

Nanopore Assay Reveals Cell-type-dependent Gene Expression of Indiana Vesiculovirus and Differential Host Cell Response

Balázs Kakuk^{1#}, András A. Kiss^{1#}, Gábor Torma¹, Zsolt Csabai¹, István Prazsák, Máté Mizik, Klára Megyeri², Dóra Tombácz¹, Zsolt Boldogkői^{1*}

¹ Department of Medical Biology, Faculty of Medicine, University of Szeged, Szeged, Hungary

² Department of Medical Microbiology and Immunobiology, Faculty of Medicine, University of Szeged, Szeged, Hungary.

* Corresponding author: Zsolt Boldogkői (boldogkoi.zsolt@med.u-szeged.hu)

These three authors contributed equally to this work

BK: kakuk.balazs@med.u-szeged.hu

GT: torma.gabor@med.u-szeged.hu

AK: Kis.andras@med.u-szeged.hu

ZC: csabai.zsolt@med.u-szeged.hu

IP: prazsak.istvan@med.u-szeged.hu

MM: mizikmate@gmail.com

KM: megyeri.klara@med.u-szeged.hu

DT: tombacz.dora@med.u-szeged.hu

ZB: boldogkoi.zsolt@med.u-szeged.hu

keywords: vesicular stomatitis virus, IVV, transcriptome, nanopore sequencing, long-read sequencing

Abstract

Indiana Vesiculovirus (IVV; formerly as Vesicular stomatitis virus and Vesicular stomatitis Indiana virus) causes a disease in livestock that is very similar to the foot and mouth disease thereby an outbreak may lead to significant economic loss. Long-read sequencing (LRS) -based approaches revealed a hidden complexity of the transcriptomes in several viruses already. This technique was utilized already for the sequencing of the IVV genome, but our study is the first for the application of this technique for the profiling of IVV transcriptome. Since LRS is able to sequence full-length RNA molecules, and thereby providing more accurate annotation of the transcriptomes than the traditional short-read sequencing methods. The objectives of this study were to assemble the complete transcriptome of using nanopore sequencing, to ascertain cell-type specificity and dynamics of viral gene expression and to evaluate host gene expression changes induced by the viral infection. We carried out a time-course analysis of IVV gene expression in human glioblastoma and primate fibroblast cell lines using a nanopore-based LRS approach and applied both amplified and direct cDNA sequencing, as well as cap-selection for a fraction of samples. Our investigations revealed that, although the IVV genome is simple, it generates a relative complex transcriptomic architecture. In this study, we also demonstrated that IVV transcripts vary in structure and exhibit differential gene expression patterns in the two examined cell types.

Introduction

Indiana Vesiculovirus (IVV) is a negative single-stranded RNA virus belonging to the Rhabdoviridae family[1]. The virus causes a zoonotic disease, the infection spreads between mammalian hosts via insect bites or direct contact [2]. Although the virus causes only mild symptoms in humans [3,4], including fever, myalgia, headache, vomiting [5], enlarged lymph nodes and conjunctivitis, it causes vesicular disease in animals including horses, cattle and especially pigs [6], which is the natural host of the virus [7]. The vesicular disease is very similar to the foot and mouth disease, and thus can lead to losses in domestic livestock and thus to significant economic loss. VSV infection used to be common among laboratory workers and animal handlers as well.

The IVV genome is small (11,161 kb) [2] encoding five polypeptides: N, P, M, G and L. The large L protein is an RNA-dependent RNA polymerase (RdRp), which is also responsible for capping and polyadenylation of IVV mRNAs [8,9]. The nucleoprotein (N) [10] surrounds the RNA molecule. The phosphoprotein (NS) is a catalytic cofactor for the L protein [11]. The matrix protein (M) performs many functions, including assembly, packaging, apoptosis, and blocking of host RNAs [12,13]. Glycoprotein (G) is responsible for the entry into the host cells [14,15], it is required for the attachment of the virion to the cell surface Low-Density Lipoprotein (LDL) receptor [16], enabling the virus to enter the cell via receptor-mediated endocytosis [17]. The acidity of the endosome lumen causes a conformational change in the G protein, thereby activating it [18]. The fusion between the viral envelope and the endosomal membrane is then facilitated by the activated G protein, which leads to the release of the viral helical nucleocapsid into the cytoplasm of the host cell.

The (-) strand VSV RNA serves as a template for transcription of the five major mRNAs. The viral capsid contains small amounts of L-protein and phosphoproteins that can initiate viral RNA synthesis after the intrusion [19]. The transcription and replication of IVV begins at the 3'-end of the genomic RNA. The viral RdRp functions in two modes [20]: in replication, it produces a full-length RNA molecule; whereas in transcription, the individual mRNA molecules are synthesized.

Free cytoplasmic ribosomes are involved in the translation of four proteins (N, NS, M, L), but G protein is translated by another pathway that is controlled by endoplasmic reticulum-bounded ribosomes [21]. The newly synthesized proteins prepare a full-length complementary (+) strand RNA, which serves as a template for the synthesis of the (-) strand RNA genome. This in turn is incorporated into the progeny virions. The newly synthesized G protein is glycosylated and oligomerized [22], then enters the secretory pathway, and is transported to the plasma membrane. When every other viral component gets transported, G protein initiates the assembly of virions, which can now egress from the host cell, being ready for the next infection [23].

Next-generation short-read sequencing (SRS) techniques can accurately characterize viral gene expression [24], but do not provide high-resolution details of the various transcript isoforms, and multigenic and overlapping transcripts. The emerging long-read sequencing (LRS) techniques can circumvent these limitations because they are able to read full-length transcripts [25,26]. The LRS techniques have been becoming increasingly popular in viral genome and transcriptome researches, and the studies based on these platforms report an unexpectedly large complexity of the viral transcriptomes [27–30]. An LRS approach has already been used for the sequencing of the IVV genome [31], but our study is the first for the application of this technique for the profiling of IVV transcriptome.

In this work, we investigate the structural and kinetic aspects of the polyadenylated fraction of the IVV transcriptome in fibroblast and glial cell lines using the ONT MinION amplified and non-amplified cDNA sequencing techniques. Furthermore, we describe the detected significant differences in the gene expression dynamics of two host cells.

Methods

The experimental system used in this study is shown in **Additional File 1: Fig. S1**. All samples were carried out in duplicates.

Cells and Viral Infection

Strain Indiana of IVV was propagated on African green monkey kidney fibroblast (Vero) and on human glioblastoma (T98G) cell lines (ECACC). Cells were grown in DMEM (Gibco/Thermo Fisher Scientific), supplemented with 5% fetal bovine serum (Gibco/Thermo Fisher Scientific) and 80 µg/ml gentamycin (Gibco/Thermo Fisher Scientific) at 37°C in the presence of 5% CO₂. Both cell types were infected with a high multiplicity of infection (MOI = 5), and samples were taken at multiple time points (1h, 6h, 15h, 24h).

Isolation of RNA

Total RNAs were isolated using NucleoSpin® RNA kit (Macherey-Nagel) according to the manufacturers recommendation. Samples were treated with Ambion® TURBO DNA-free™ kit to eliminate residual DNA contamination. The concentration of RNA samples was determined using a Qubit® 4.0 Fluorometer and the Qubit RNA BR Assay Kit (Life Technologies). The poly(A)+ RNA fraction was isolated applying Oligotex mRNA Mini Kit (Qiagen).

Oxford Nanopore MinION sequencing

For the preparation of cDNA libraries, the polyA(+) RNA fraction was reverse transcribed using an oligo(d)T-containing primer [(VN)T20 (Bio Basic, Canada)]. The RT reaction carried out using SuperScript IV enzyme (Life Technologies), a strand-switching oligo [containing three O-methylguanine RNA bases (PCR_Sw_mod_3G; Bio Basic, Canada)] added to the sample. The cDNAs were amplified using LongAmp Taq 2X Master Mix (New England Biolabs) and Ligation Sequencing Kit Primer Mix according to the ONT Kit's manual. End repair was made on the samples using the NEBNext End repair / dA-tailing Module (New England Biolabs). The "barcoding" made by the specific barcode (ONT PCR Barcoding Kit 96; EXPPBC096), and ligated to the sample according to the 1D PCR barcoding genomic DNA (SQK-LSK109) protocol. Barcoded samples were amplified by PCR using LongAmp Taq 2X Master Mix. The PCR product was end-repaired, then it was followed by adapter ligation utilizing the sequencing adapters supplied in the kit and NEBNext Quick Ligation Module (New England Biolabs). The cDNA sample was purified between each step using Agencourt AMPure XP magnetic beads (Beckman Coulter). To avoid the analysis of potential false PCR products Non-amplified cDNA libraries were also prepared using the ONT's Direct cDNA (dcDNA) Sequencing Kit (SQK-DCS109), according to the manufacturer's recommendations / as we described earlier (Tombácz et al. 2021 Front Genet). The amplified libraries were run on MinION SpotOn Flow Cells (R9.4), while the dcDNA samples were load onto ONT Flongle Flow Cells.

Cap Selection Protocol

For capturing the 5'-cap structure, a specific adapter was ligated to the cDNAs using the Lexogen's TeloPrime Full-Length cDNA Amplification Kit (25°C, overnight). The samples were amplified by PCR using the Enzyme Mix and the Second-Strand Mix from the TeloPrime Kit. Detailed protocol can be found in our earlier publication (<https://www.nature.com/articles/sdata2018119>). The reactions were performed in a Veriti Thermal Cycler, and the samples purified on silica membranes (TeloPrime Kit) after the enzymatic reactions. The sequencing ready libraries were loaded onto R9.4 SpotON Flow Cells.

Bioinformatic analyses

Guppy software v3.3.3 (ONT) was used for base calling of the data from MinION sequencing. The raw reads were aligned to the Vesicular stomatitis Indiana virus reference genome (NCBI Nucleotide accession: NC_001560.1) using *minimap2* with the following options: -ax splice -Y -C5 --cs. The

LoRTIA software (<https://github.com/zsolt-balazs/LoRTIA>), which can filter out false products, was used to find TESs and TSSs and to annotate viral transcripts. The LoRTIA toolkit is also. We used additional criteria for the annotations to eliminate potentially spurious transcripts: only those features were accepted as true that were detected in least two amplified cDNA and in one dcDNA library.

Host cell's gene abundance estimation was carried out with salmon [32] on the GCA_000409795.2 and GCA_000001405.28 genome assemblies for the Vervet Monkey fibroblast and Human glia cells, respectively. Transcript counts were summed per gene and then translated to the SYMBOL database, as gene SYMBOLS for most genes are shared between *H. sapiens* and *C. sabaeus*. In order to be able to assess the differences between the two host cell lines, those genes whose gene symbol was not found in the other cell line were filtered out, i.e. only the intersection of gene SYMBOLS were used in the downstream analysis. On average ~73% of all transcript counts from each sample could be assigned into gene SYMBOLS. ImpulseDE2 [33] (utilizing DESeq2 [34]) was used to identify genes that showed differential kinetic profiles between the two host cell lines and within each cell line (Differentially Expressed Genes, DEGs). Clusterprofiler [35] was used to assess, which KEGG pathways were significantly different, based on the DEGs. Complexheatmap [36], Gviz [37], and the packages of the tidyverse [38] were used for data analysis and visualization in R [39].

Results

Time-course long-read sequencing of the IVV transcriptome reveals novel transcripts

Our investigations revealed that the simple IVV genome encodes a relatively complex transcriptomic architecture, which differs in the two investigated cell lines with regard of both the structure and the kinetics of transcripts. Two technical replicates were used from the amplified cDNA-Seq samples at each time point in both cell lines. Direct cDNA sequencing (dcDNA-Seq) was used to confirm transcript identity. The detailed sequencing statistics of IVV and host cells are shown in **Additional file 2: Table S1**. The obtained reads were analyzed using LoRTIA for the identification of transcription start sites (TSSs), transcription end sites (TESs), and the transcripts themselves. The data calculated using this program is shown in **Additional file 3: Table S2/LoRTIA**. The read length distribution is illustrated in **Additional File 1, Figure S2**.

Novel transcripts of IVV

In this work, we identified 20 nested RNAs. These contain 5'-truncated in-frame ORFs embedded into longer canonical ORFs and thus are the products of putative nested genes and might encode N-terminally truncated polypeptides. Six of these are longer TSS variants of truncated transcripts containing in-frame ORFs. Bi- and multicistronic viral mRNAs were also detected in both host cells. In fibroblast cells only bicistronic mRNAs were detected, however in the glial cells, we detected 5 multigenic transcripts as well. Three different bicistronic mRNAs were found in both cell types: N-NS, NS-M, and M-G.

The kinetic transcription profile of IVV genes

The kinetic analysis of the annotated IVV transcripts were carried out in both cell types at four sampling time points (1, 6, 15, 24 hpi) (**Figure 1**). We found a differing structural and temporal expression pattern of the IVV genes in the two cell lines, including the 5'-truncated RNAs. There are differences in the sets of IVV transcripts that are produced in the two host cells during the viral infection.

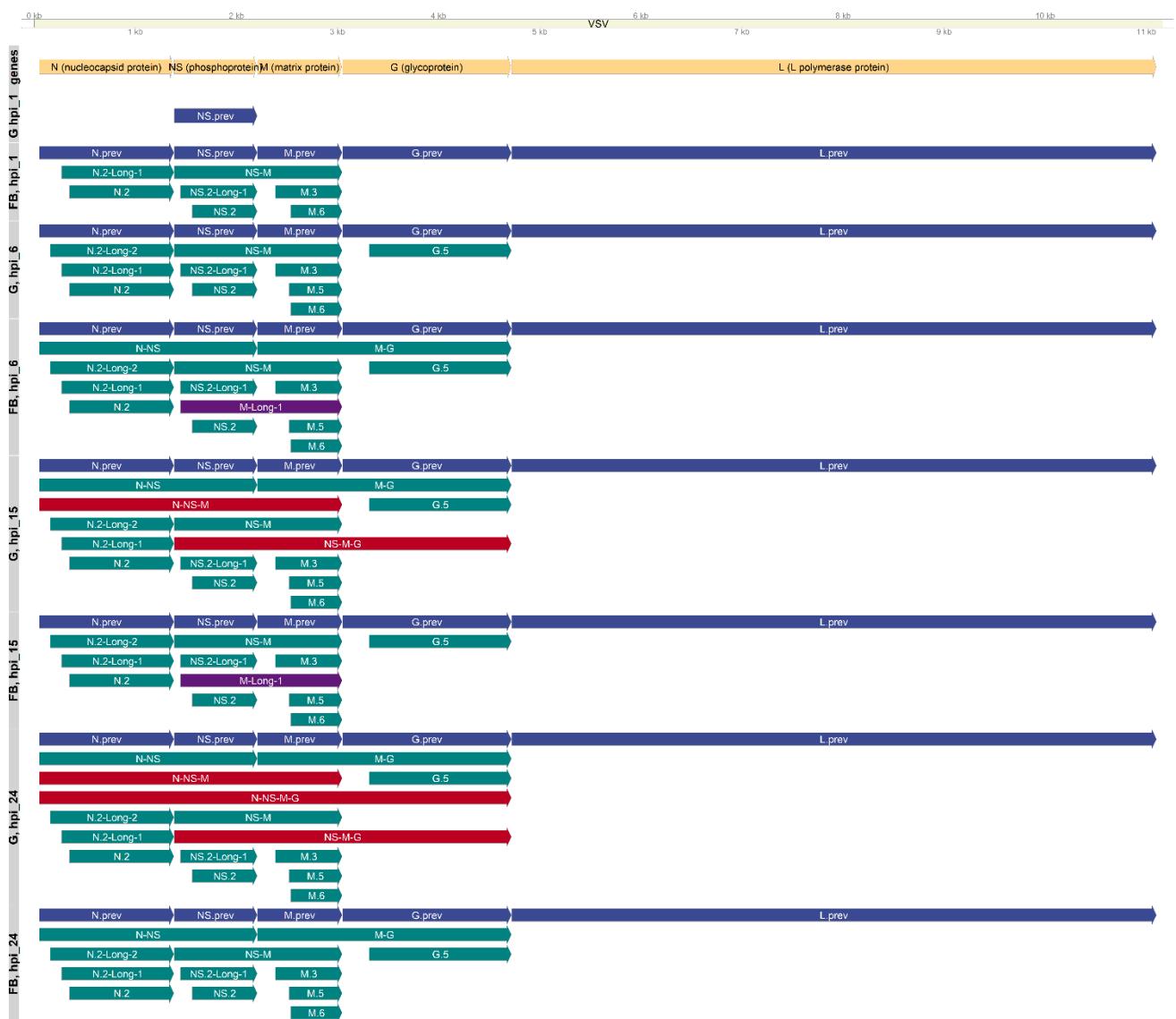


Figure 1. Kinetic transcriptome profiles of IVV. The set of annotated transcripts in each time point and in each cell line is shown. Yellow arrows indicate the IVV genes; blue arrows illustrate the previously described canonical transcripts; green arrows illustrate transcripts that are identical in both cell types; indigo arrows indicate transcripts found only in glial cells, while dark red arrows indicate transcripts found only in fibroblast cells.

There is a remarkable difference between the proportions of 5'-truncated transcripts and the canonic transcripts of the genes in each time point in the two cell types (Figure 2). Generally, fibroblast cells produced more embedded transcripts. Except for the L gene (where no high-confidence isoform was detected), each gene expressed relatively high proportions of embedded genes. Interestingly, the proportion of these truncated transcripts followed a similar pattern: very low percentage at the start of the infection, a peak at either 6 or 15 hpi; and decrease at 24 hpi. These can be viewed as isoform-switching events. The highest proportion of 5'-truncated transcripts was observed in the case of M and N genes: here 60-75% percent of the gene's expression was composed of these truncated transcripts. Even more, this pattern was seen in the case of Glia cells too, only here the proportions were obviously lower, but the shape of curve is clearly similar. This suggests that the expression of the expression of these truncated transcripts is regulated differently from their host (canonical) transcripts and also differentially in the two cell lines.

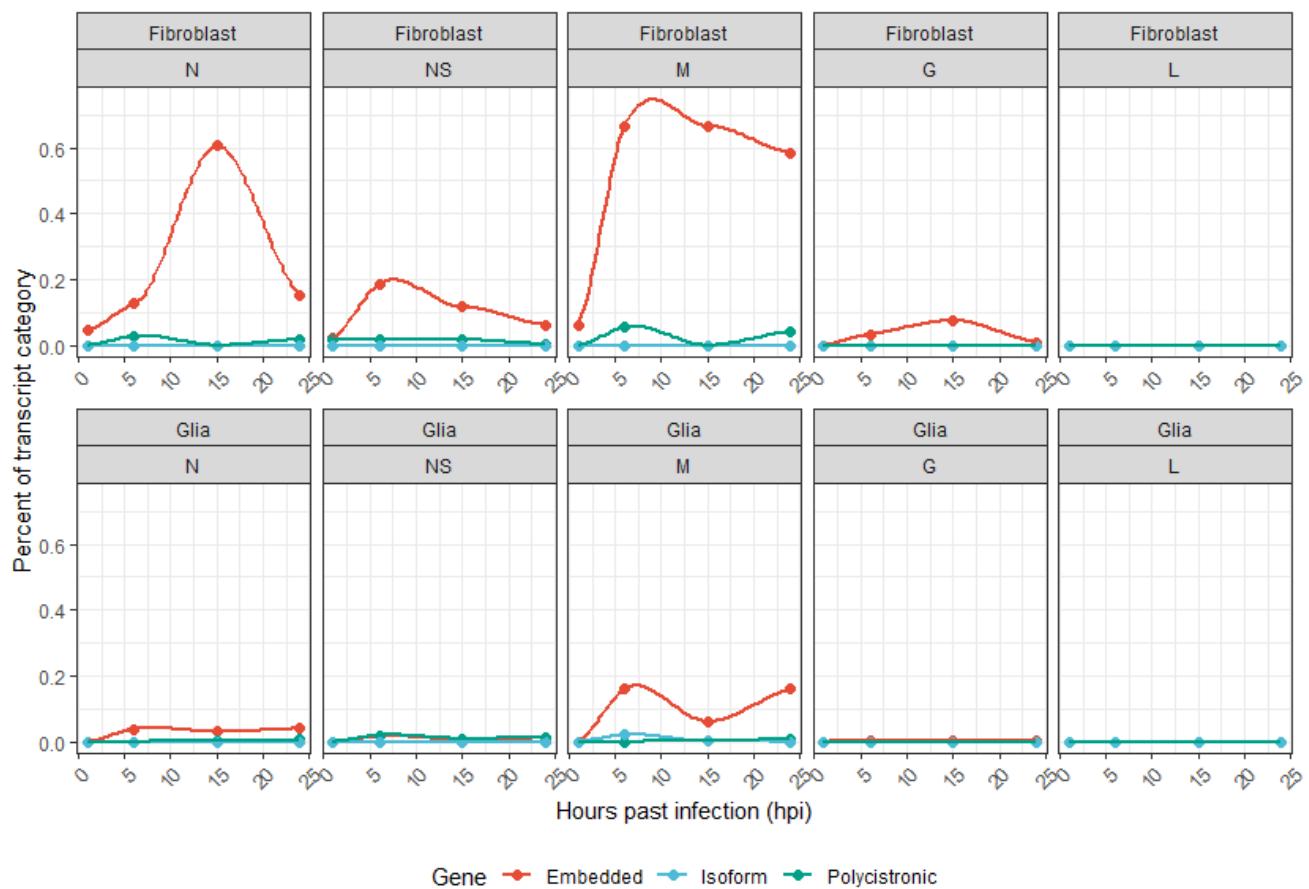


Figure 2. Relative transcript category expression. This chart illustrates the proportion of different transcript categories compared to the canonic transcripts. The values were calculated as the sum of the expression values of all the transcripts of each transcript category in each gene and in each sample divided by the expression values of the canonic transcript in the respective gene and sample.

The proportions of the polycistronic RNAs were generally low. In both cell lines, the highest values were detected in the NS gene, 6 hpi samples: 1.8% in glial cells and 5.5% in fibroblast cells. The proportions fluctuated below these values without a clear trend their expression profile.

Viral gene-level expression kinetics

Viral gene expression values were estimated with *salmon*, and the resulting count matrices were evaluated with ImpulseDE (which uses DESeq2's normalization approach), to analyze each gene's expression level as the function of time (gene expression difference between samples).

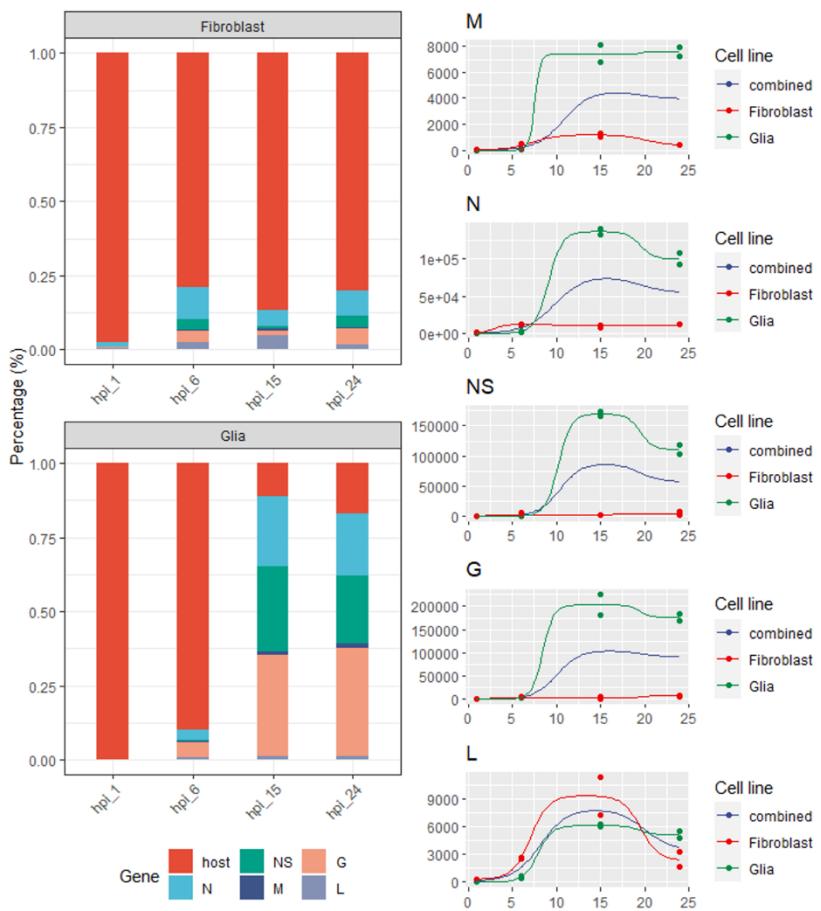


Figure 3. Gene-level expression kinetics. Left panel (stacked bar chart) shows the proportion of reads in each library as estimated by *salmon*; the filling colors shows the origin of the reads (host cell or one of the viral genes). Right panel shows the expression trajectory of each viral gene across the time points and the impulse model fitted on the data. Colors represent the two cell lines and the combined model from both cell lines.

The N, NS, M and G genes showed a somewhat similar expression pattern in the Glia cells: only a very low amount of viral reads were obtained in the 1 h post infection (hpi) samples (a total 30 and 31 reads could be pseudomapped to the viral transcriptome in the two replicates, respectively); viral transcription kick started in the hpi 6 samples, which was followed by considerable increase and a peak at hpi 15; and finally expression decreased at hpi 24 in the case of NS and N, decreased only slightly in the case of G or remained more or less the same in the case of L and M genes. The shape of the gene expression curve in the case of fibroblast cells was similar except for NS and G genes, where it fluctuated in the 6, 15 and 24 hpi samples. The gene expression levels showed a striking difference, however: in glia cells the proportion of the viral reads elevated from only 11% in the hpi 6 samples – which is comparable to 21% in the fibroblast cells, to 89% in the hpi 15 samples and decreased only slightly in the hpi 24 samples (83%). This phenomenon may be due to the IVV M protein, which blocks the escape of host mRNAs from the nucleus by blocking the nucleopores and preventing host RNAs from entering the cytoplasm [40,41]. In contrast, fibroblast cells followed a different trend. The number of viral mRNA at 1 hpi was 2.5% of all obtained reads. At 6 hpi, this was increased to 20% but at the following timepoints it did not increase further. It seems that the Vero cells were able to form some kind of a balance with the virus and didn't allow it to effectively inhibit their own gene expression, at least not until 24 hpi. These results show that there is a significant difference in viral gene expression levels and virus:to:host expression ratios between the two cell lines. On exception is the viral L mRNA, as its abundance levels showed no difference between the two cell lines: in both the glia and fibroblast cells, the L gene showed low expression values but in the fibroblast cells this accounted for a larger proportion of reads.

VSV-N and VSV-NS mRNAs. The proteins formed from them are cofactors for the virus's RdRp. This may explain why the expression levels of viral transcripts are higher in glial cells.

Host gene expression

The gene expression count matrices as produced by salmon and translated to gene symbols (only shared symbols between *H. sapiens* and *C. sabaeus* were kept) were analyzed using ImpulseDE2, first in case-control mode. In this test, genes with a low p-value exhibit different kinetic profiles between the two cell lines, but not necessarily a difference between the initial (hpi 1) and subsequent time-points within the cell line. This showed that out of the 16,133 shared genes, 1,370 were significantly differentially expressed (DEG) between fibroblast and glial cells under FDR-corrected p-value cutoff of 0.01. This DEG list was supplied as an input for Clusterprofiler to identify KEGG pathways that are differentially expressed by analyzing number of DEGs that are involved in the pathways. As a result, 35 pathways were found. Many of these are associated with viral diseases (i.e. Coronavirus, Influenza and Epstein-Barr infections), but we found several that are associated with more general cell function, i.e. Carbon metabolism and Ribosome **Additional File 1, Figure S3**. This is because many genes with widespread functions were affected by the viral infection (were differentially expressed between the two host cells).

We also carried out an analysis separately for the two cell-lines, which tests whether the gene expression deviates from a constant model (case-only mode). This showed that gene expression profiles changed significantly in 460 genes in glial cells, and in 176 genes in fibroblast cells. Overall, glia cells showed significantly more DEGs as a function of time. 71 genes overlapped in this comparison; these are the genes whose expression seem to be affected the most by the viral infection in both cell lines. There were about 3 times more DEGs in the fibroblast vs glia comparison than in the within-sample comparisons (**Additional File 1, Figure S4**). The list of DEGs for each comparison is provided in **Additional File 3**. The DEGs in each cell line (case-only analyses) were clustered together according to their expression pattern into 5 clusters. The z-score normalized gene expression values in the two cell lines for these genes are shown in **Additional File 1, Figure S5/A-B** (heatmap), and in **Additional File 1, Figure S6** (scatterplot). **Additional File 1, Figure S7** shows the z-score normalized gene expression for the viral genes. In fibroblast cells, cluster_3 showed a similar mean expression trajectory to that of the N gene and clusters 4 and 5 to that of M and L genes, while in glia cells, the expression pattern described above was similar to that of the host genes only in cluster_4 but came earlier in cluster_3 and with a delay in cluster_5. These results overall show that the effect of infection on host gene expression is completely different in the two cell types.

Discussion

In this work, several novel TSSs and TESs and associated transcripts of IVV were identified using amplified and non-amplified cDNA nanopore sequencing. We analyzed the kinetics of the transcripts in two cell lines during viral infection. Different bi- and polycistronic mRNAs and novel 5'-truncated mRNAs that are embedded in the longer canonical genes of the virus types were identified in the two cell types. Polycistronic mRNAs were expressed in low abundance only, however, in some cases, compared to the canonic transcripts, the newly identified 5'-truncated mRNAs were expressed in very high proportions. Viral gene-level expression values were also determined and interestingly, their curves are similar to what was observed in the proportions of embedded and canonic transcripts of each gene. We found a significant difference in the effectiveness of viral infection between the two cell types. Glia cells are much more sensitive to infection, while fibroblast cells are more resistant. Even though the 5'-truncated transcripts may not encode functional proteins, they may play a role in regulating gene expression. Indeed, it is possible that the isoform switching events, i.e., the change in the proportions of the truncated and canonic transcripts in the genes, contribute to the apparent resistance of the fibroblast cells against the viral infection, although this needs to be tested

experimentally. A large number of differentially expressed host genes were also identified both within each cell line (as a function of time), and between the two.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Raw datasets are available in European Nucleotide Archive: PRJEB46127. The LoRTIA pipeline is available at GitHub: <https://github.com/zsolt-balazs/LoRTIA>

Competing interests

The authors declare that they have no conflict of interest

Funding

This study was supported by the National Research, Development, and Innovation Office grants FK 128252 to DT and K 128247 to ZBo.

Authors' contribution

Conception and design: BK, DT, KM, and ZB. Contribution in laboratory work: GT, DT, ZC, MM, IP and KM. Data analysis: BK, AAK, GT, MM, IP, DT, ZB. Manuscript drafting: KB, DT and ZB. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Payne S. Family Rhabdoviridae. *Viruses*. 2017. doi:10.1016/b978-0-12-803109-4.00019-2
2. Wasik BR, Bhushan A, Ogbunugafor CB, Turner PE. Delayed transmission selects for increased survival of vesicular stomatitis virus. *Evolution (N Y)*. 2015. doi:10.1111/evo.12544
3. Rihn SJ, Aziz MA, Stewart DG, Hughes J, Turnbull ML, Varela M, et al. TRIM69 Inhibits Vesicular Stomatitis Indiana Virus. *J Virol*. 2019. doi:10.1128/jvi.00951-19
4. Fields BN, Hawkins K. Human infection with the virus of vesicular stomatitis during an epizootic. *N Engl J Med*. 1967. doi:10.1056/NEJM196711092771901
5. Johnson KM, Vogel JE, Peralta PH. Clinical and serological response to laboratory-acquired human infection by Indiana type vesicular stomatitis virus (VSV). *Am J Trop Med Hyg*. 1966. doi:10.4269/ajtmh.1966.15.244
6. Cargnelutti JF, Olinda RG, Maia LA, de Aguiar GMN, Neto EGM, Simões S V.D., et al. Outbreaks of Vesicular stomatitis Alagoas virus in horses and cattle in northeastern Brazil. *J Vet Diagnostic Investig*. 2014. doi:10.1177/1040638714553428
7. Velazquez-Salinas L, Naik S, Pauszek SJ, Peng KW, Russell SJ, Rodriguez LL. Oncolytic

Recombinant Vesicular Stomatitis Virus (VSV) Is Nonpathogenic and Nontransmissible in Pigs, a Natural Host of VSV. *Hum Gene Ther Clin Dev.* 2017. doi:10.1089/humc.2017.015

8. Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, et al. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell.* 2015. doi:10.1016/j.cell.2015.06.018
9. Heilmann E, Kimpel J, Geley S, Naschberger A, Urbiola C, Nolden T, et al. The Methyltransferase region of vesicular stomatitis virus L polymerase is a target site for functional intramolecular insertion. *Viruses.* 2019. doi:10.3390/v11110989
10. Patil G, Xu L, Wu Y, Song K, Hao W, Hua F, et al. TRIM41-Mediated Ubiquitination of Nucleoprotein Limits Vesicular Stomatitis Virus Infection. *Viruses.* 2020. doi:10.3390/v12020131
11. Banerjee AK, Abraham G, Colonna RJ. Vesicular stomatitis virus: mode of transcription. *Journal of General Virology.* 1977. doi:10.1099/0022-1317-34-1-1
12. Pan W, Song D, He W, Lu H, Lan Y, Tong J, et al. The matrix protein of vesicular stomatitis virus inhibits host-directed transcription of target genes via interaction with the TFIIH subunit p8. *Vet Microbiol.* 2017. doi:10.1016/j.vetmic.2017.07.020
13. Redondo N, Madan V, Alvarez E, Carrasco L. Impact of vesicular stomatitis virus M proteins on different cellular functions. *PLoS One.* 2015. doi:10.1371/journal.pone.0131137
14. Nikolic J, Belot L, Raux H, Legrand P, Gaudin Y, Albertini AA. Structural basis for the recognition of LDL-receptor family members by VSV glycoprotein. *Nat Commun.* 2018. doi:10.1038/s41467-018-03432-4
15. Li R, Qiao S, Chen X, Xing G, Li X, Zhang G. Vesicular stomatitis virus glycoprotein suppresses nuclear factor kappa-B- and mitogen-activated protein kinase-mediated pro-inflammatory responses dependent on sialic acids. *Int J Biol Macromol.* 2020. doi:10.1016/j.ijbiomac.2020.02.322
16. Finkelshtein D, Werman A, Novick D, Barak S, Rubinstein M. LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. *Proc Natl Acad Sci U S A.* 2013. doi:10.1073/pnas.1214441110
17. Sun X, Yau VK, Briggs BJ, Whittaker GR. Role of clathrin-mediated endocytosis during vesicular stomatitis virus entry into host cells. *Virology.* 2005. doi:10.1016/j.virol.2005.05.006
18. Roche S, Bressanelli S, Rey FA, Gaudin Y. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science (80-).* 2006. doi:10.1126/science.1127683
19. Jenni S, Bloyet LM, Diaz-Avalos R, Liang B, Whelan SPJ, Grigorieff N, et al. Structure of the Vesicular Stomatitis Virus L Protein in Complex with Its Phosphoprotein Cofactor. *Cell Rep.* 2020. doi:10.1016/j.celrep.2019.12.024
20. Qanungo KR, Shaji D, Mathur M, Banerjee AK. Two RNA polymerase complexes from vesicular stomatitis virus-infected cells that carry out transcription and replication of genome RNA. *Proc Natl Acad Sci U S A.* 2004. doi:10.1073/pnas.0401449101
21. Katz FN, Rothman JE, Knipe DM, Lodish HF. Membrane assembly: Synthesis and intracellular processing of the vesicular stomatitis viral glycoprotein. *J Supramol Cell Biochem.* 1977. doi:10.1002/jss.400070308
22. Balch WE, McCaffery JM, Plutner H, Farquhar MG. Vesicular stomatitis virus glycoprotein is sorted and concentrated during export from the endoplasmic reticulum. *Cell.* 1994. doi:10.1016/0092-8674(94)90359-X
23. Odenwald WF, Arnheiter H, Dubois-Dalcq M, Lazzarini RA. Stereo images of vesicular

stomatitis virus assembly. *J Virol.* 1986. doi:10.1128/jvi.57.3.922-932.1986

24. Neidermyer WJ, Whelan SPJ. Global analysis of polysome-associated mRNA in vesicular stomatitis virus infected cells. *PLoS Pathog.* 2019. doi:10.1371/journal.ppat.1007875

25. Byrne A, Beaudin AE, Olsen HE, Jain M, Cole C, Palmer T, et al. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat Commun.* 2017. doi:10.1038/ncomms16027

26. Tilgner H, Jahanbani F, Blauwkamp T, Moshrefi A, Jaeger E, Chen F, et al. Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat Biotechnol.* 2015. doi:10.1038/nbt.3242

27. Moldován N, Szucs A, Tombácz D, Balázs Z, Csabai Z, Snyder M, et al. Multiplatform next-generation sequencing identifies novel RNA molecules and transcript isoforms of the endogenous retrovirus isolated from cultured cells. *FEMS Microbiol Lett.* 2018. doi:10.1093/femsle/fny013

28. Moldován N, Balázs Z, Tombácz D, Csabai Z, Szűcs A, Snyder M, et al. Multi-platform analysis reveals a complex transcriptome architecture of a circovirus. *Virus Res.* 2017. doi:10.1016/j.virusres.2017.05.010

29. Szucs A, Moldován N, Tombácz D, Csabai Z, Snyder M, Boldogkui Z. Long-read sequencing reveals a GC pressure during the evolution of porcine endogenous retrovirus. *Genome Announc.* 2017. doi:10.1128/genomeA.01040-17

30. Moldován N, Tombácz D, Szucs A, Csabai Z, Balázs Z, Kis E, et al. Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. *Sci Rep.* 2018. doi:10.1038/s41598-018-26955-8

31. Wongsurawat T, Jenjaroenpun P, Taylor MK, Lee J, Tolardo AL, Parvathareddy J, et al. Rapid sequencing of multiple RNA viruses in their native form. *Front Microbiol.* 2019. doi:10.3389/fmicb.2019.00260

32. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides accurate, fast, and bias-aware transcript expression. *Nat Methods.* 2017.

33. Sander J, Schultze JL, Yosef N. ImpulseDE: Detection of differentially expressed genes in time series data using impulse models. *Bioinformatics.* 2017. doi:10.1093/bioinformatics/btw665

34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15: 550. doi:10.1186/s13059-014-0550-8

35. Yu G, Wang LG, Han Y, He QY. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi A J Integr Biol.* 2012;16: 284–287. doi:10.1089/omi.2011.0118

36. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016. doi:10.1093/bioinformatics/btw313

37. Hahne F, Ivanek R. Visualizing genomic data using Gviz and bioconductor. *Methods in Molecular Biology.* 2016. doi:10.1007/978-1-4939-3578-9_16

38. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. *J Open Source Softw.* 2019;4: 1686. doi:10.21105/joss.01686

39. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2019. Available: <https://www.r-project.org/>

40. Quan B, Seo HS, Blobel G, Ren Y. Vesiculoviral matrix (M) protein occupies nucleic acid binding site at nucleoporin pair (Rae1•Nup98). *Proc Natl Acad Sci U S A.* 2014.

doi:10.1073/pnas.1409076111

41. Rajani KR, Pettit Kneller EL, McKenzie MO, Horita DA, Chou JW, Lyles DS. Complexes of Vesicular Stomatitis Virus Matrix Protein with Host Rae1 and Nup98 Involved in Inhibition of Host Transcription. *PLoS Pathog.* 2012. doi:10.1371/journal.ppat.1002929