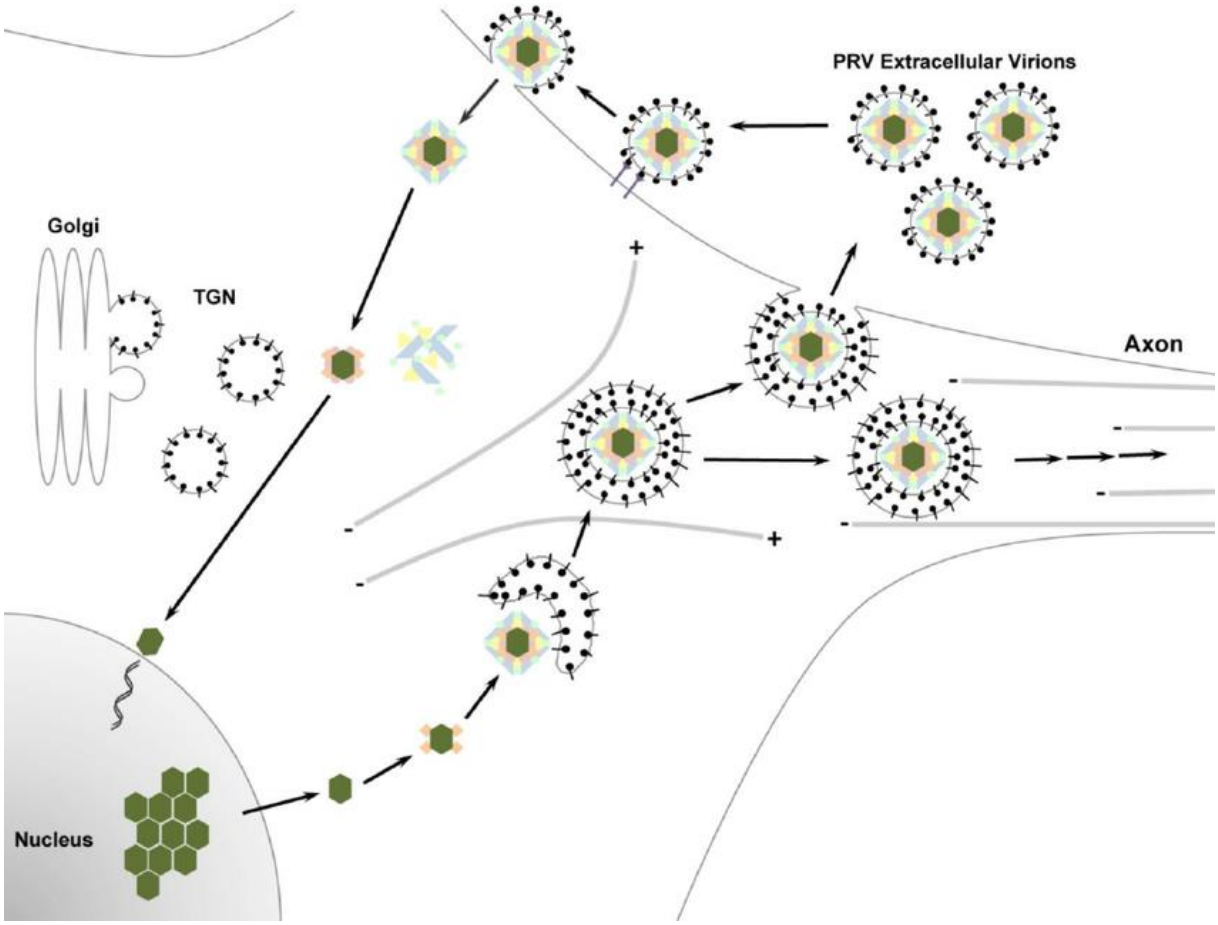
Viral life cycle

Viral infection starts on the periphery and spreads to the central nerve system, where the virus can establish latent infection in sensory ganglions (like other alpha herpes viruses), from which it can reactivate to its lytic reproductive phase. The prototypic global viral lifecycle is summarized in Figure 4. Viral entry to host cells is mediated by viral glycoproteins found in the envelope of the virus. After the envelope fuses with host cell plasma membrane, the tegument coated nucleocapsid is released into the cytoplasm of the infected cell. After entry herpes virus capsids are found to be associated with dynein and are transported along microtubules from the cell periphery to the nuclear pore. This mechanism was observed in both HSV-1 and PRV (Dodding & Way 2011). After herpesvirus capsids dock at the nuclear pores, the viral genomic DNA is released into the nucleus of the infected cell. Viral genome replication takes place in the nucleus. Replicated copies of the genome are packaged into freshly synthetized capsids. These new viruses must first exit from the nucleus to the cytoplasm, where the tegument is formed around the new nucleocapsids. The last step in the production of mature

virions is acquiring the envelope from the host cell's membranes. Visual schematics of a virus infecting a cell is displayed in Figure 5.



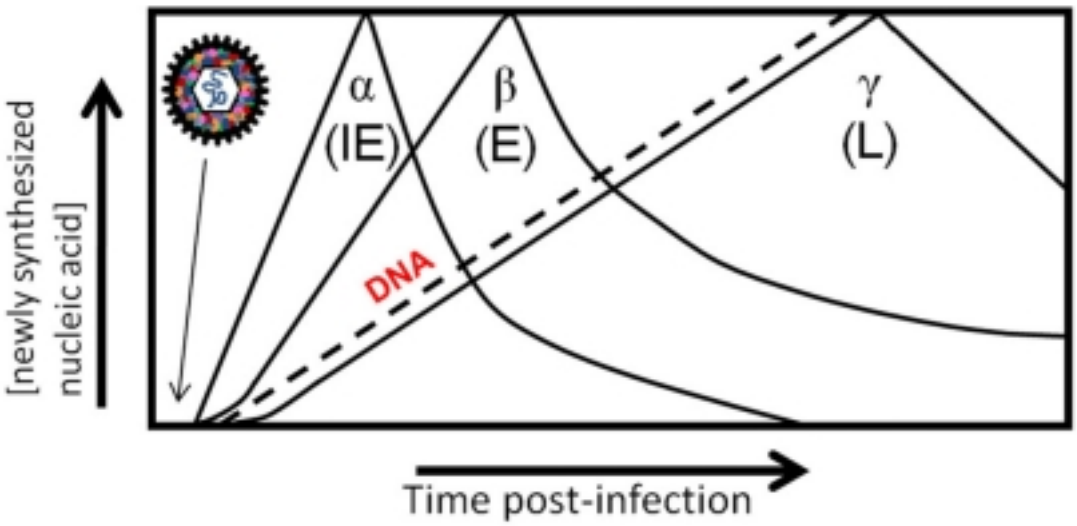
4. Figure. The scheme of the prototypic herpes virus lifecycle. A, Infections is initiated in peripheral cells with lytic infection. B, Mature virions from the primary infection site spread first to the sensory neurons through axonal transport, then to the central nervous system. C, In the central nervous system the virus enters into latency with only minimal transcription from the viral genome to maintain this phase. D, Upon certain stimuli lytic phase is reactivated, and mature infectious virions are produced, that travel back through axonal transport (near) to primary infection site and epidermal sings of the infection occur due to lytic reproduction of the virus in peripheral cells. Original figure (Frampton et al. 2005)



5. Figure. Schematics of the events in the case of a PRV virion infecting a cell. Cells are attached on the cell surface with the help of viral glycoproteins embedded in the envelope. Fusion of the envelop with cell membrane allows for the viral nucleocapsid and the tegument to enter into the cytosol. As the nucleocapsid is transported towards the cell nucleus the tegument is left behind. Upon arrival of the nucleocapsid to a nucleopor the viral genome is released into the cell nucleus and the transcription cascade is initiated. New viral DNA is produced and is packaged into forming capsids producing the nucleocapsids of new viruses. These nucleocapsids acquire tegument proteins, and envelope as they travel towards the outer regions of cytosol. Enveloped virions are either released out of the cell or transported through axons to be transmitted at neural junctions. Original figure from (Kramer & Enquist 2013).

Transcription cascade

Once the viral genome enters the host nucleus, the transcriptional cascade starts. To mature virions be produced, tight temporal regulation of gene expression is required. See Figure 6 for the schematics of gene expression timing for the classical temporal classes of herpes virus genes (IE, E, L). The promoter of the only immediate early gene in PRV (IE180) is recognized by host cell transcription machinery and thus does not require any viral protein synthesis (Cheung 1989). IE180 acts as a transactivator and is required for the expression of early genes (Ihara et al. 1983). Immediate early genes in other alpha herpes viruses also serve to modulate the host's antiviral defense mechanism or to utilize cell physiology for viral reproduction.



6. Figure. The schematics of the temporal expression pattern of classical herpes virus gene classes. α genes, also known as immediate early genes, express early on after infection. Their expression usually does not require de-novo protein synthesis. Expression of β genes (E) are dependent on the function of IE genes and typically involved in viral DNA replication. γ genes (L) are expressed several hours after infection and these genes mostly encode for structural proteins.

Early genes are mostly required for viral DNA replication while E/L and L genes are typically encode for structural proteins of the virus. It was hypothesized that the switch between the early and late gene expression is controlled by the DNA replication but it was recently show that this is not the case (Takács et al. 2013). Disruptions in the cascade have severe impact on the viral gene expression that may lead to for example reduced infectious virion production or reduced spreading ability.

The roles of the gE/gI complex

Genes *us8* and *us7* encode for glycoprotein E (gE) and glycoprotein I (gI) respectively, that play role in cell to cell spread at neural junctions. Both proteins are type I transmembrane proteins and form heterodimers thus often mentioned as gE/gI. The *us7* and *us8* genes are preserved across alpha herpes viruses and various functions have been attributed to them, for example: binding to Fc receptors at mucosal surfaces (Knapp et al. 1997) or sorting virion particles to cell junctions (David C Johnson et al. 2001), as well as promoting anterograde spreading between neurons (Husak et al. 2000). It has also been shown that both proteins are essential in axonal transport (Kratchmarov et al. 2013) of the virus. Even though DNA replication of the virus is not affected in *US7/US8* mutant viruses, their ability of cell-to-cell spreading and virion formation is impaired, thus the mutant viruses are attenuated (Snyder et

al. 2008). The Bartha strain of PRV is naturally deleted for the genomic region encoding these proteins and is therefore used as live vaccine. Recently it has been shown that infection with this strain results in greater type I interferon (TI-IFN) response, and that the absence of the gE/gI complex contributes to this increased response (Lamote et al. 2017). Since these two proteins play important role at many stages of the viral life cycle they are fairly well researched. By understanding viral spreading better, we can not only expand our knowledge of the virus, but may provide better anti-viral drugs and vaccines (Gu et al. 2015). Due to the above mentioned functions and roles we were interested in the effects of the deletion of these two genes on the viral expression pattern, an area that is not well documented in the history of the protein complex.

The roles of ICP22

The protein ICP22 is encoded by the gene *us1* with IE kinetics in HSV-1, but not in PRV where it is expressed with early kinetics. The protein is thoroughly researched in HSV-1 (but not in PRV), and several functions have been attributed to it, for example inhibiting transcription of several promoters (Guo et al. 2012; Prod'hon et al. 1996), assisting proper localization of viral proteins at nuclear egress (Prod'hon et al. 1996), modifying Pol II activity (Bastian & Rice 2009). ICP22 deleted mutant viruses are able to replicate in permissive cell lines such as HeLa and Vero cells, but efficiency of the replication is reduced (Poffenberger et al. 1993). Since it is known for ICP22 to play a central role in viral gene expression and to interact with several viral and host cell proteins, we were interested in the effects of deletion to the global viral gene expression pattern.

Research outlines

Our aim was to characterize the effects of the deletion of certain key genes in respect of the global gene expression levels and to provide a more detailed description of the viral transcriptome. Specifically we sought answers to the following questions / aims, with the third being the main focus of this thesis:

1. How does the deletion of the gene us1 affect the global gene transcription / expression levels?
2. Acquire more detailed data of the transcriptome with Next-Generation Sequencing methods.
3. What is the impact of the deletion of the gE/gI protein complex in respect of the global transcript levels?

We have analyzed the role of three key proteins namely ICP22, gI and gE, and conducted a complete transcriptome analysis to better understand the differentiated genetic regulation of viral genes. Gene deleted mutant viruses were cultured in porcine kidney epithelial cells (PK-15) and were harvested at different time points post infection. Gene activity were measure at each time point.

Methods and materials

Cell cultures

For all experiments porcine kidney epithelial cells (PK-15) were cultured under the same conditions to support PRV propagation. Dulbecco's modified Eagle's medium (DMEM) (Sigma Aldrich / Gibco / Thermo Fisher Scientific), supplemented with 5% fetal bovine serum (FBS) (Gibco / Invitrogen / Thermo Fisher Scientific) and gentamicin (Invitrogen / Thermo Fisher Scientific) at 80 µg/ml concentration was used as growth medium. Incubation occurred at 37°C in the presence of 5% CO_2.

Wild type virus stock

In all three experiments the Kaplan strain was used as wild type (wt) and as the progenitor for the mutant viruses. Stocks were generated using low-dose infection (0.1 plaque forming unit (pfu) per cell) on rapidly growing semi-confluent cells. Incubation was carried out until a complete cytopathic effect was observed.

Generation of the viruses used in the experiments

Construction of the *us1* knock-out (us1-KO) virus

First, we isolated the BamHI-10 fragment of PRV from agarose gel after digestion with restriction endonuclease BamHI. The isolated fragment was subcloned to pRL525 (Elhai & Wolk 1988), that resulted in pRL525-B10 plasmid. The pRL525-B10 served as template for amplifying the homologous sequences of the target viral region by PCR. Using the unique Ecl136II we inserted a gene expression cassette containing a green-fluorescent protein (GFP) (Clonetech). The resulting pUS1-gfp plasmid was used as the transfer plasmid for the generation of the knockout virus. Linearized transfer plasmid was co-transfected with purified wt viral DNA into PK-15 cells. The recombinant virus was generated by homologous recombination. Isolation and plaque-purification of the recombinant virus was based on fluorescence of the GFP. For the generation of rescued viruses, pUS1 was used as transfer plasmid, that was co-transfected with purified DNA of us1-KO into PK-15 cells. Revertant

viruses were selected based on the non-fluorescent plaque phenotype. Both mutant and rescued viruses were validated with DNA sequencing.

Construction of *us7/us8* knock-out (us7/us8-KO) virus

The BamHI-7 fragment of the viral genome was isolated after digestion with restriction endonuclease and subcloned to pRL525 cloning vector resulting in the pRL525-B7 plasmid. The 1,855 bp long DNA fragment designated by the StuI–AgeI sites was replaced with an EcoRI linker. EcoRI sites were introduced to the GFP expression cassette at both ends. The modified cassette was inserted at the EcoRI site to replace the StuI–AgeI fragment yielding the pUS7/US8-gfp transfer plasmid. Mature mutant viruses were generated by co-transfecting the linearized transfer plasmid with wt viral DNA into PK-15 cells. Mutant viruses were selected on the basis of the green fluorescent phenotype. Deletion of the 1,855 bp fragment from the PRV genome resulted in the inactivation of both us7 and us8 genes of the virus.

Production of the virus stock for sequencing

Initial steps for the virus stock for this experiment was generated by the method described earlier. After showing full cytopathic symptoms, cells were frozen and thawed three times and subsequent low speed centrifugation (10,000 g) for 20 min was applied. After removing cell debris, the supernatant was concentrated and further purified by ultracentrifugation through a 30 % sugar cushion at 24,000 rpm for 1 h, using a Sorvall AH-628 rotor.

Infections and measurements

Assessment of the effects of US1 deletion

In order to measure the influence of the us1 gene deletion on global transcription kinetics of PRV, we infected with the wt or us1-KO virus at a high multiplicity of infection (MOI; 10 plaque forming units (pfu)/cell). Cells were incubated for 1 h with the virus suspension, after which the suspension was removed and the cells were washed with phosphate-buffered saline (PBS) solution. Subsequently, fresh culture medium was added to the cells, and cultivation continued for an additional 0.5, 1, 4, 6, 8, 12, 18 or 24 h. At each time point, cells were washed with PBS, harvested and were frozen for later RNA isolation.

Assessment of the impact of *us7/us8* deletion

Rapidly growing, semi-confluent PK-15 cells were infected with three different doses of the mutant virus: low (0.1 pfu/cell), medium (1 pfu/cell), and high (10 pfu/cell) multiplicity of infection. The same protocol was used for initial virus suspension incubation, washing and post infection incubation as in the *us1* experiment. PRV-infected cells were washed in PBS and harvested similarly for RNA purification as in the *us1* experiment at 1, 2, 4, 6, 8, 12, 18 and 24 h post infection.

Cell cultures for transcriptome sequencing

In this experiment high multiplicity of infection (10 pfu/cell) was used for the infection of PK-15 cells. The same protocol was used as in the previously described cases. Time points for cell harvest were 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 and 24 h p.i.. Mock-infected cells that were treated in the same way as infected cells, were used as controls.

Isolation of RNAs

Total RNA was isolated with the use of NucleoSpin RNA II Kit (Macherey-Nagel) as per manufacturer's instructions. Previously harvested cells were centrifuged at low speed then and lysed in the supplied lysis buffer. rDNase (supplied as RNAase-free) treatment was applied to the samples to remove potential genomic DNA contamination. To remove accidental residual DNA contamination we utilized Turbo DNase (Ambion Inc.) treatment. Subsequently, RNA samples were eluted in RNase-free water (also supplied with the kit), resulting in a total volume of 60 µl of RNA solution. RNA concentrations were determined spectrophotometrically in a BioPhotometer Plus (Eppendorf) for the US1 experminent and by Qubit 2.0 Fluorometer (by using the Qubit RNA BR Assay Kit (Life Technologies/Thermo Fisher Scientific) for the *us7/us8* experiments. The RNA solutions were stored at −80°C until use.

Reverse transcription coupled real-time PCR (RT/RT PCR or RT^2 PCR) TBC

Reverse-transcription was carried with gene specific primers and SuperScript III enzyme (Invitrogen) on total RNA samples Reverse-transcription reaction mixtures (total RNA, primer, SuperScript III enzyme, buffer, dNTP mix and RNase inhibitor (RNAsin, Promega)) were incubated at 55°C for 1 h. The cDNA synthesis of the first-strand was terminated by

putting the samples to 70°C for 15 min. The cDNAs were diluted 10-fold with nuclease-free water (Promega Corp.) and the solutions were stored at −80°C until use. Real-time PCR reaction were carried out using Absolute QPCR SYBR Green Mix (Thermo Scientific), in Rotor-Gene 6000 machines (Corbett Life Sciences). Empty (no reverse-transcription sample), no-primer and no-template along with loading control of 28S rRNA of swine were used to ensure accuracy.

Preparation of the cDNA library

We used the Illumina compatible ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicenter) to create a strand-specific (random hexamer-primed) total RNA library for 100 bp paired-end sequencing. For polyA-sequencing (PA-seq), a single-end library was constructed using custom anchored adaptor-primer oligonucleotides with an oligoT_{10}(VN) primer sequence (anchored primers containing standard Illumina strand-specific adaptor sequences).

Illumina sequencing

Transcriptome sequencing was carried out using an Illumina HiScanSQ platform at the Genomic Medicine and Bioinformatic Core Facilty of the University of Debrecen. Quality of raw read data was assessed with FastQC version 0.10.1. Reads alignment was done with Tophat version 2.09 (Trapnell et al. 2009),first to the respective host genome (*Sus scrofa*, assembly: Sscrofa10.2), then to the PRV genome (KJ717942.1). Ambiguous reads were discarded. As for PA-Seq, we used Bowtie version 2 to carry out the mapping (Langmead & Salzberg 2012), followed by peak detection using HOMER in strand-specific mode, with adjustments for the peak qualities of oligo(dT) primed libraries and a cutoff of 50 reads per base position. We used in-house scripts to assign peak categories. The following criteria was used: the presence or absence of a poly-adenylation signal (polyA signal (PAS)) in the 50 bp region upstream from the poly-adenylation (PA) site and the presence of at least 2 consecutive adenine mismatches in at least 10 independent reads at the PA site. Annotation and visualization were carried out in the Artemis Genome Browser v15.0.0 (Rutherford et al. 2000) and IGV v2.2 (Robinson et al. 2011).We examined the GC bias in the alignments using the Bioconductor R package. The prediction of canonical and non-canonical PAS was carried out using PolyApred (Ahmed et al. 2009).

Data analysis

Calculation of relative expression ratios (R values)

For calculating the relative expression ratios (R values) of genes we utilized the following formula (Formula 1.):

$$R = \frac{\overline{(E_{sample6h})^{Ct_{sample6h}}}}{(E_{sample})^{Ct_{sample}}} : \frac{\overline{(E_{ref6h})^{Ct_{ref6h}}}}{(E_{ref})^{Ct_{ref}}}$$

Formula 1. Legend: E – efficiency, calculated by the software. Ct – cycle threshold, the cycle number in which the amplification reaction enters the exponential phase. In the counter of the quotients are the averages of the values measured at 6h in wt viruses for the samples and reference respectively. The denominators represent the values from each of the real-time PCR reactions.

The cDNAs were normalized to the cDNAs of the 28S rRNAs of the host cells by using the Comparative Quantitation module of the Rotor-Gene 6000 software Version 1.7.28 / Rotor-Gene Q software v.2.3.1.49 (Corbett Research / Qiagen respectively) for ICP22 and gE/gI experiments respectively. The software automatically sets the cycling thresholds and calculates the efficiency of PCR reactions sample-by-sample. Each gene's expression is represented relative to the average of the 6h levels in wild type virus (the counter of the first quotient) and normalization is achieved by division with the second quotient, that is the relative expression of the reference (in our case 28S rRNA of the host cells) in that same sample.

Effects of the deletion of genes in question on the global gene expression were calculated using the R_r value, that is the ratio of the R values of the mutant virus and the wt virus, explained in Formula 2.

$$R_r = \frac{R_{mutant}}{R_{wt}}$$

Formula 2. To assess the effects of mutations on the global expression, R values for each gene in the mutant virus at a given time point is divided by the R value of the same gene at the same time point from the wild-type virus

An R_r value greater than one indicates an inhibitory effect, while R_r less than one represent excitatory effect of the deleted protein(s) on the expression of a given gene in the wt PRV. The more the deviation is from the value of 1 the higher the influence is. Furthermore we used the mean expression value (E^{Ct}) of all examined mRNAs (total) in a given sample as normalization

factor for the transcripts (sample) real-time quantitative PCR data as described earlier (Mestdagh et al. 2009) to obtain R_x values and their derivatives. The following formulas (Formula 3, 4 and 5) describe the method used to obtain these values.

$$R_x = \frac{{E_{sample}}^{Ct_{sample}}}{{E_{other}}^{Ct_{other}}}$$

Formula 3. Rx values are meant for normalization of the real-time PCR data for samples. The E^{ct} values represent the efficiency of the transcription for a gene in a RT-PCR experiment. The value E is the computed efficiency by the RT-PCR machine's software, and Ct stands for Cycle-threshold – the cycle in which the fluorescence level in the PCR reaction exceeds the background fluorescence in the early exponential phase.

$$R_{x'} = \frac{R_x sample}{\overline{R_x all}}$$

Formula 4. Rx' values are derived from the Rx values by dividing a samples Rx value with the average of all the Rx values. This yields the relative contribution of a gene's transcripts to the total transcriptome.

$$R_{x'}r = \frac{R_x \Delta us7/us8}{\overline{R_{x'}wt}}$$

Formula 5. This formula represents ratio of the relative contribution of a gene between the mutant and wild type genetic background.

To compare the gene expression patterns in the wild-type and mutant virus we calculated the Pearson's correlation of the two using the following formula (Formula 6.):

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_x S_y}$$

Formula 6. The formula used to calculate the Pearson's correlation coefficient. The variables X and Y are the R_r values of a gene in the mutant and wild type genetic background. $\bar{X}$ and $\bar{Y}$ represent the averages of the R_r values of the corresponding genetic background, n is the number of samples. Finally, S_x and S_z are the standard error of the mean for variable X and Y respectively.

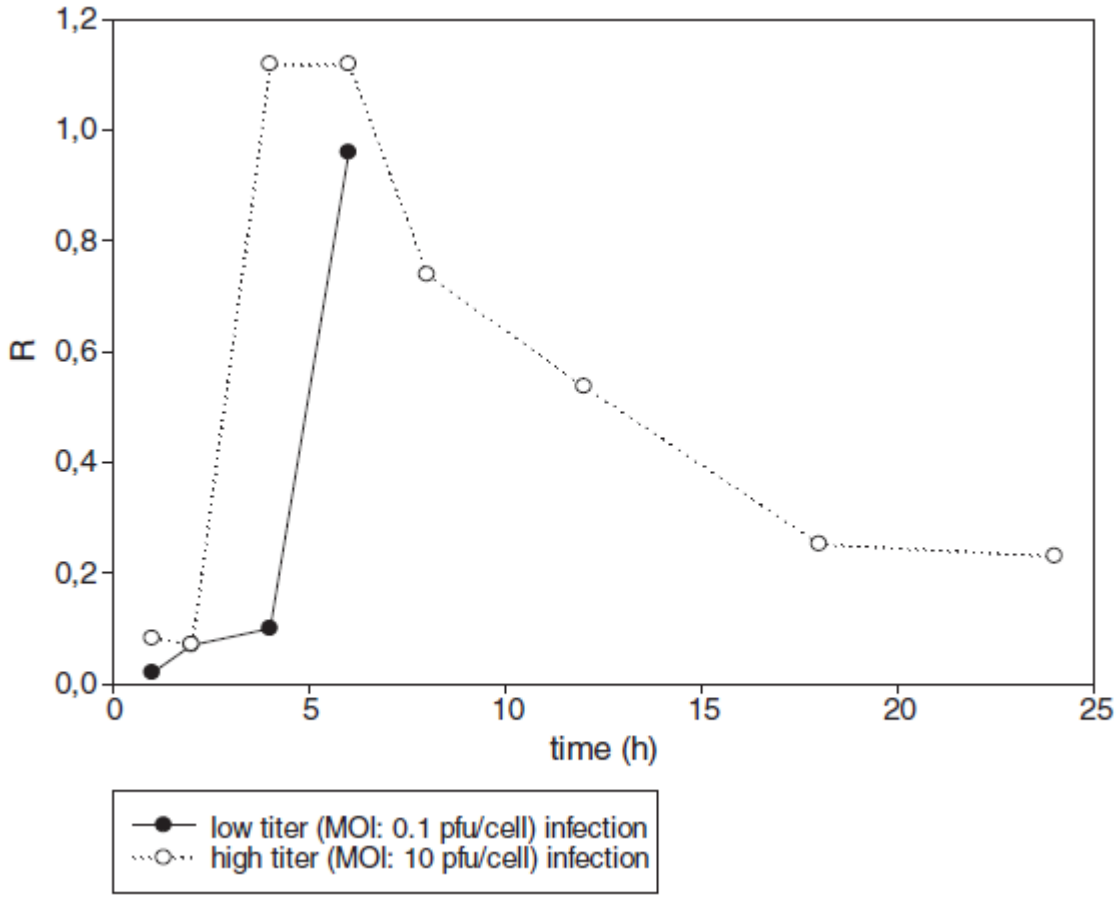
Pearson's correlation coefficient measures linear relationship between two variables. Its value is between +1 and -1. The positive values mean that the two variables have similar pattern, while negative values imply opposite behavior. The absolute value of the correlation coefficient determines the strength of the relationship.

Results

Results of the us1 deletion experiment

Transcription kinetics of the us1 in wild type virus

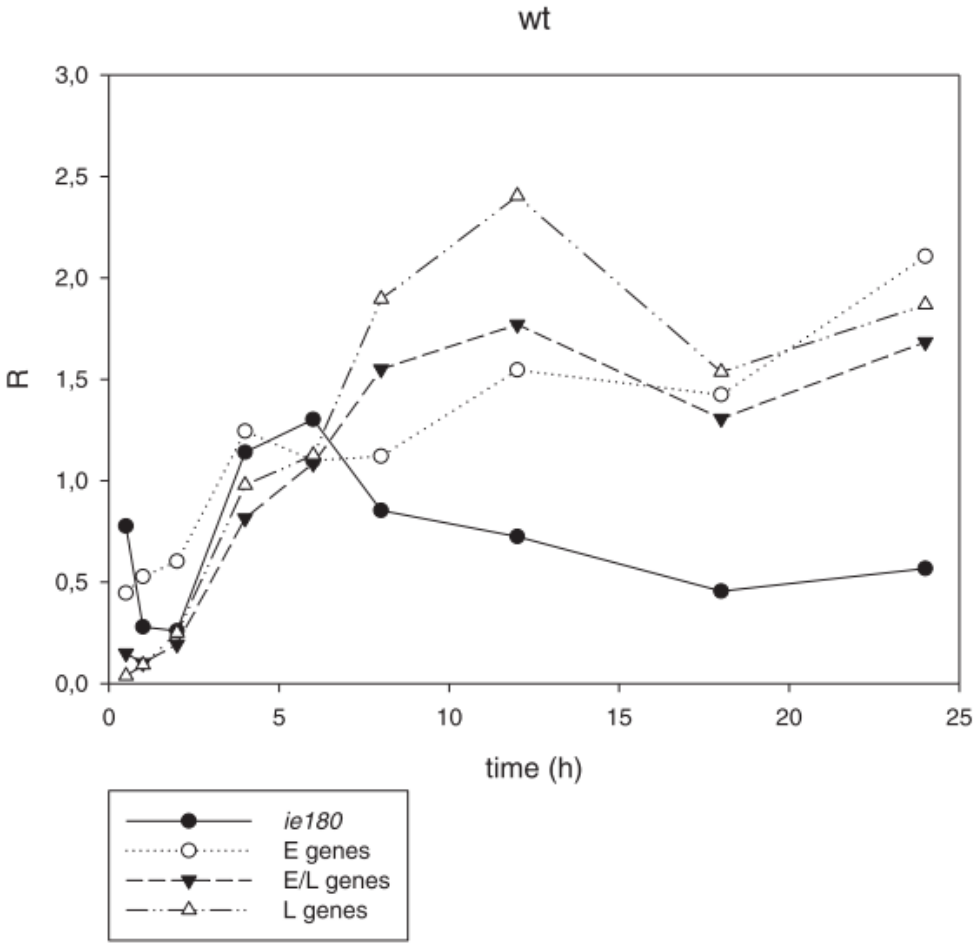
Infection of PK-15 cells with wt virus using low (MOI = 0.1 pfu/cell), and high titer (MOI = 10 pfu/cell), revealed that us1 expression is titer dependent. In case of low titer its expression resembles that of late genes, with expression steeply increasing after 4h. Contrary to this in the event of high titer infection, the same rapid increase in expression is observable after 2h post infection. This is summarized in Figure 7.



7. Figure. The expression kinetics of us1 gene in the wild type PRV. In low titer infections the expression shows that us1 gene belong to the late kinetic class, however high titer infection data sorts the gene to the early genes. We have not observed this level of titer dependence with other genes. Low titer infection were not evaluated after 6h due to secondary infections caused by virions released from cells from the primer infection

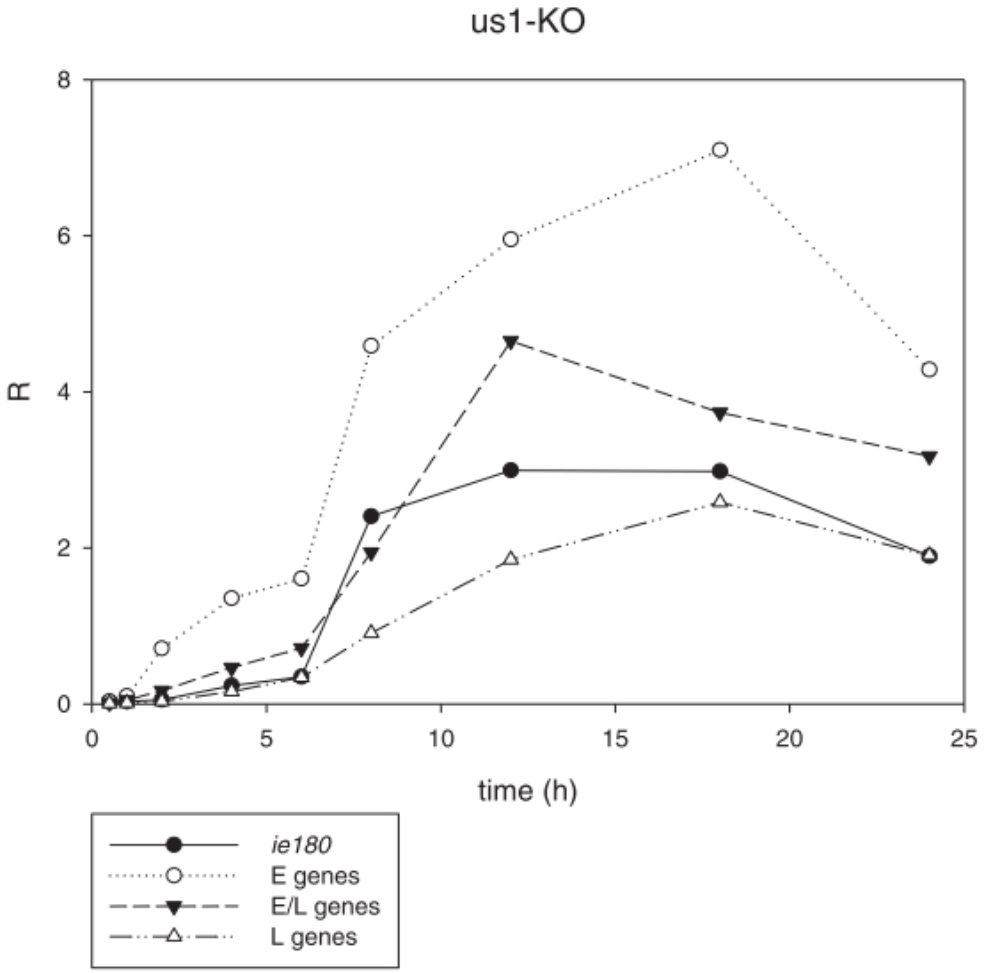
The effects of the *us1* gene deletion

Relative expression ratios (R values) were calculated for the mutant (us1-deleted) and wild type virus, and relative gene expression was assessed compared to the wild type using the R_r values. The heat-map visualization of the results are shown in the Figure 10, and the average expression levels for the different kinetic classes are shown in Figure 8 and Figure 9 for the wild-type and the us1-KO virus respectively. Our data shows that in the first hour of infection, most of the examined genes' expression (with few exceptions like *ul53* or *ul24*) decreased drastically compared to wt. This decrease is most significant in the E kinetic class of genes. However by 6h post infection the average expression level for this class approximates that of wild type virus. Late genes on the other hand were expressed at lower levels even at 8h post infection in average. Early/late genes were expressed between the two previous classes' levels,



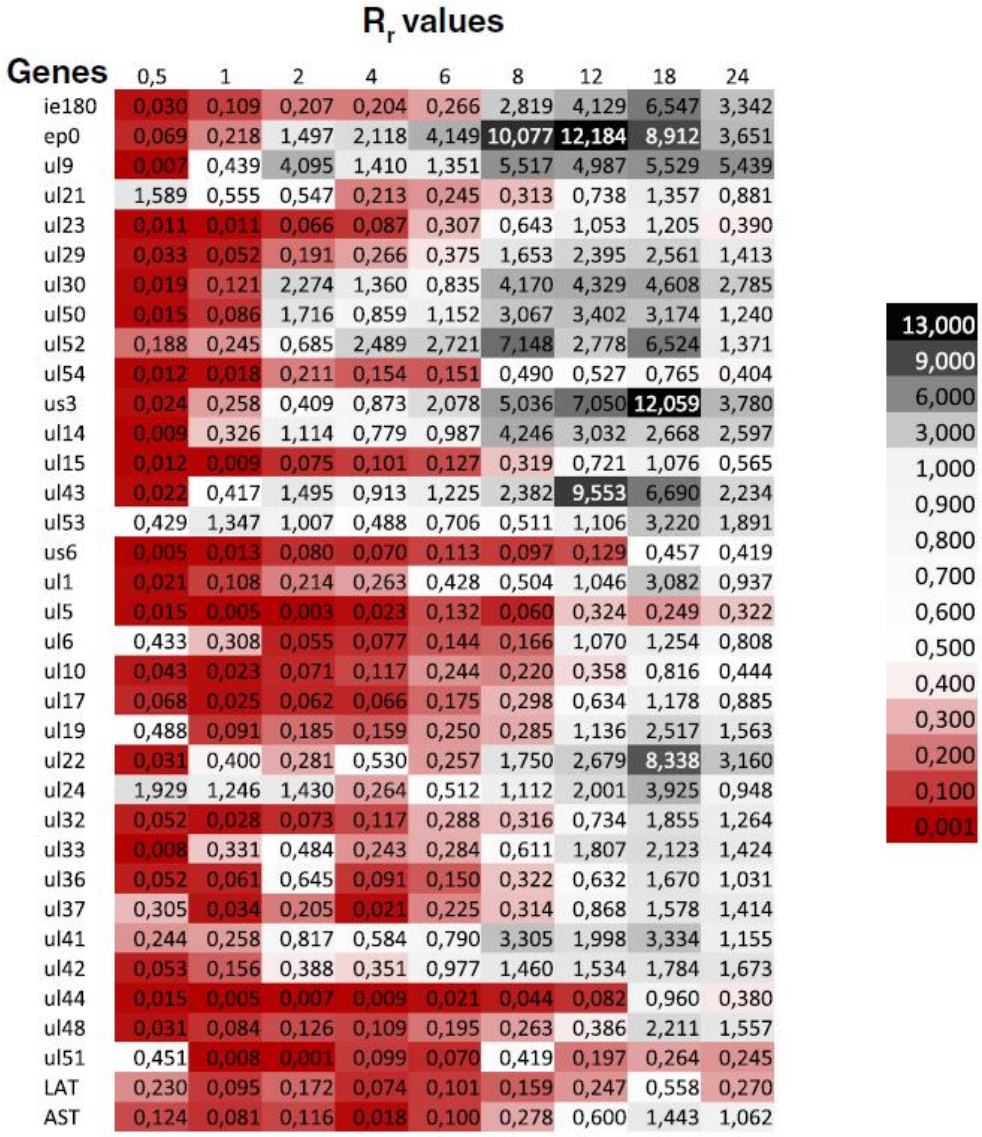
8. Figure. The average expression levels of the different kinetic classes of genes of the PRV on wild type genetic background. Early genes start at a relatively high expression, but E/L and L genes show much steeper increase in their expression as viral infection progresses.

and the only IE gene (*ie180*) was expressed at lower level in the first 6h of the infection in the mutant virus.



9. Figure. The average expression levels for the different kinetic classes in the us1 mutant virus. E/L and L genes suffered severe downregulation, but not early genes, whose average expression is even multiple times greater than in the wild-type

The average rate of transcription of the IE, E and E/L genes from the mutant viral genome exceeded the rate of transcription from the wt genome after 6 h pi. This was in contrast with the transcript levels of the average L genes of the mutant virus, which remained below or equal to that of the wt virus, except at 18 h pi. The expressions of the *us3*, *ul22*, *ul43*, and *ul52* genes were significantly elevated at certain time points in the late stage of infection. The *ul5*, *ul51* and LAT genes were the only examined PRV genes whose expressions were always lower in the us1- KO virus than in the wt virus. Overall, the above data suggest that the ICP22 protein exerts a selective effect on the expressions of PRV genes belonging in different kinetic classes.



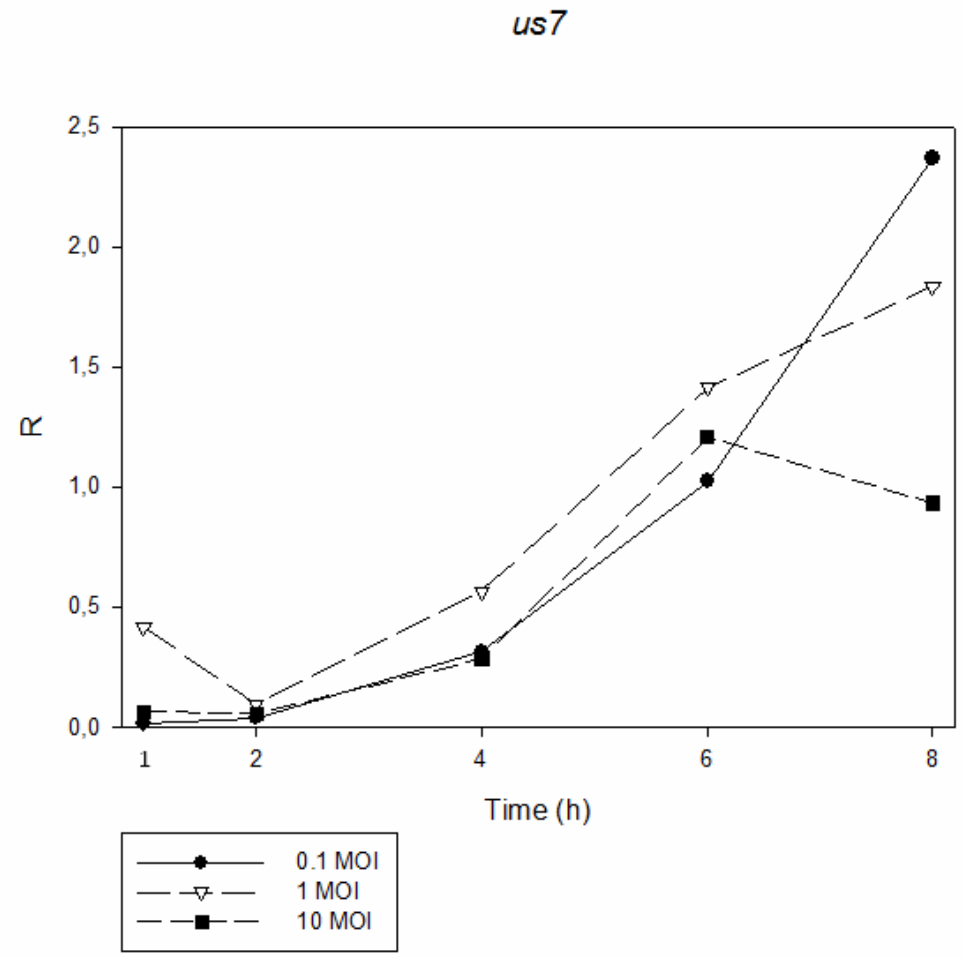
10. Figure. The heatmap visualization of the R_r values for the selected genes with exact values. Values less than one denotes that in the mutant virus the given gene was underexpressed at that given time point, while values greater than one mean that the given gene is overexpressed compared to the wild-type.

Results of the us/us8 experiment

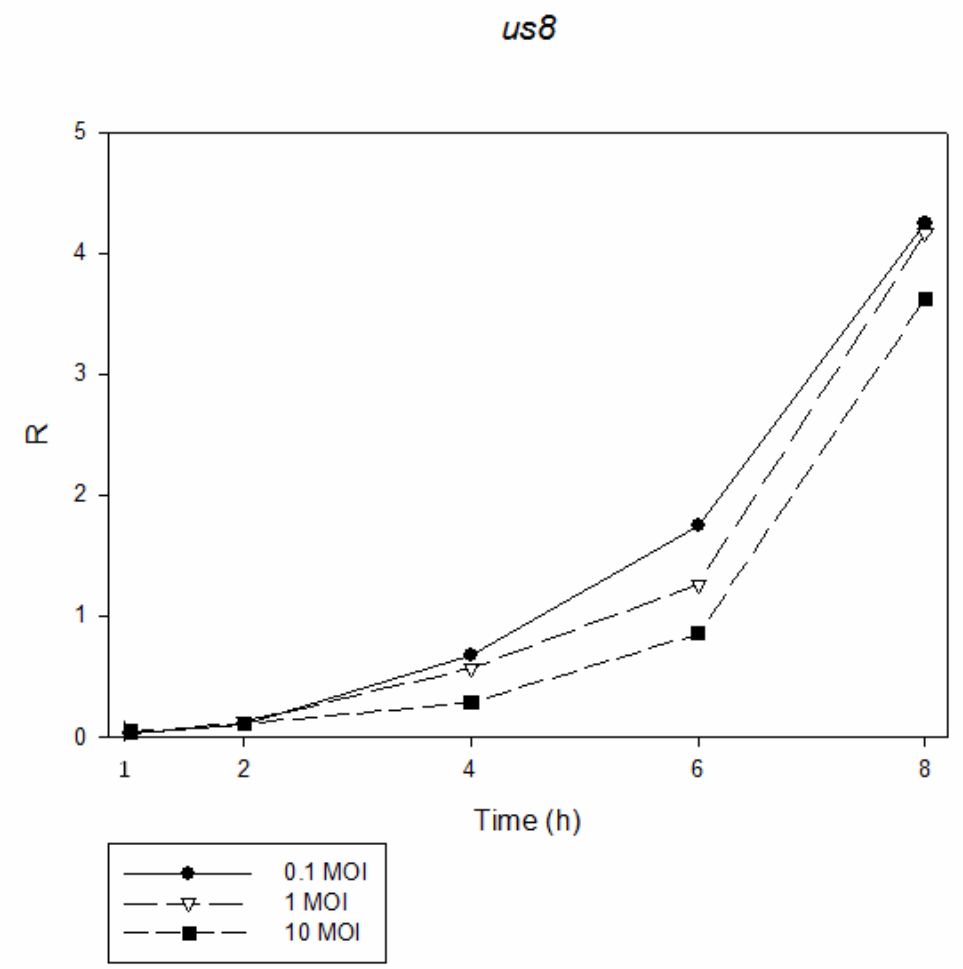
Transcriptional kinetics of *us7* and *us8* genes in wild-type PRV

As an initial step in our study we measured the transcription kinetics of *us7* and *us8* genes in the wt virus under different conditions: high (MOI=10), medium (MOI=1) and low (MOI=0.1) multiplicity of infection. Results of these measurements are summarized in Figure 4 and Figure 5. Since the products of these genes form heterodimers, their expression kinetics are highly similar – both display E/L kinetics. Notable expression starts between 2h and 4h post

infection, with expression peaking at 8h p.i.. In wild type virus as secondary infections occur by virions released from initially infected cells after 6-8h at low titer infections thus later hours were not taken into consideration for the sake of comparability between different titer levels. Figure 11 and Figure 12 depicts the expression levels of us7 and us8 genes respectively for the wild type virus at different titer levels.



11. Figure. The expression kinetics of the us7 gene under different infection conditions in wild type virus. In agreement with it's functions the us7 gene shows E/L kinetics with minimal titer dependence in its expression.



12. Figure. The expression kinetics of the us8 gene under different infection conditions in wild type virus. Similarly to its functional partner us7, the us8 gene shows E/L kinetics with minimal titer dependence in its expression.

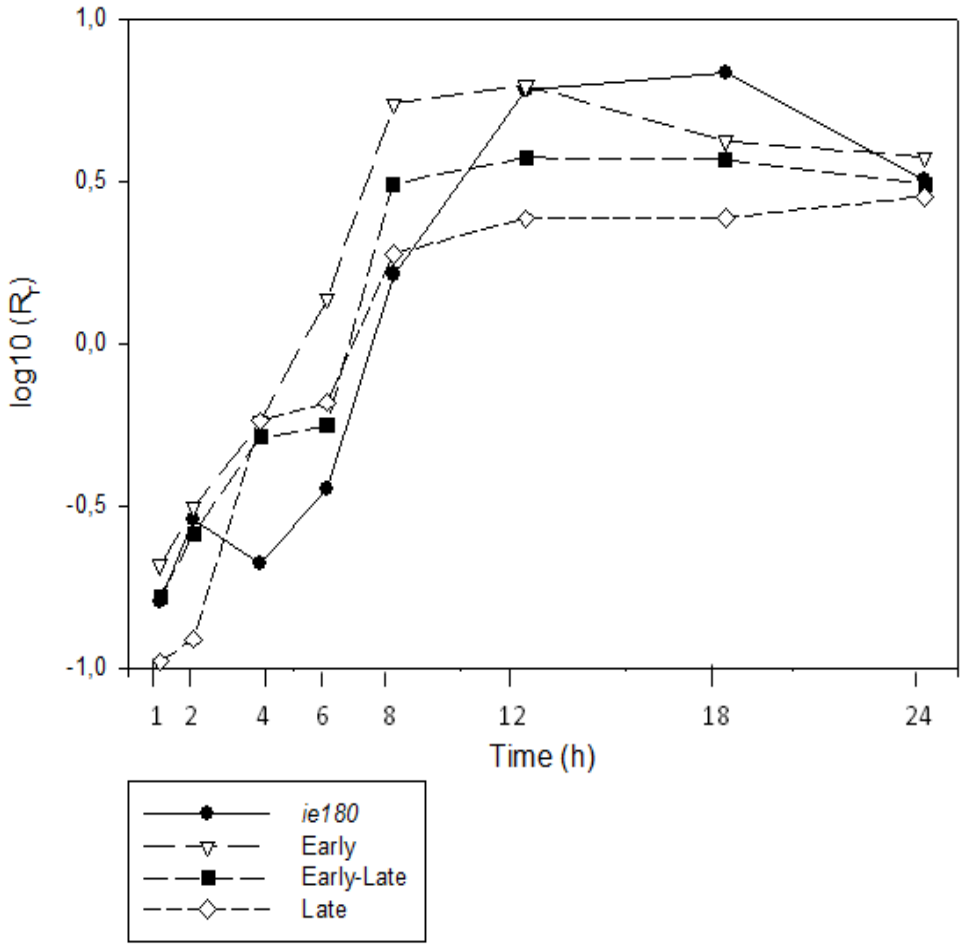
The effect of the *us7/us8* deletion on the expression of PRV genes

Next, we made a comparison between the expression levels of 37 PRV genes in wt and mutant (*us7/us8* deleted) genetic backgrounds over a 24 h period of viral infection. Results of this experiment are summarized in Table 1. High titer (MOI=10) was used for both viruses to infect host PK-15 cells. Results indicate relatively higher level of gene expression in the first 6 h for all of the three kinetic classes in the wt virus genes (with few exceptions). On the other hand, the situation is reversed in the late phase of infection: with some exceptions, the genes of the mutant virus were expressed in relatively higher levels. The genes *us6* and *us9* are special cases since they are adjacent to *us7* and *us8*, thus their expressions are likely to be affected by the genetic modification itself. This may account for the irregular behavior. These two genes were excluded from further analysis.

1. Table. R_r values of the examined PRV genes in the *us7/us8* experiment. With only few exceptions, nearly all genes are under-expressed in the mutant virus compared to the wt in the first 6 hours of infection. However this trend changes in the later stages, as after 8h almost all genes show higher expression in the mutant than in the wild type.

Kinetics	Gene	EARLY/EARLY-LATE PHASE OF INFECTION				LATE PHASE OF INFECTION			
		1h	2h	4h	6h	8h	12h	18h	24h
IE	*ie180*	0,16	0,29	0,21	0,35	1,63	6,02	6,82	3,19
E	*ul54*	0,03	0,15	0,43	0,42	4,84	3,19	4,85	5,02
E	*ul52*	0,12	0,25	1,21	1,15	12,13	6,01	2,54	1,49
E	*ul50*	0,09	0,55	0,51	0,74	4,26	3,18	6,01	4,78
E	*ul29*	0,25	0,25	0,31	0,95	10,18	10,83	5,85	4,21
E	*ul30*	0,11	0,12	0,48	0,69	4,94	5,27	2,83	4,44
E	*ul36.5*	0,12	2,38	0,34	0,8	4,03	2,13	0,63	3,23
E	*ul23*	0,05	0,24	0,25	0,54	2,54	7,51	3,12	3,94
E	*ul21*	0,4	0,35	0,39	0,51	1,31	2,78	2,09	2,68
E	*ul9*	0,09	0,23	0,31	0,74	2,4	2,89	1	1,88
E	*ep0*	0,33	0,52	1,05	5,58	10,2	23,66	23,11	18,51
E	*us3*	0,08	0,05	0,1	0,49	1,8	3,29	2,85	3,43
E/L	*ul43*	0,22	0,36	0,57	0,7	2,64	3,85	6,64	2,26
E/L	*ul20*	0,25	0,13	0,85	0,37	3,16	2,63	4,77	5,05
E/L	*ul15*	0,08	0,3	0,18	0,35	1,82	3,65	4,37	4,94
E/L	*ul14*	0,12	0,24	0,4	0,82	4,86	4,56	2,02	2,52
L	*ul51*	0,01	0,19	2,79	1,68	5,64	2,64	3,89	3,94
L	*ul49.5*	0,11	0,3	0,28	0,37	1,31	1,55	3,2	4,4
L	*ul48*	0,1	0,13	0,27	0,32	1,52	1,71	2,08	1,72
L	*ul32*	0,19	0,14	0,25	0,49	1	1,73	1,61	3
L	*ul33*	0,22	0,33	0,62	0,84	3,07	4,01	3,99	2,7
L	*ul37*	0,11	0,24	0,29	0,57	1,46	2,13	1,68	1,33
L	*ul38*	0,01	0,02	0,17	0,28	0,58	1,8	1,21	2,03
L	*ul41*	0,22	0,2	2,06	2,73	9,99	7,04	5,44	1,58
L	*ul42*	0,1	0,14	0,15	0,33	1,13	1,7	1,53	3,47
L	*ul44*	0,11	0,19	0,11	0,17	0,38	0,36	1,16	0,81
L	*ul24*	0,66	1,1	0,44	0,78	2,7	3,13	2,97	3,29
L	*ul22*	0,24	0,25	1,18	0,6	3,39	3,42	4,39	3,14
L	*ul19*			0,31	0,44	0,71	2,21	1,04	0,76
L	*ul17*	0,05	0,11	0,18	0,44	1,35	2,29	1,81	1,45
L	*ul10*	0,09	0,09	0,53	0,35	0,76	1,61	1,93	2,14
L	*ul6*	0,11	0,04	0,35	0,64	1,55	3,57	3,1	7,53
L	*ul5*	0,01	0,01	0,1	0,72	0,81	4,73	0,99	6,76
L	*ul1*	0,2	0,42	0,48	0,55	1,26	2,21	2,22	3,4
L	*us1*	0,03	0,22	0,12	0,25	1,89	4,56	4,41	8,76
E/L	*us6*	0,04	0,1	0,05	0,08	0,09	0,17	0,21	0,4
E/L	*us9*	0,07	0,07	0,19	0,22	0,67	1,34	2,09	1,93

30
20
15
12
8
5
3
2
1,5
1
0,8
0,6
0,4
0,2
0



13. Figure. The $\log_{10}$ values for the average expression levels of the different kinetic classes of PRV genes in the *us7/us8* deleted virus. Expression of late gene is barely detectable at all at the initial phase if the infection. The other kinetic classes suffer less from the mutation. All off the kinetic classes are expressed at higher level in the later phase of the infection with IE and E genes benefiting the most, while late gene exhibit lesser increase in their expression.

Comparison of the averages of the expression values (Rr) for each of the kinetic classes showed, that in early hours of infection genes with L kinetics were the most affected, showing the lowest average transcript level, while at the later hours, the mutation influences the genes belonging to the IE and E kinetic classes with these kinetic classes displaying the highest average transcript levels (see Figure 13).

The impact of the deletion was the highest for the *ep0* gene that showed 23.66 fold increase in its expression in the mutant virus compared to wt. To assess the effect of the *us7/us8* deletion on expression dynamics we calculated the Pearson's correlation coefficient between the mutant and wt virus for the E, E/L and L kinetic classes. The average of R values for each kinetic class were used as input. The results are shown in Table 2. The coefficient values shows

that genes of the late class show similar expression kinetics in the mutant as in the wild type virus. As for E and E/L genes, their expression kinetics deviated to a moderate level in the mutant virus compared to the wt virus.

2. Table. The Pearson's correlation coefficient values for the different kinetic classes of the PRV. The deletion affected early and earl/late genes to a moderate level, while late genes were affected to little extent.

	Pearson's coefficient
$\overline{R}_{\Delta gEgI\ E\ genes}$ vs. $\overline{R}_{wt\ E\ genes}$	0.804
$\overline{R}_{\Delta gEgI\ E/L\ genes}$ vs. $\overline{R}_{wt\ E/L\ genes}$	0.825
$\overline{R}_{\Delta gEgI\ L\ genes}$ vs. $\overline{R}_{wt\ L\ genes}$	0.915

Analysis of the relative contribution of RNA molecules to the global transcripts

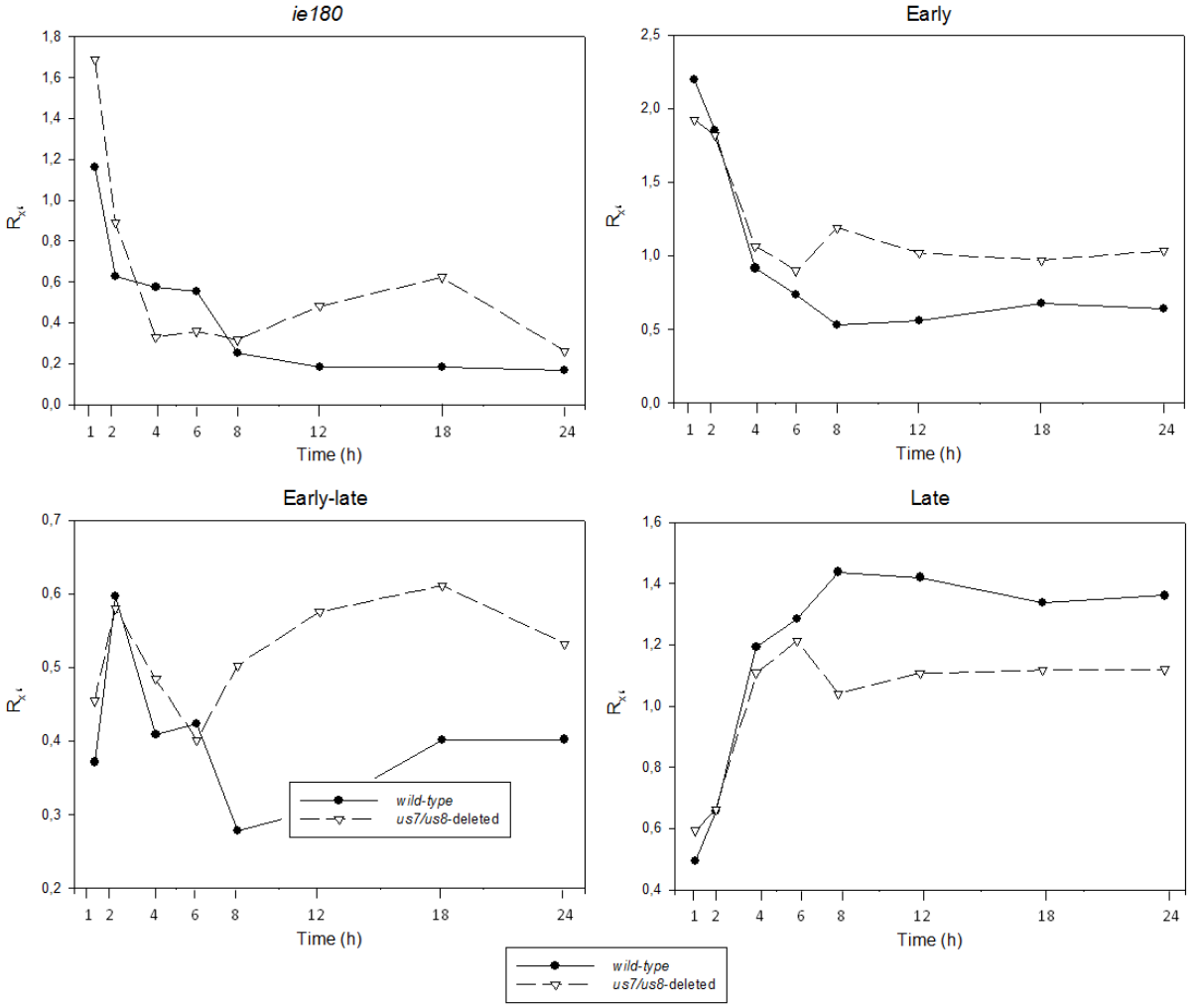
We calculated $R_{x'}$ values to see the changes of relative contribution of each transcript or a group of transcripts to the whole viral transcriptome. We found that L gene products' relative contribution is much less to the whole transcriptome at late phase of the infection in the mutant virus in comparison to wt, despite they are expressed at higher levels in this phase in the mutant virus than in the wild type. Opposing to this the other kinetic classes contribute higher proportion to the global transcriptome of the mutant PRV in the later phase of infection. Figure 14 demonstrates the relative contribution of genes belonging to different kinetic classes to the whole transcriptome.

We have also calculated the $R_{x'}$r values and found that the gene ie180 is affected oppositely to other PRV genes. While the sole IE gene is relatively abundant in the first two hours and at the late phase of infection, other kinetic classes are produced at relatively higher levels in the mid-phase (see Figure 15). The effect of the deletion is the highest on the relative amount of ie180 gene moderate for E genes, low for E/L genes, and minimal for the L kinetic class.

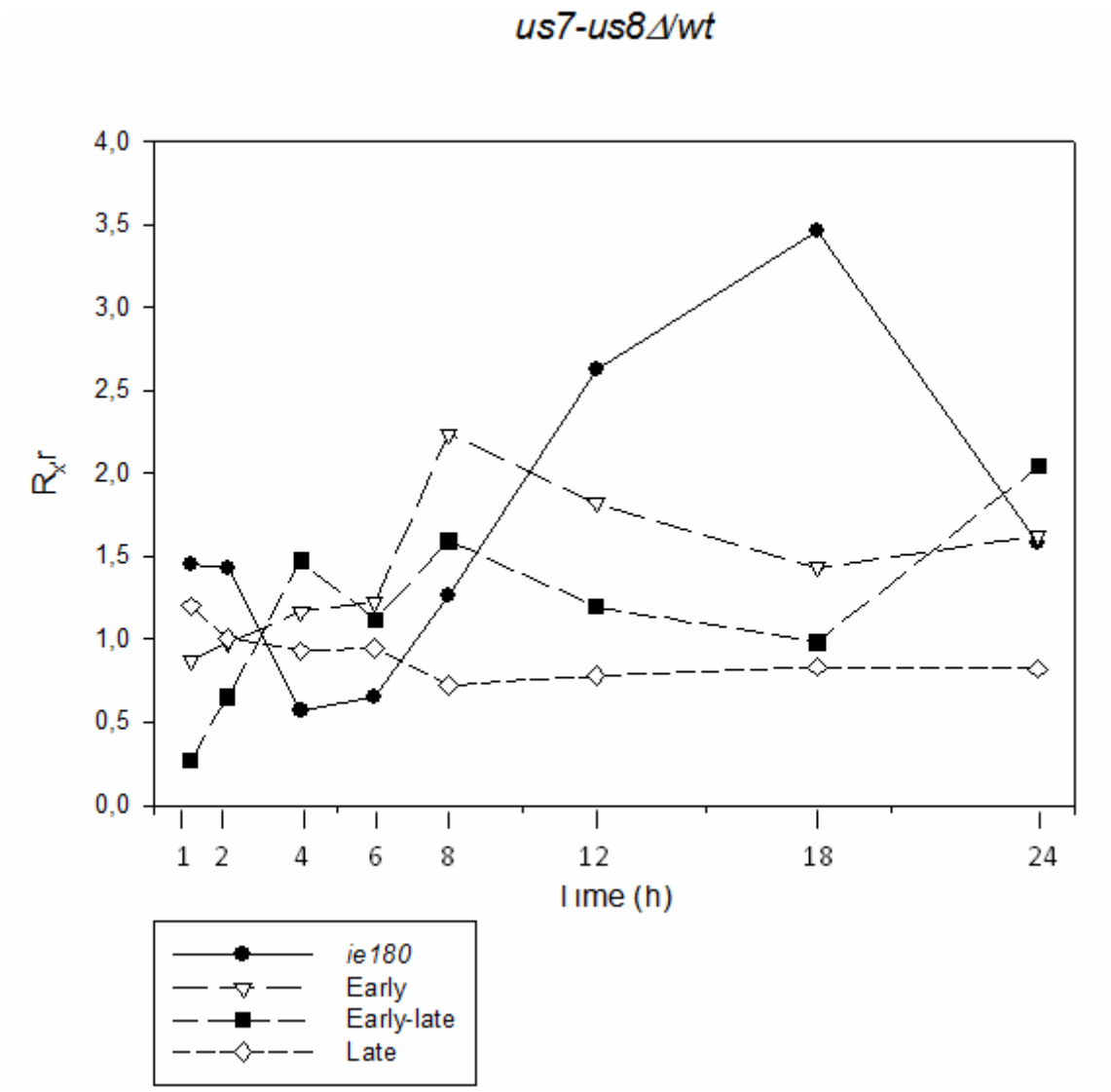
The expression of the *ie180* gene is correlated with the expressions of PRV genes in the *us7*/*us8* mutant

Our institute has showed in one of our previous work that the expression the *ie180* gene correlated with the expression of other genes in the wild type virus weakly while strong

correlation was observable for the mutant virus used in that work. (Pearson's correlation coefficient, r= 0.330, and *r* = 0.893 for the wild type and mutant respectively) (Tombácz et al. 2011). Similar results were found in the case of the us7/us8 deleted virus. In the mutant virus, correlation of the R values for the *ie180* and other genes show strong correlation (r=0,904), while in the wt virus the same comparison shows weak correlation as we have seen before (r=0.343).



14. Figure. The average relative contribution of each of the kinetic classes for both wild type and mutant (*us7/us8*) genetic backgrounds. The IE class is only represented by the gene *ie180* since this is the only gene in PRV that belongs to this class. For the wild type virus the relative contribution of each kinetic class follows the expected behavior, eg. early genes contributing to high levels during the early phase, but later the late genes take over. However for the mutant virus this is not the case



15. Figure. The (average) relative amounts of genes of different kinetic classes. While the only immediate early gene *ie180* is abundant at early and late stages, the rest of the kinetic classes are more abundant in the early-mid phase of the infection.

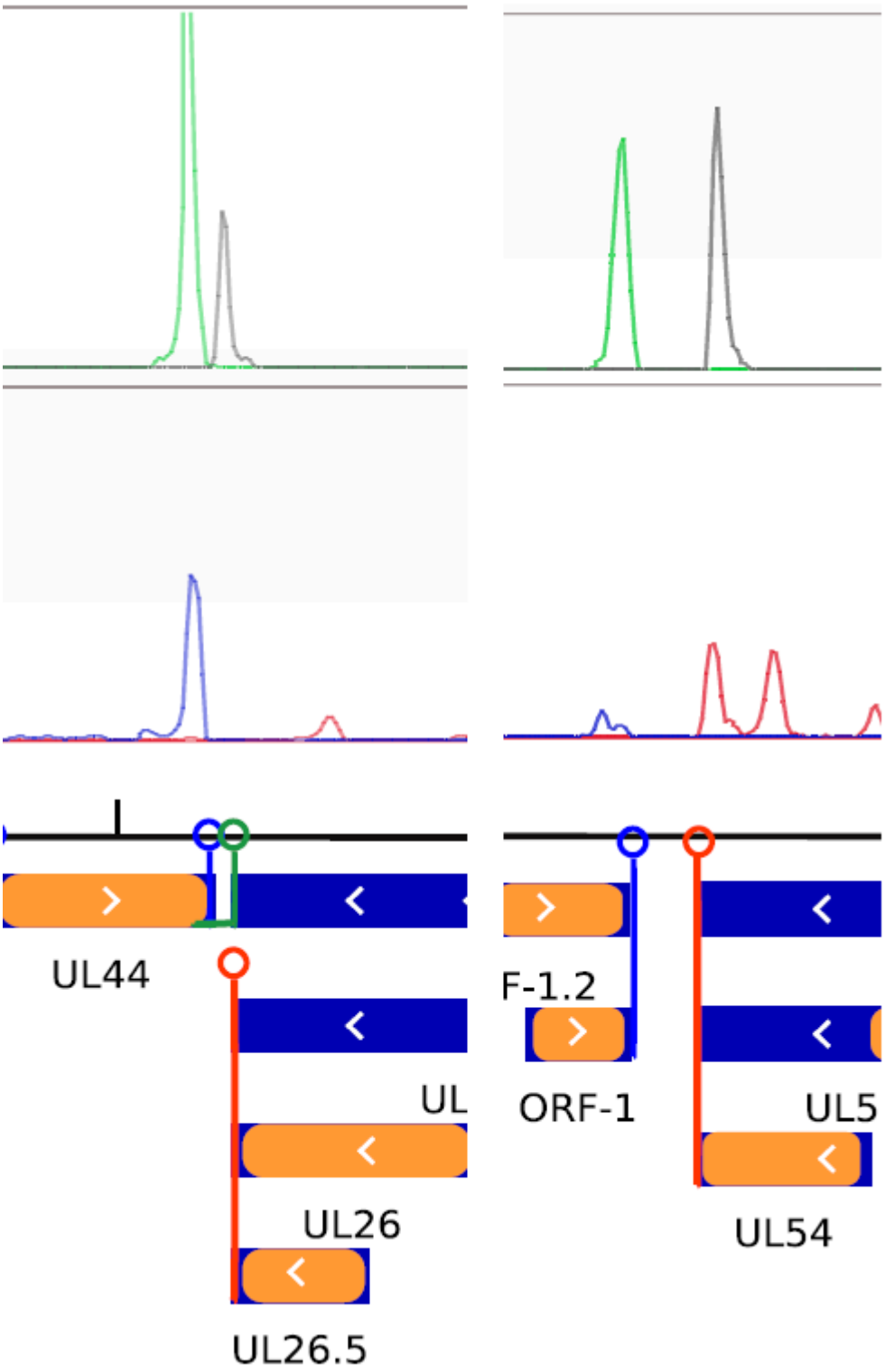
Summary of the transcriptome profiling

Detailed results of this experiment can be found in (Oláh et al. 2015), here we only show a brief summary of the transcriptome profiling.

Total RNA sequencing resulted in a data set of ~ 208 million 100 bp paired-end reads for the random hexamer-primed library, of which 1.3 million reads (achieving nearly 1000 times average coverage, although with highly unequal distribution of reads) aligned to the viral genome version KJ717942.1.The majority of the remaining sequences aligned to the host

organism genome *Sus scrofa* 10.2. We had ~103 million single- end, 50 bp reads, with 10 million reads aligning to the above-mentioned viral reference from the PA-seq.

Apart from the terminal and internal repeat sequences, that are highly repetitive, nearly all of the viral genome was transcribed. Intergenic repeat regions also did not show detectable transcript levels, with the exceptions described below. These were earlier predicted to be transcriptional insulators (Klupp et al. 2004). The exceptional insulator sequences that exhibited observable transcription are the two convergently oriented gene pairs, *ul44-ul26* and, to a lesser extent, *ul35-ul36* that showed leaky transcription into the intergenic repeat regions of 109 bp and 443 bp, respectively. In other cases (eg. between *ORF-1* and *ul54*; *ul46* and *ul27*) the intergenic repeat regions acted as transcription insulators and in these boundary regions, no expression was observable. Examples of a leaky transcription and an insulator region is shown in Figure 16. Considerable amount of the transcripts belongs to a previously not identified non-coding RNA, CTO ("close to oriL"), located between the *ul21* gene and the *oriL*, between bases 63673–63958 on the complementary strand of genome KJ717942.1 (Oláh et al. 2015). The CTO (Reads Per Kilobase Million - RPKM= 1.6×10^{6}) in the total RNA library and *us1* (RPKM=2.32×10^{5}) were the most abundant transcripts. Despite lytic infection used in the experiment, the two latency associated transcripts (LAT and LLT) were found to be expressed at a low level, but we were not able to determine splicing donor and acceptor sites due to insufficient coverage at these regions. The hypothetical *ORF1.2* (Baumeister et al. 1995) sequence was also found to be transcribed, involving 5′upstream regions. However single-base localization of the transcription start site is hindered due to numerous repeats being present in the genomic sequence in the interval 730–960 bp. We have found polyadenylation peaks on both the positive and negative strands, mostly in accordance with already existing ORF annotations, and also long noncoding RNAs, including the latency-associated transcript (LAT) and the long-latency transcript (LLT).



16. Figure Example of a leaky transcription and an insulator intergenic repeat sequence. Legend: orange: coding sequences, blue: transcripts, blue circles: PA site on + strand, red circles: PA site on –strand, green circles: alternative PA site on + strand. Expression levels (in coverage per base): upper box: PA-Seq expression, green: +strand read coverage, black: −strand read coverage, lower box: totalRNA sequencing, blue: +strand coverage, red: −strand coverage.

Discussion

In this work we presented our findings of our research that aimed to find how deletion of key genes of the Pseudorabies virus affects its global gene expression levels. We constructed gene deleted mutants for the *us1* gene (that encodes ICP22), and for the *us7/us8* gene pair that encode for the gI and gE proteins respectively, which form heterodimers (the gE/gI complex). Gene expression levels of the mutant viruses were investigated at different times post infection using reverse transcription coupled real-time PCR (also known as RT-RT PCR or RT^2-PCR). The transcriptome of the wild type virus was also investigated with Illumina NGS platform, to obtain a better picture of the viral transcriptome.

In both gene deletion experiments we used high titer infections to prevent secondary infections from newly produced viruses which would occur at low titers. The time points at which samples were taken were chosen to reflect the rapid life-cycle of the virus. For the us7us8 experiment we have increased the frequency of samples to get better resolution at later stages of infection.

First, we assessed the effects of the deletion of the gene us1 which encodes the ICP22 protein. The protein has many functions, including regulation of gene expression, and is essential for proper virion composition (Orlando et al. 2006). Functions of this protein has been fairly well researched in HSV-1 and abundant literature is available, but not for PRV. The expression kinetics of *us1* showed an interesting titer dependency, showing early kinetics in high-titer and late kinetics in low-titer infections. It has been shown in *Equine herpesvirus I*, that is a relative of PRV, that the open reading frame harboring the ICP22 homolog gene of this virus can be expressed in two different forms of RNAs, one with early and one with late kinetics (Holden et al. 1992). Similar in-frame expression from the us1 open reading frame (ORF) was shown in HSV-1 (Carter & Roizman 1996). It was not in the focus of our investigations to further pursue the reasons of this phenomenon at the time of the experiment, and the latter sequencing of the transcriptome did not indicate the presence of a novel RNA transcript at this region. Our findings of ICP22 deletion in PRV selectively affecting different kinetic classes of genes, are in contrast to what was shown in HSV-1 (Guo et al. 2012), namely that in HSV-1 ICP22 inhibits transcription of promoters of the IE, E and L classes. It appears that ICP22 has

excessive stimulatory effect on global gene expression of the virus in the early phase of the infection (30 min and 1h post infection). This could be either by directly enhancing transcription of viral promoters or inhibiting host cell genes that would otherwise digest viral RNAs. In HSV-1, ICP22 was shown to interact with Cyclin-Dependent Kinase (CDK) (Zaborowska et al. 2014), through which it can inhibit host cell gene expression. Similar function could also be attributed to ICP22 in PRV. Either possible function of the protein would require to be readily present at early times of infection, which is in agreement with the fact that ICP22 was shown to be present in the tegument of PRV (Kramer et al. 2011). These results were published (Takács et al. 2013) with many other findings that are not part of this thesis, contributed greatly to the literature of ICP22 in PRV.

Similar measurements were conducted with the *us7/us8* deleted mutant virus, and reported in greater details, as this series of experiments is the main focus of this thesis. The heterodimer of the products of these genes play important role in cell-to-cell spreading of the virus and other connected functions (Snyder et al. 2008; Kratchmarov et al. 2013; Howard et al. 2014; D C Johnson et al. 2001). However we were interested in its part in viral gene expression, an area not well documented. Unlike *us1*, neither of the two genes in question showed titer dependence in terms of kinetics, both showed typical E/L kinetics at all titer levels. The genes were expected to have the same kinetics, since they form heterodimers. Deletion of the gene pair led to heavily reduced gene expression levels for all kinetic classes in the first 6h of the infection (with few exceptions). On the other hand, the opposite was observable for the later stages of infection (again with few exceptions). Late genes were less affected by this late stage excitatory effect compared to E and E/L genes. This is in agreement, with our finding that relative contribution of a single gene's product is less for L genes in the later stages of infection. The mutation affected the *ep0* gene in extreme measures, having this gene vastly overexpressed in late stages of infection in the mutant virus. *ep0* is a known transcription factor of the virus and in our earlier work we assumed that it may be responsible for viral gene expression synchronization (Tombácz et al. 2012). Furthermore we found that the expression of *ie180* (that is the major transactivator in PRV) in the mutant virus is correlated to the expression of other genes while similar phenomenon is only weakly present in wild-type. This, together with the considerable overexpression of the *ep0* could explain why did the expression individual genes' expression followed the global trends better in the mutant virus compared to the wild type. Our

hypothesis is that *us7* and *us8* have yet an unknown effect on *ep0* transcription, which in turn affects the expression and/or the activity of *ie180* and/or other gene products .Since it is known that *us7/us8* deleted mutants are not hindered in DNA reproduction (Snyder et al. 2008), and our results do not show significant changes in DNA replication (data not shown here), thus our opinion is that it is not the root cause behind the characteristics of the mutant virus. The literature of the gE/gI protein complex is focused around the function of the protein in cellular sorting, viral spreading and related, or the use of gE/gI mutant viruses as vaccine (due to its attenuation). Thus, our recently published work (Póka et al. 2017) with results that were shown in this thesis along with others that are not detailed here, sheds light to a new area of PRV research.

The purpose of the sequencing experiment was dual: we wanted to have a general picture of the transcriptome, that is basically the results reported here briefly, and the more important focus of this particular experiment was to identify elements of the transcriptional interference network (Boldogköi 2012), and to provide basis for further experiments in that topic.

Read alignment for the total RNA library with random-hexamer primed reads had almost 1000 times genome wide coverage, however the distribution of the reads was not homogenous. The maximum number of possible random hexamers is 4096 (4^6 – 4 possible bases in 6 independent positions), that is already a limited set of combinations, of which many do not have great complexity which reduces the chance of annealing. To worsen things the PRV has extremely high GC ratio, further filtering the set of useful combinations. These, altogether with the difference of expression levels between different genes, results in an unequal distribution of read mappings showing peaks consuming several tens of thousands of reads, leaving other regions with only few hundreds or thousands. The high-resolution transcriptome mapping resulting from the total RNA sequencing confirmed that the PRV genome is indeed extensively transcribed, and also confirmed that no laboratory environment induced mutations were present in our strain. The use of anchored primers – that is adding a dV (short-hand denomination for either dG (deoxyguanine), dA (deoxyadenine) or dC (deoxycytosine) and a dN (short-hand for either of the four deoxy-bases), hence the name oligotdT(VN) – ensure s that the primers are annealed to the 5’ end of the polyA tail. This compensate for the loss in depth due to the high fraction of reads containing solely adenine bases on the use of

conventional oligo(dT)primers. Results of the transcriptome profiling mostly confirmed available knowledge, but also provided new details about the virus, such as the newly discovered non-coding RNA, termed CTO or the evidence of the leaky transcription termination sites, and others that are not part of this thesis but thoroughly documented in our publication (Oláh et al. 2015).

Summary

Our subject of investigation was the Pseudorabies virus that is a natural pathogen of swin, causing serious economic loss at pig farms. On the other hand, PRV also serves as model organism for human pathogen alpha herpes viruses such as *Herpes simplex virus 1* (HSV-1). The ability of the virus to spread through synaptic junctions, makes it a natural neural tracer, further increasing its research value.

In this work we presented the effects of deletion of genes *us1*, and the gene pair *us7/us8* on the global transcriptome of the virus along with the summary of transcriptome profiling obtained via RNA-sequencing. The expression level of the selected genes were measured using quantitative reverse transcriptase-based real-time PCR technique at multiple post-infection time-points. To extend our knowledge on the PRV transcriptome we have carried out RNA sequencing experiment using Illumina platform. Our experiments yielded the following key results (among others):

- The expression of *us1* gene is titer dependent in the wild type Pseudorabies virus
- The deletion of us1 affects kinetic classes selectively
- The average expression for all kinetic classes were higher in the *us7/us8* mutant than in the wild type virus in late stages of infection
- The different kinetic classes are affected differently by the deletion of *us7/us8*
- The expression of the *ie180* gene is correlated with the expressions of PRV genes in the *us7/us8* mutant
- Summary of a high resolution profile of the PRV transcriptome

The results of these experiments were published in various journals and contributed to the literature of the PRV.

Acknowledgements

I would like to express my gratitude for all the help and support to:

Prof. BOLDOGKŐI, Zsolt

Prof. DUDA, Ernő

CSABAI, Zsolt

OLÁH, Péter

PRAZSÁK, István

Dr. TAKÁCS, Irma

Dr. TOMBÁCZ, Dóra

And all members of the Department of Medical Biology at University of Szeged.

References

Ahmed, F., Kumar, M. & Raghava, G.P.S., 2009. Prediction of polyadenylation signals in human DNA sequences using nucleotide frequencies. *In silico biology*, 9(3), pp.135–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19795571 [Accessed May 21, 2017].

Ajueszky, A., 1902. A contagious disease, not readily distinguishable from rabies, with unknown origin. *Veterinarius*, 12, pp.387–396.

Anderson, L. a et al., 2008. Pseudorabies (Aujeszky's Disease) and Its Eradication. *USDA:APHIS Technical Bulletin #1923*.

Bastian, T.W. & Rice, S.A., 2009. Identification of sequences in herpes simplex virus type 1 ICP22 that influence RNA polymerase II modification and viral late gene expression. *Journal of virology*, 83(1), pp.128–39. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18971282 [Accessed April 30, 2017].

Baumeister, J., Klupp, B.G. & Mettenleiter, T.C., 1995. Pseudorabies virus and equine herpesvirus 1 share a nonessential gene which is absent in other herpesviruses and located adjacent to a highly conserved gene cluster. *Journal of virology*, 69(9), pp.5560–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7637001 [Accessed May 28, 2017].

Boldogkoi, Z. et al., 2009. Genetically timed, activity-sensor and rainbow transsynaptic viral tools. *Nature methods*, 6(2), pp.127–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19122667 [Accessed January 19, 2016].

Boldogköi, Z., 2012. Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. *Frontiers in Genetics*, 3, p.122. Available at: http://journal.frontiersin.org/article/10.3389/fgene.2012.00122/abstract [Accessed April 30, 2017].

Carter, K.L. & Roizman, B., 1996. The promoter and transcriptional unit of a novel herpes simplex virus 1 alpha gene are contained in, and encode a protein in frame with, the open reading frame of the alpha 22 gene. *Journal of virology*, 70(1), pp.172–8. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/8523523 [Accessed May 6, 2017].

Cheung, A.K., 1989. DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies virus. *Nucleic acids research*, 17(12), pp.4637–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2546124 [Accessed April 30, 2017].

Dodding, M.P. & Way, M., 2011. Coupling viruses to dynein and kinesin-1. *The EMBO Journal*, 30(17), pp.3527–3539. Available at: http://emboj.embopress.org/cgi/doi/10.1038/emboj.2011.283 [Accessed January 11, 2016].

Elhai, J. & Wolk, C.P., 1988. A versatile class of positive-selection vectors based on the nonviability of palindrome-containing plasmids that allows cloning into long polylinkers. *Gene*, 68(1), pp.119–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2851487 [Accessed May 21, 2017].

Frampton, A.R. et al., 2005. HSV trafficking and development of gene therapy vectors with applications in the nervous system. *Gene Therapy*, 12(11), pp.891–901. Available at: http://www.nature.com/doifinder/10.1038/sj.gt.3302545 [Accessed April 30, 2017].

Gu, Z. et al., 2015. A novel inactivated gE/gI deleted pseudorabies virus (PRV) vaccine completely protects pigs from an emerged variant PRV challenge. *Virus Research*, 195, pp.57–63. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0168170214003712 [Accessed December 6, 2015].

Guo, L. et al., 2012. Herpes simplex virus 1 ICP22 inhibits the transcription of viral gene promoters by binding to and blocking the recruitment of P-TEFb. *PloS one*, 7(9), p.e45749. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23029222 [Accessed April 30, 2017].

Holden, V.R. et al., 1992. ICP22 homolog of equine herpesvirus 1: expression from early and late promoters. *Journal of virology*, 66(2), pp.664–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1370553 [Accessed May 6, 2017].

Howard, P.W. et al., 2014. Herpes Simplex Virus gE/gI Extracellular Domains Promote Axonal Transport and Spread from Neurons to Epithelial Cells. *Journal of virology*, 88(19),

pp.11178–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25031334 [Accessed January 16, 2016].

Husak, P.J., Kuo, T. & Enquist, L.W., 2000. Pseudorabies virus membrane proteins gI and gE facilitate anterograde spread of infection in projection-specific neurons in the rat. *Journal of virology*, 74(23), pp.10975–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11069992 [Accessed May 28, 2017].

Ihara, S. et al., 1983. Characterization of the immediate-early functions of pseudorabies virus. *Virology*, 131(2), pp.437–454. Available at: http://linkinghub.elsevier.com/retrieve/pii/004268228390510X [Accessed April 30, 2017].

Johnson, D.C. et al., 2001. Herpes simplex virus gE/gI sorts nascent virions to epithelial cell junctions, promoting virus spread. *Journal of virology*, 75(2), pp.821–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11134295 [Accessed August 23, 2016].

Johnson, D.C., Webb, M. & Wisner, T.W., 2001. Herpes Simplex Virus gE / gI Sorts Nascent Virions to Epithelial Cell Junctions , Promoting Virus Spread Herpes Simplex Virus gE / gI Sorts Nascent Virions to Epithelial Cell Junctions , Promoting Virus Spread. *Journal of Virology*, 75(2), pp.821–833. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=113978&tool=pmcentrez&rendertype=abstract [Accessed December 6, 2015].

Klupp, B.G. et al., 2004. Complete, annotated sequence of the pseudorabies virus genome. *Journal of virology*, 78(1), pp.424–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14671123 [Accessed May 20, 2017].

Knapp, a C., Husak, P.J. & Enquist, L.W., 1997. The gE and gI homologs from two alphaherpesviruses have conserved and divergent neuroinvasive properties. *Journal of virology*, 71(8), pp.5820–5827. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=191837&tool=pmcentrez&rendertype=abstract [Accessed December 6, 2015].

Kramer, T. et al., 2011. Proteomic Characterization of Pseudorabies Virus Extracellular

Virions. *Journal of Virology*, 85(13), pp.6427–6441. Available at: http://jvi.asm.org/cgi/doi/10.1128/JVI.02253-10 [Accessed January 11, 2016].

Kramer, T. & Enquist, L., 2013. Directional Spread of Alphaherpesviruses in the Nervous System. *Viruses*, 5(2), pp.678–707. Available at: http://www.mdpi.com/1999-4915/5/2/678/ [Accessed April 30, 2017].

Kratchmarov, R. et al., 2013. Glycoproteins gE and gI are required for efficient KIF1A-dependent anterograde axonal transport of alphaherpesvirus particles in neurons. *Journal of virology*, 87(17), pp.9431–40. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3754139&tool=pmcentrez&rendertype=abstract [Accessed January 17, 2016].

Lamote, J.A.S. et al., 2017. The Pseudorabies Virus Glycoprotein gE/gI Complex Suppresses Type I Interferon Production by Plasmacytoid Dendritic Cells R. M. Longnecker, ed. *Journal of Virology*, 91(7), pp.e02276-16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28122975 [Accessed May 28, 2017].

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), pp.357–359. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22388286 [Accessed May 21, 2017].

Mestdagh, P. et al., 2009. A novel and universal method for microRNA RT-qPCR data normalization. *Genome biology*, 10(6), p.R64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19531210 [Accessed May 27, 2017].

Oláh, P. et al., 2015. Characterization of pseudorabies virus transcriptome by Illumina sequencing. *BMC Microbiology*, 15(1), p.130. Available at: http://www.biomedcentral.com/1471-2180/15/130 [Accessed March 19, 2017].

Orlando, J.S. et al., 2006. ICP22 is required for wild-type composition and infectivity of herpes simplex virus type 1 virions. *Journal of virology*, 80(19), pp.9381–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16973544 [Accessed May 7, 2017].

Poffenberger, K.L., Raichlen, P.E. & Herman, R.C., 1993. In vitro characterization of a herpes

simplex virus type 1 ICP22 deletion mutant. *Virus genes*, 7(2), pp.171–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8396283 [Accessed April 30, 2017].

Póka, N. et al., 2017. Deletion of the us7 and us8 genes of pseudorabies virus exerts a differential effect on the expression of early and late viral genes. *Virus Genes*, pp.1–10. Available at: http://link.springer.com/10.1007/s11262-017-1465-8 [Accessed May 8, 2017].

Pomeranz, L.E., Reynolds, A.E. & Hengartner, C.J., 2005. Molecular Biology of Pseudorabies Virus: Impact on Neurovirology and Veterinary Medicine. *Microbiology and Molecular Biology Reviews*, 69(3), pp.462–500. Available at: http://mmbr.asm.org/cgi/doi/10.1128/MMBR.69.3.462-500.2005 [Accessed October 18, 2015].

Prod'hon, C. et al., 1996. Characterization of Regulatory Functions of the HSV-1 Immediate-Early Protein ICP22. *Virology*, 226(2), pp.393–402. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8955059 [Accessed April 30, 2017].

Robinson, J.T. et al., 2011. Integrative genomics viewer. *Nature biotechnology*, 29(1), pp.24–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21221095 [Accessed May 21, 2017].

Rutherford, K. et al., 2000. Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England)*, 16(10), pp.944–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11120685 [Accessed May 21, 2017].

Snyder, A., Polcicova, K. & Johnson, D.C., 2008. Herpes simplex virus gE/gI and US9 proteins promote transport of both capsids and virion glycoproteins in neuronal axons. *Journal of virology*, 82(21), pp.10613–24. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2573198&tool=pmcentrez&rendertype=abstract [Accessed January 17, 2016].

Strack, A.M. & Loewy, A.D., 1990. Pseudorabies virus: a highly specific transneuronal cell body marker in the sympathetic nervous system. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 10(7), pp.2139–47. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/1695943 [Accessed December 27, 2015].

STRAUSS, J.H. et al., 2008. CHAPTER 7 – DNA-Containing Viruses. In *Viruses and Human Disease*. pp. 261–323. Available at: http://www.sciencedirect.com/science/article/pii/B9780123737410500106 [Accessed May 21, 2017].

Takács, I.F. et al., 2013. The ICP22 protein selectively modifies the transcription of different kinetic classes of pseudorabies virus genes. *BMC molecular biology*, 14, p.2. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3599583&tool=pmcentrez&rendertype=abstract.

Tombácz, D., Tóth, J.S. & Boldogkői, Z., 2011. Deletion of the virion host shut-off gene of pseudorabies virus results in selective upregulation of the expression of early viral genes in the late stage of infection. *Genomics*, 98(1), pp.15–25. Available at: http://www.sciencedirect.com/science/article/pii/S0888754311000784 [Accessed May 4, 2017].

Tombácz, D., Tóth, J.S. & Boldogkői, Z., 2012. Effects of deletion of the early protein 0 gene of pseudorabies virus on the overall viral gene expression. *Gene*, 493(2), pp.235–242. Available at: http://www.sciencedirect.com/science/article/pii/S0378111911007190 [Accessed May 8, 2017].

Trapnell, C., Pachter, L. & Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), pp.1105–1111. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19289445 [Accessed May 21, 2017].

Verin, R. et al., 2014. Serologic, molecular, and pathologic survey of pseudorabies virus infection in hunted wild boars (Sus scrofa) in Italy. *Journal of wildlife diseases*, 50(3), pp.559–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24779458 [Accessed January 19, 2016].

Verpoest, S., Cay, A.B. & De Regge, N., 2014. Molecular characterization of Belgian pseudorabies virus isolates from domestic swine and wild boar. *Veterinary microbiology*, 172(1–2), pp.72–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24908275

[Accessed January 19, 2016].

Wang, Y. et al., 2015. Molecular epidemiology of outbreak-associated pseudorabies virus (PRV) strains in central China. *Virus Genes*, 50(3), pp.401–409. Available at: http://link.springer.com/10.1007/s11262-015-1190-0 [Accessed October 18, 2015].

Woźniakowski, G. & Samorek-Salamonowicz, E., 2015. *Animal herpesviruses and their zoonotic potential for cross-species infection*, Available at: http://www.ncbi.nlm.nih.gov/pubmed/26094506 [Accessed April 7, 2017].

Wu, R. et al., 2013. Emergence of virulent pseudorabies virus infection in northern China. *Journal of veterinary science*, 14(3), pp.363–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23820207 [Accessed May 21, 2017].

Yamane, I., Ishizeki, S. & Yamazaki, H., 2015. Aujeszky's disease and the effects of infection on Japanese swine herd productivity: a cross-sectional study. *The Journal of veterinary medical science / the Japanese Society of Veterinary Science*, 77(5), pp.579–82. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4478737&tool=pmcentrez&rendertype=abstract [Accessed January 19, 2016].

Yu, T. et al., 2016. Growth characteristics and complete genomic sequence analysis of a novel pseudorabies virus in China. *Virus Genes*, 52(4), pp.474–483. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27012685 [Accessed April 5, 2017].

Zaborowska, J. et al., 2014. Herpes Simplex Virus 1 (HSV-1) ICP22 protein directly interacts with cyclin-dependent kinase (CDK)9 to inhibit RNA polymerase II transcription elongation. *PloS one*, 9(9), p.e107654. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25233083 [Accessed May 7, 2017].

Appendix

I.

II.

III.