

RESEARCH ARTICLE

Open Access



Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus

István Prazsák^{1†}, Norbert Moldován^{1†}, Zsolt Balázs¹, Dóra Tombácz¹, Klára Megyeri², Attila Szűcs¹, Zsolt Csabai¹ and Zsolt Boldogkői^{1*}

Abstract

Background: Varicella zoster virus (VZV) is a human pathogenic alphaherpesvirus harboring a relatively large DNA molecule. The VZV transcriptome has already been analyzed by microarray and short-read sequencing analyses. However, both approaches have substantial limitations when used for structural characterization of transcript isoforms, even if supplemented with primer extension or other techniques. Among others, they are inefficient in distinguishing between embedded RNA molecules, transcript isoforms, including splice and length variants, as well as between alternative polycistronic transcripts. It has been demonstrated in several studies that long-read sequencing is able to circumvent these problems.

Results: In this work, we report the analysis of the VZV lytic transcriptome using the Oxford Nanopore Technologies sequencing platform. These investigations have led to the identification of 114 novel transcripts, including mRNAs, non-coding RNAs, polycistronic RNAs and complex transcripts, as well as 10 novel spliced transcripts and 25 novel transcription start site isoforms and transcription end site isoforms. A novel class of transcripts, the nroRNAs are described in this study. These transcripts are encoded by the genomic region located in close vicinity to the viral replication origin. We also show that the ORF63 exhibits a complex structural variation encompassing the splice sites of VZV latency transcripts. Additionally, we have detected RNA editing in a novel non-coding RNA molecule.

Conclusions: Our investigations disclosed a composite transcriptomic architecture of VZV, including the discovery of novel RNA molecules and transcript isoforms, as well as a complex meshwork of transcriptional read-throughs and overlaps. The results represent a substantial advance in the annotation of the VZV transcriptome and in understanding the molecular biology of the herpesviruses in general.

Keywords: Herpesvirus, Varicella zoster virus, Transcriptome, Oxford Nanopore Technologies, Long-read sequencing, RNA editing, nroRNA

Background

Varicella zoster virus (VZV) is one of the nine herpesviruses that infect humans. It is the etiological agent of chickenpox (varicella) caused by primary infection and shingles (zoster) due to reactivation from latency, which is established in the spinal and trigeminal ganglia [1]. VZV belongs to the *Varicellovirus* genus of the

subfamily *Alphaherpesvirinae*, and is closely related to the pseudorabies virus (PRV) a veterinary *Varicellovirus* and to the herpes simplex virus type-1 (HSV-1) belonging to the *Simplexvirus* genus of alphaherpesviruses. The investigation of the VZV pathomechanism is not easy due to its highly cell-associated nature and the lack of animal models, except the SCID mouse model [2, 3]. The VZV virion is ~ 80–120 nm in diameter and is composed of an icosahedral nucleocapsid surrounded by a tegument layer [4]. The outer virion component is an envelope derived from the host cell membrane with

* Correspondence: boldogkoi.zsolt@med.u-szeged.hu

[†]István Prazsák and Norbert Moldován contributed equally to this work.

¹Department of Medical Biology, Faculty of Medicine, University of Szeged, Szeged 6720, Hungary

Full list of author information is available at the end of the article



incorporated viral glycoproteins [5]. The VZV genome is composed of a ~125-kbp double-stranded DNA molecule containing at least 70 annotated open reading frames (ORFs). The viral DNA consists of two unique genomic regions (UL and US), and a pair of inverted repeats (IRs) surrounding the US region. Three genes (ORF62, ORF63 and ORF64) located at the IR regions are therefore in duplicate [6, 7]. VZV encodes six genes (ORFs S/L, 1, 2, 13, 32, and 57), which are not present in HSV [8]. It has been long held that VZV does not express the RNA molecule homologues to the HSV latency-associated transcripts (LAT), however, a recent study reported the identification of VZV latency-associated transcript (VLT) overlapping the ORF61 gene [9]. It has also been thought that VZV lacks the α -TIF protein encoded by the HSV's vp16 gene that activates the transcription of immediate-early (IE) genes during the initial events of the virus life cycle [10]. VZV ORF10 together with ORF4, 62 and 63 are transcriptional transactivators that are all present in the virus tegument [11, 12]. ORF10 is supposed to substitute the function of VP16 in the transactivation of ORF62 (homologue of HSV ICP4), but unlike VP16, it is unable to control the expression of ORF4, ORF63, and ORF61 [13, 14].

The viral replication is primarily controlled by the regulation of transcription, which is carried out through a sequential activation of viral genes. First, the IE genes are expressed, which is followed by the activation of the early (E) and then the late (L) kinetic classes of genes [15]. The IE protein ORF62 is the major transactivator of the viral genes, which recruits the general transcriptional apparatus of the host cell and thereby controls the expression of other viral genes [16, 17]. The VZV E genes encode the proteins required for the DNA replication, while L genes code for the structural elements of the virus. Polycistronism represents an inbuilt characteristic of the herpesvirus genome [18, 19]. However, the herpesvirus multigenic transcripts differ from the bacterial polycistronic transcripts in that the downstream genes on the viral RNA molecules are untranslated. Exceptions to this rule were found in the Kaposi's sarcoma-associated herpesvirus [19, 20] and in HSV-1 [21]. The functional significance of this organization of the herpesviral transcripts has not yet been ascertained. Alternative splicing expands the RNA and protein repertoire with respect of the one gene, one RNA/protein situation. In contrast to the beta and gammaherpesviruses, splicing events are rare among the alphaherpesviruses [22, 23]. In VZV only a few genes have been shown to produce spliced mRNAs so far, which include the ORF0 (also referred as ORFS/L) located at the termini of UL region, the UL15

homologue ORF42/45, ORF50 the glycoprotein M (gM), and the newly discovered, multiple spliced VLT [6, 9, 23–26].

RNA editing includes the adenosine deaminase acting on RNA1 (ADAR1) enzyme that catalyzes the C-6 deamination of adenine (A) to inosine (I). The I is recognized as guanine (G) by the reverse transcriptase (RT) enzyme [27], therefore it can be detected with cDNA sequencing methods. In hyper-edited sites the ADAR1 transforms multiple adjacent As on an RNA strand. A G-enriched neighborhood has been described at the RNA editing sites [28] with an upstream uracil (U) exerting a stabilizing effect on the RNA-ADAR complex [29]. Additionally, RNA hyper-editing has been shown to play a crucial role in the viral life cycle by dodging the host's immune system [30] and also in the control of DNA replication [31].

The translation of eukaryotic mRNAs occurs according to the scanning model, where the 40S ribosomal subunits scans the RNA strand in 5' to 3' direction and initiates the translation at the first AUG they encounter [32]. Mammalian mRNAs contain an essential sequence context for translation initiation, known as the Kozak consensus sequence [33]. A purine at -3 and G at +4 position has the strongest binding effect for the translational machinery, while AUGs with a different context tend to be overlooked by the ribosomal subunits (leaky scanning) [34]. Upstream ORFs in the 5'-untranslated regions (UTRs) of the RNAs have been shown to exert a regulatory effect on the protein synthesis through a process called translation reinitiation [32, 35]. In short upstream (u)ORFs translation reinitiation shows a positive correlation with distance between stop codon of uORF and the AUG of downstream ORF [36].

The hybridization-based microarray and the second-generation short-read sequencing (SRS) techniques have revolutionized genome and transcriptome research, including herpesviruses [37–40]. However, both techniques have limitations compared to long-read sequencing (LRS), including Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) platforms. Microarray and SRS techniques perform poorly in the detection of multi-intronic transcripts, transcript length variants, and polycistronic RNA molecules, as well as transcriptional overlaps. Additionally, the LRS sequencing platforms are capable of determining the transcript ends with base-pair precision, without the need of any supplementary method [41, 42].

Similar to other LRS techniques, ONT cDNA sequencing is afflicted by sample degradation, resulting in false transcriptional start sites (TSSs). This problem can be mitigated by using cap selection. Another source of non-specificity is the presence of nucleotide sequences complementary to the MinION strand switching

oligonucleotide, which can lead to either template switching [43, 44] or false priming [45]. False priming with the oligo(dT) primer can also occur at A-rich regions of the transcripts. These errors contribute to artifactual 3'-ends of the reads.

LRS techniques have already been used in genome and transcriptome studies of viruses [46–49], including herpesviruses [18, 22, 23, 50–52]. These reports have revealed a hidden transcriptome complexity, which included the discovery of long RNA molecules and a large variation of transcript isoforms. ONT and PacBio-based studies have also detected a number of embedded transcripts with in-frame truncated (t)ORFs in several herpesviruses, such as PRV [18, 46], HSV-1 [53], and human cytomegalovirus (HCMV) [22]. Additionally, a large extent of transcriptional read-through and overlaps has also been uncovered by these studies [18, 22, 23]. It has been proposed that the transcriptional overlaps may be the result of a transcriptional interference mechanism playing a role in genome-wide regulation of gene expressions [54].

In this work, we used the ONT MinION sequencing technique to investigate the structural aspects of the polyadenylated VZV transcripts. We report numerous novel transcripts, transcript isoforms, and yet unknown splice events. This study also explores a complex meshwork of transcriptional overlaps. Additionally, we report and analyze a hyper-editing event in a VZV transcript.

Methods

Cells and viral infection

Human primary embryonic lung fibroblast cell line (MRC-5) was obtained from the American Type Culture Collection (ATCC) and grown in DMEM supplemented with antibiotic/antimycotic solution and 10% fetal bovine serum (FBS) at 37 °C in a 5% CO₂ atmosphere. The live attenuated OKA/Merck strain of varicella zoster virus (VZV) was cultured at 37 °C in MRC-5 cell line, the cells were harvested by trypsinization, when the monolayers had displayed specific cytopathic changes. For subsequent propagation of the virus, infected cells were used to inoculate MRC-5 cultures previously grown to full confluence at a ratio of 1:10 infected to uninfected cells. The cultures were then incubated at 37 °C for 5 days, when the cytopathic effect was near 100%.

RNA purification

Total RNA was isolated immediately after collecting the infected cells, using the Nucleospin RNA Kit (Macherey-Nagel) according to the manufacturer's instruction. Briefly, infected cells were collected by centrifugation and the cell membrane was disrupted using the lysis

buffer provided in the kit. Genomic DNA was digested with the RNase-free rDNase solution (supplied with the kit). Samples were eluted in a total volume of 50 µl nuclease-free water. To eliminate residual DNA contamination a subsequent treatment with the TURBO DNA-free Kit (Thermo Fisher Scientific) was conducted. The RNA concentration was measured with a Qubit 2.0 Fluorometer, with the Qubit RNA BR Assay Kit (Thermo Fisher Scientific).

Poly(a) selection

Thirty-five µg of total RNA was pipetted in separate DNA LoBind Eppendorf tubes (Merck). The poly(A)⁺ RNA fraction was isolated from the samples using the Oligotex mRNA Mini Kit (Qiagen). RNA samples were stored at – 80 °C until use.

Oxford Nanopore MinION cDNA sequencing and barcoding

The cDNA library was prepared using the Ligation Sequencing kit (SQK-LSK108; Oxford Nanopore Technologies) according to the modified 1D strand switching cDNA by ligation protocol. In short, the polyA(+) RNA fraction was reverse transcribed using an oligo(d)T-containing primer [(VN)T20 (Bio Basic, Canada)]. The RT reaction was carried out as recommended by the 1D protocol, using SuperScript IV enzyme (Life Technologies); a strand-switching oligo [containing three O-methyl-guanine RNA bases (PCR_Sw_mod_3G; Bio Basic, Canada)] was added to the sample. The cDNA was amplified by using KAPA HiFi DNA Polymerase (Kapa Biosystems) and Ligation Sequencing Kit Primer Mix (part of the 1D Kit) following the ONT 1D Kit's manual. End repair was carried out on the samples using NEBNext End repair / dA-tailing Module (New England Biolabs), which was followed by “barcoding”: the C11 barcode (ONT PCR Barcoding Kit 96; EXP-PBC096) was ligated to the sample according to the 1D PCR barcoding (96) genomic DNA (SQK-LSK108) protocol, Barcode Adapter ligation step. The barcoded- sample was amplified by PCR using KAPA HiFi DNA Polymerase, as well as the C11 PCR barcode according to the 1D PCR barcoding protocol. The PCR product was end-repaired, then it was followed by adapter ligation using the sequencing adapters supplied in the kit and NEB Blunt/TA Ligase Master Mix (New England Biolabs). The cDNA sample was purified between each step using Agencourt AMPure XP magnetic beads (Beckman Coulter). The concentration of cDNA library was determined using a Qubit 2.0 Fluorometer through use of the Qubit (ds)DNA HS Assay Kit (Thermo Fisher Scientific). Samples were loaded on R9.4 SpotON Flow Cells, while the base calling was performed using Albacore v2.1.10.

Cap selection, cDNA synthesis and sequencing

For the precise determination of TSSs, the ONT's 1D strand switching cDNA by ligation protocol was combined with a 5'-cap specific protocol. The cDNA sample was prepared from total RNA using the TeloPrime Full-Length cDNA Amplification Kit (Lexogen). Reverse transcription (RT) reaction was performed at 46 °C for 50 min according to the kit's recommendations by using an oligo(d)T-containing primer (5' -> 3': TCTC AGGCGTTTTTTTTTTTTTTTTTTT). The RT product was purified by using spin column-based method (silica columns are from the Lexogen kit). A double-strand specific ligase enzyme (Lexogen Kit) was used to ligate an adapter to the 5'C of the cap of the RNA. The reaction was carried out at 25 °C, overnight. The sample was washed by applying the silica-column method. The Second-Strand Mix and the Enzyme Mix (both from the TeloPrime Kit) as well as a Verity Thermal Cycler (Applied Biosystems) were used to produce double-stranded cDNAs according to the Kit's guide. In order to generate sufficient amount of cDNAs for MinION library preparation, samples were amplified by endpoint PCR following the TeloPrime Kit's manual. The generation of the sequencing-ready library from this sample is based on the 1D strand switching cDNA by ligation protocol from ONT. The Ligation Sequencing kit (SQK-LSK108, ONT) and NEBNext End repair / dA-tailing Module (New England Biolabs) was used to repair the cDNA ends. This step was followed by the 1D adapter ligation, which was carried out according to the 1D protocol, using the NEB Blunt/TA Ligase Master Mix (New England Biolabs). We used barcoding for the better in silico identification of the transcripts' ends. We found this method useful because it helped the base-pair precision mapping of TSSs. The library was sequenced on an ONT R9.4 SpotON Flow Cells.

Targeted sequencing of putative ONT transcripts

Validation of the TSSs of three 'near the transcription origin' (NTO) transcripts, and the spliced isoforms of the region was carried out using targeted long-read sequencing. The cDNA library was prepared using the ONT 1D protocol, as described in the 'Oxford Nanopore MinION cDNA sequencing and barcoding' section, with the following modifications: we used ribodepletion instead of poly(A) selection, and target-specific primers (Integrated DNA Technologies) were applied instead of oligo(d)T-priming for the cDNA preparation. The target-specific primers used in this study are listed in Table 1.

Data analysis and alignment

Reads resulting from ONT sequencing were aligned to the reference genome of VZV (GeneBank accession: NC_001348.1) and the host cell genome (*Homo sapiens* - GRCh37, BioProject number: PRJNA31257) using

Table 1 Target-specific primers used for the detection of NTO transcripts

Primer name	Sequence
pNTO2	ACTTGCCTGTCGCTCTATCTTCCGGTCAGGATCT
pNTO3	ACTTGCCTGTCGCTCTATCTTCGACTGTTGGAAC
pNTO4	ACTTGCCTGTCGCTCTATCTTCATGTGAGACAACC
pNTO1_3'	ACTTGCCTGTCGCTCTATCTTCTCGTTCGACG
pNTO1-ex2	ACTTGCCTGTCGCTCTATCTTCTCGACGCTG

The primers are composed of a sequence complementary to the MinION cDNA adapter and a target-specific sequence (in italics)

GMAP v2017-04-24 with the default settings [55]. In order to annotate the 5' ends of the transcripts, we used our custom script (<https://github.com/zsolt-balazs/LoR-TIA>). In short: the last 16 nt of the MinION 5' strand switching adapter or the last 16 nt of the cap selection adapter was aligned in a window of -10 nt and +30 nt from the first mapped position of a read using the Smith-Waterman algorithm with a match cost of +2 or -3, a gap opening cost of -3, and a gap extension cost of -2. Read ends with a score below 14 were considered false 5' ends and were discarded. A 5'-end position was considered a TSS if the number of reads starting at this position was significantly higher than at other nucleotides in the region surrounding this start position. For this the Poisson-probability (Poisson $[k_0; \lambda]$) of k_0 read starting at a given nucleotide in the -50 nt to +50 nt window from each local maximum was calculated

with $\lambda = \frac{\sum_{i=-50}^{50} k_i}{101}$. The 5'-ends of the rare long reads were inspected individually using the Integrative Genome Viewer (IGV) [56]. Poly(A) tails were defined using the same algorithm and parameters, by mapping 15 nt of homopolymer As or Ts to the soft-clipped region on the read's end. Read ends with a score below 14 or with more than 5 As/Ts directly upstream their positions were considered artefacts of false priming, and were discarded. Transcription end sites (TESs) were defined using the same criteria for the Poisson-probability as in case of TSSs. The computational pipeline is outlined in Additional file 1.

Reads passing both criteria and present in more than three copies were considered as "certain transcript isoforms". Those with less than three copies or the longest unique reads with questionable 5' ends were considered "uncertain transcript isoforms". Reads with sequencing adapters or poly(A) tails on their both ends were discarded, except for complex transcripts, which were individually inspected using IGV. Reads with a larger than 10 nt difference in their 5' or 3' ends were considered novel length isoforms (L: longer 5' UTR, S: shorter 5'

UTR, AT: *alternative 3' termination*). Short length isoforms harboring a truncated version of the known open reading frame (ORF) were considered novel *putative protein coding* transcripts, and designated as '5'. If multiple putative protein coding transcripts were present, then the one with the longest 5' UTR was designated '5' and its shorter versions were labelled in an ascending order. Transcripts with TESs located within the ORFs of genes (therefore lacking STOP codons) or with TSSs within the coding regions without in-frame ORFs were both considered *non-coding* transcripts.

Multigenic transcripts containing at least two genes standing in opposite were named *complex* transcripts. If a TSS was not obvious at these transcripts, we assumed that they start at the closest upstream annotated TSSs. Splice junctions were accepted if the intron boundary consensus sequences (GT and AG) were present in at least ten sequencing reads and if the frequency of introns was more than 1% at the given region. The relative abundance of the transcript isoforms was calculated by dividing the number of reads with the total number of reads of the dataset; both resulting from the computational pipeline described before. A ± 10 nt-variation was allowed at both the TSS and TES of sequencing read for considering them as certain transcript isoforms. To assess the homology of protein products possibly translated from the ORFs of novel transcript isoforms we used the online BLASTP suite [57], with an expected threshold of 10.

To evaluate the effect of hyper-editing on the secondary structure of RNAs, we used the 'RNAstructure Software' suite [58] with the following parameters: temperature = 310,15 K, Max % Energy Difference = 10, Max Number of Structures = 20, Window Size = 0. To simulate the presence of inosine, we changed the edited adenines to guanine. To assess the secondary structure of the NTO3-ORF62 dsRNA hybrid we used the first 467 bases of the ORF62 labeled ORF62-5' (from genomic position 109,204 to 108,738) as one of our starting sequences, and the whole sequence of NTO3 as the other one. Reads were visualized using the Geneious [59] software suite and IGV. GC-, CCAAT- and TATA-boxes were annotated using the Smith-Waterman algorithm for canonical motifs [60].

Results

Analysis of the VZV transcriptome with third-generation sequencing

In this work, the ONT MinION RNA-Seq technique was used for profiling the genome-wide expression of the lytic VZV transcripts (Fig. 1). Both cap-selected and

non-cap-selected protocols were applied for the analyses of the whole transcriptome. Additionally, targeted sequencing was used with the non-cap-selected protocol for validating the TSS and splice junctions of the NTO transcripts. Sequencing of the non-cap-selected non-targeted sample yielded 619,687 reads of which 66,455 mapped to the VZV reference genome (GeneBank Accession: NC_001348.1) with an average mapped read length of 1349 bp, and an average coverage of 625, while 553,232 reads mapped to the host cell genome (*Homo sapiens* - GRCh37, BioProject number: PRJNA31257).

Targeted sequencing yielded 1,792,059 reads of which 328,112 mapped to the VZV genome, with an average mapped read length of 651 bp, and an average coverage of 1710, and 1,463,947 reads mapped to the host cell genome. The 2.41% of the targeted reads mapping to the VZV genome included the target-specific primers and mapped to the correct position, whereas, 24.112% of the reads included the primer but mapped off-target, while for 73.477% the target-specific primer was missing. These latter two categories could be a result of nonspecific priming or template switching.

Sequencing of the cap-selected samples yielded 6,918,054 reads, of which 614,192 mapped to the viral genome, with an average coverage of 1200, and 6,303,862 reads mapping to the host cell. This latter technique performed poorly in VZV, which was indicated by the short average aligned read length (294 bp) (Fig. 2 panel a). Intriguingly, we obtained the same poor result with other herpesviruses, including PRV [23] HSV-1 (our unpublished results), however, the cap-selection technique performed very well in the baculovirus *Autographa californica* multiple nucleopolyhedrosis virus (AcMNPV), and vaccinia virus (our unpublished results). These results together with an analysis of the nucleotide distribution at the vicinity of the 5'-ends in the cap-selected datasets (Fig. 2 panel b) demonstrate that the read length of cap-selected RNA-s is not determined by the GC-content (PRV: 73%, HSV-1: 68%, VZV: 46%), but by a yet unknown factor.

In total, we detected 1124 5'-ends and 255 3'-ends in the non-cap-selected dataset, and 1428 5'-ends and 279 3'-ends in the cap-selected dataset (Additional file 2: Table S1-S4). The number of 5'-ends qualifying as a putative TSS was 10.86% of the total 5'-end positions, while 32.95% of the total 3'-end positions turned out as TES (Table 2). We excluded 49 5'-ends and 16 3'-ends from the non-cap-selected and 21 5'-ends and 16 3'-ends from the cap-selected datasets because they proved to be the products of false priming.

Earlier results with primer extension, S1 nuclease assay and Illumina sequencing determined the transcript ends of some of the RNA molecules of VZV. Using the ONT sequencing platform, we confirmed 18 previously known TSSs and nine TESs. Additionally, we annotated

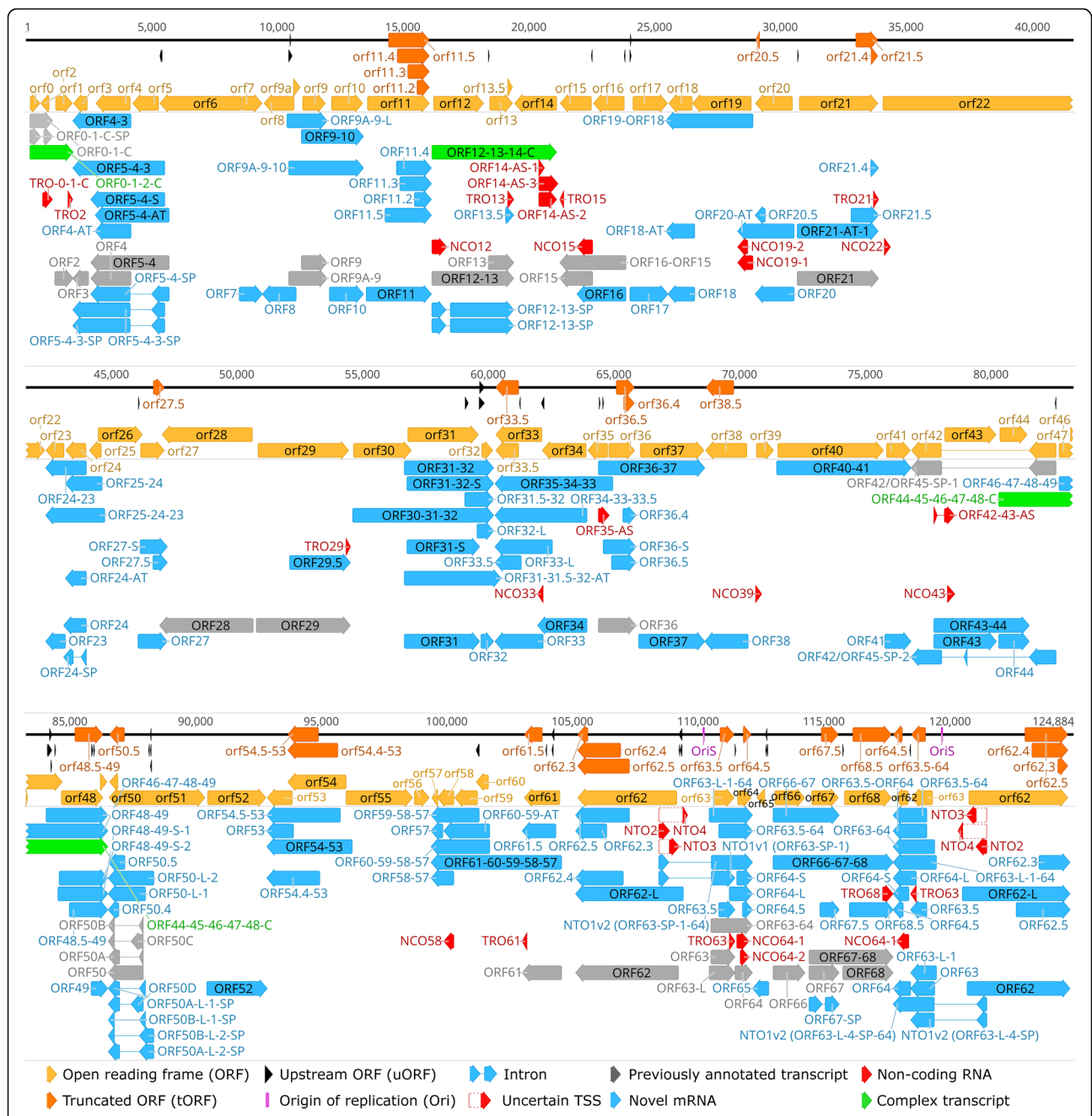


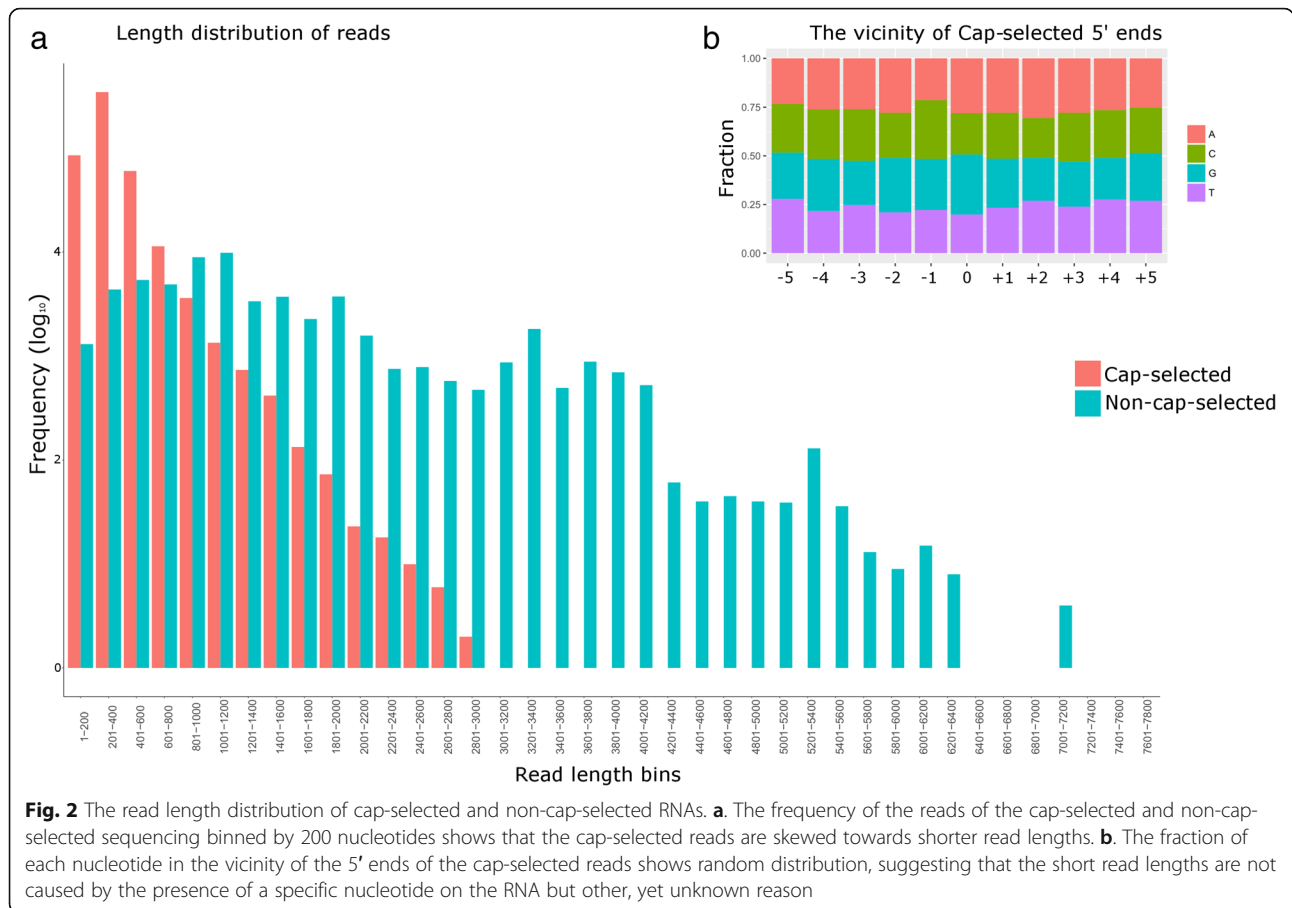
Fig. 1 The VZV transcriptome. The transcriptome of the varicella zoster virus was analyzed using long-read cDNA sequencing. The reads were annotated using our custom script (<https://github.com/zsolt-balazs/LORTIA>). Transcriptional start sites and transcriptional end sites of the annotated transcripts are the 5' and 3'-ends with significantly higher abundance than any other read end in their ± 50 vicinity, according to Poisson-probability, and those that are not a result of strand switching or false priming

124 new putative TSSs and 71 TESs listed in Additional file 2: Table S5. The sequences upstream of the TSSs and TESs were analyzed in silico to detect putative GC-, CCAAT- and TATA-box motifs, and Poly(A) signals (PASS), by motif searching (Table 3).

Twenty-two and five transcript isoforms were identified with the non-cap-selected, and cap-selected dataset,

respectively with a relative abundance higher than 1% (Additional file 2: Table S5).

In this work, we detected an enrichment of As and Us in an interval of up to - 10 nt upstream of the putative TES, and an abundant GT-rich region downstream in the + 10 nt interval (Fig. 3). We annotated altogether 240 novel transcripts of which 143 were



confirmed by at least three reads, additionally, we mapped the TSS and TES of 37 previously detected transcripts lacking former mapping of RNA end positions (Additional file 2: Table S5-S6).

Putative mRNAs

Transcripts embedded into larger RNA molecules are easily detected by the LRS techniques. If such a transcript contains an in-frame ORF shorter [truncated (t)ORF] than that of the host ORF, it can be considered as a putative mRNA potentially coding for an N-terminally abridged polypeptide. Earlier in silico

approaches have annotated the VZV ORFs [6, 7] one of which (the orf33.5) is a tORF. In this study, we report the identification of 25 embedded viral transcripts containing tORFs. We could identify promoters for 10 of these transcripts, located at a mean distance of 93.85 ± 32.18 nt (mean \pm SD) from their TSSs (Additional file 2: Table S7). Three of these transcripts (ORF13.5, ORF54.5–53 and ORF62.3) contain strong Kozak consensus sequences near their AUGs, while 19 have at least one of the two essential nucleotides present on their -3 or +4 positions. Eighteen of the possibly protein coding VZV transcripts contain a PAS at a 19.44 ± 8.91 (mean \pm SD) distance upstream their TESs (Additional file 2: Table S7).

Table 2 The number of read ends following each step of filtering

Filter	Non-cap-selected		Cap-selected	
	5' ends	3' ends	5' ends	3' ends
None	1124	255	1428	279
Peak analysis and Poisson probability	151	92	202	116
Template switching/false priming	102	76	181	100

Determination of the ends of mRNAs with annotated ORFs

In this work, we determined the TSSs and TESs of 25 mRNAs whose ORFs had earlier been described (Fig. 1, Additional file 2: Table S5). The TSS and TES of these monocistronic mRNAs were not characterized before by any other means. Twelve of these transcripts have a promoter sequence 51.57 ± 33.31 (mean \pm SD) upstream

Table 3 Putative promoter sequences and Poly(A) signals (PASs) of the VZV transcripts

	Search range (nt)	Nr. of motifs	Nr. of transcripts	Mean Distance	SD
GC-box	-150 to 0	41	55	93.28	33
CCAAT-box	-150 to -40	19	27	93.91	23
TATA-box	-35 to -25	26	44	31.67	1,5
PAS	-50 to 0	61	169	16.11	8.86

The search range position 0 is the position of the TSS in GC-, CCAAT- and TATA-boxes and is the position of the TES for PAS. The distance was calculated between the position of the TSS/ TES and the last nucleotide of the motif

their TSSs, and 19 have a canonical PAS upstream their TESs at a mean distance of 14.96 ± 11.59 (mean \pm SD).

Non-coding transcripts

The ncRNAs are transcripts without the ability to encode proteins. They can overlap the coding sequences of the genes in the same orientation, or in the opposite orientation [antisense (as)RNAs], or they can be located at the intergenic regions. The 5'-truncated (TR) transcripts have their own promoters but share the PASs and TESs with the 'host' transcript, while the 3'-truncated (NC) RNAs are controlled by the same promoters as the canonical transcript in which they are embedded but have no in-frame ORFs. Twenty-three of the novel non-coding transcripts are long non-coding (lnc)RNAs exceeding 200 bps in length per definition, while five of them are small non-coding (snc)RNAs with a size below 200 bps.

In this work, we identified 18 embedded ncRNAs, of which eight were 5'-truncated, while ten were 3'-truncated transcripts (Additional file 2: Table S5). We also detected one intergenic and seven antisense ncRNAs, all of them being controlled by their own promoters. In total, 17 of the novel ncRNAs contain a canonical promoter sequence 60.4 ± 35.19 nt (mean \pm SD) upstream their TSSs, while 21 have a PAS at a 16.35 ± 8.87 (mean \pm SD) distance of their TESs. ORF42-43-AS and ORF35-AS overlap multiple mRNAs. ORF42-43-AS stands in antisense orientation with respect of ORF42/ORF45-SP-1 and ORF42/ORF45-SP-2,

while in tail-to head orientation with ORF43 and ORF43-44. ORF35-AS stands in antisense orientation with the polycistronic transcript ORF35-34-33 and in tail-to-head orientation with ORF36 and ORF36-S and ORF36-37 (Fig. 1). The LAT RNA has been described in every member of the alpha-herpesvirus subfamily [61, 62], including VZV [9]. We confirmed the existence of four previously detected lytic isoforms of VLT (VLT_{ly}) of which two are TSS isoforms, one is a splice variant and one is both a TSS isoform and a splice variant (Fig. 4).

5'- and 3'-UTR isoforms

The 5'-UTR isoforms (TSS variants) start upstream or downstream of the TSS of the earlier annotated transcripts, and their expression is regulated by their own promoters, while 3'-UTR isoforms (TES variants) contain distinct PAS and their polyadenylation occurs upstream or downstream of the TES of the main transcript. In this report, we detected 18 novel 5'-UTR length variants, 7 being shorter and 11 being longer than the earlier annotated transcript isoforms. We found canonical promoter sequences in 10 of the 5'-UTR isoforms at a distance of 94.93 ± 38.58 (mean \pm SD). Additionally, we detected eight 3'-UTR length isoform, all with a canonical PAS 22 ± 7.65 (mean \pm SD) upstream their TESs. An intriguing finding is the putative TSS at the genome position +4 belonging to the rare transcripts ORF0-1-L-C and ORF0-1-2-L-C at the extreme termini of UL region (TR₁), which suggests the existence of a promoter

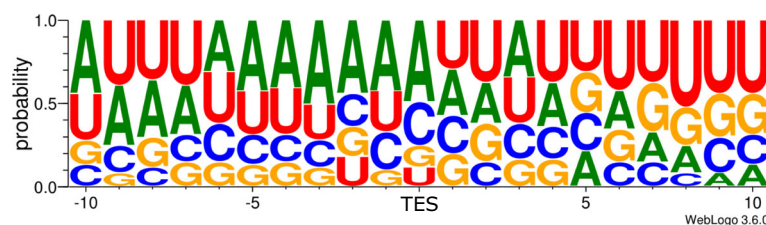


Fig. 3 The frequency of nucleotides in the vicinity of transcriptional end sites (TES). The ± 10 nt vicinity of $N = 61$ unique TESs was analyzed using the online WebLogo v3.6.0 service. The Weblogo shows an enrichment of A and U bases upstream, while an enrichment of G and U bases downstream of the TES. This pattern is akin with the sequence surroundings of mammalian TESs

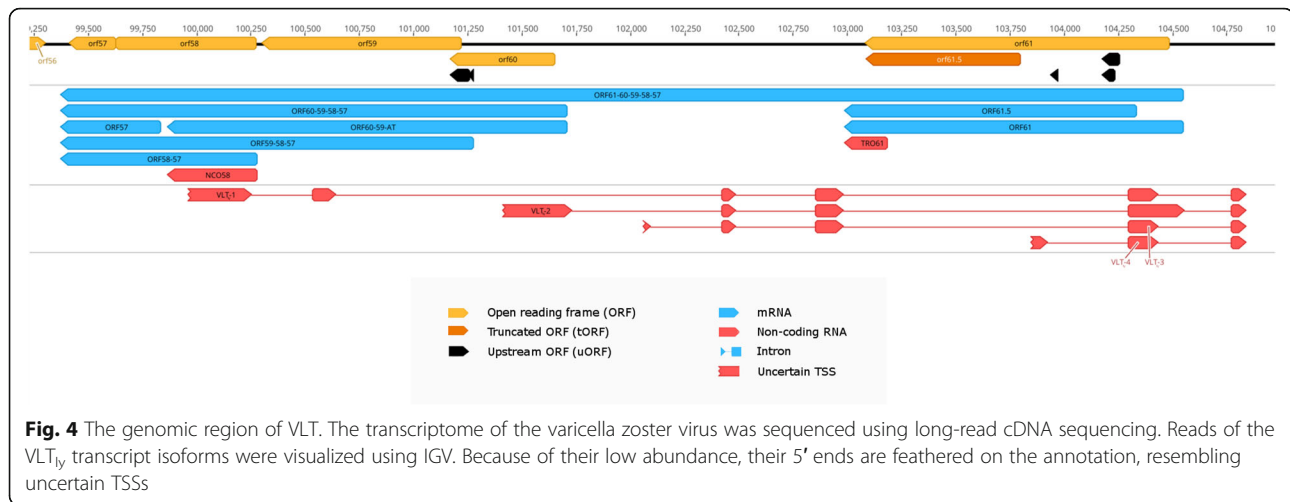


Fig. 4 The genomic region of VLT. The transcriptome of the varicella zoster virus was sequenced using long-read cDNA sequencing. Reads of the VLT_{ly} transcript isoforms were visualized using IGV. Because of their low abundance, their 5' ends are feathered on the annotation, resembling uncertain TSSs

located on the other terminal repeat (TR_S) of the genome. This putative promoter is supposed to be active in the circular genome.

Splice sites and splice isoforms

Reverse transcription can produce false introns between repetitive sequences of the template RNA due to the phenomenon of template switching. In order to exclude these artefacts, we removed sequencing reads with low abundance ($\leq 1\%$) and with a repeat of more than six nucleotides next to one of the splice junctions. From the initial set of 10,064 unique splice acceptor and donor site candidates 24 matched our criteria resulting in a total of 16 splice isoforms being above the 1% intron depth level of acceptance.

We detected two novel splice variants encoded by the ORF42/45 gene. Furthermore, we identified twelve novel splice sites and confirmed the existence of nine previously described spliced transcripts, all with a consensus GT at the splice donor site and AG at the splice acceptor site.

ORF63, homologue to the HSV-1 and the PRV *us1* gene, is one of the main regulator of VZV transcription. In this work, we discovered ten novel splice variants of ORF63. Similarly to the HSV-1 US1 mRNA, NTO1v1 and the NTO1v3 harbors one intron in its 5' UTR, while the NTO1v4 is spliced twice just as the PRV US1 mRNA [50, 53] (Fig. 5). Ten splice junctions of ORF63 splice variants coincide with those of the VLT_{ly} splice variants [9], thus they were labeled as VLT-ORF63-C (Fig. 6). These were confirmed using targeted sequencing. It is noted that one of the splice donor sites, present in both NTO1v1 and NTO1v2 is GC, which differ from the canonical splice donor sequence.

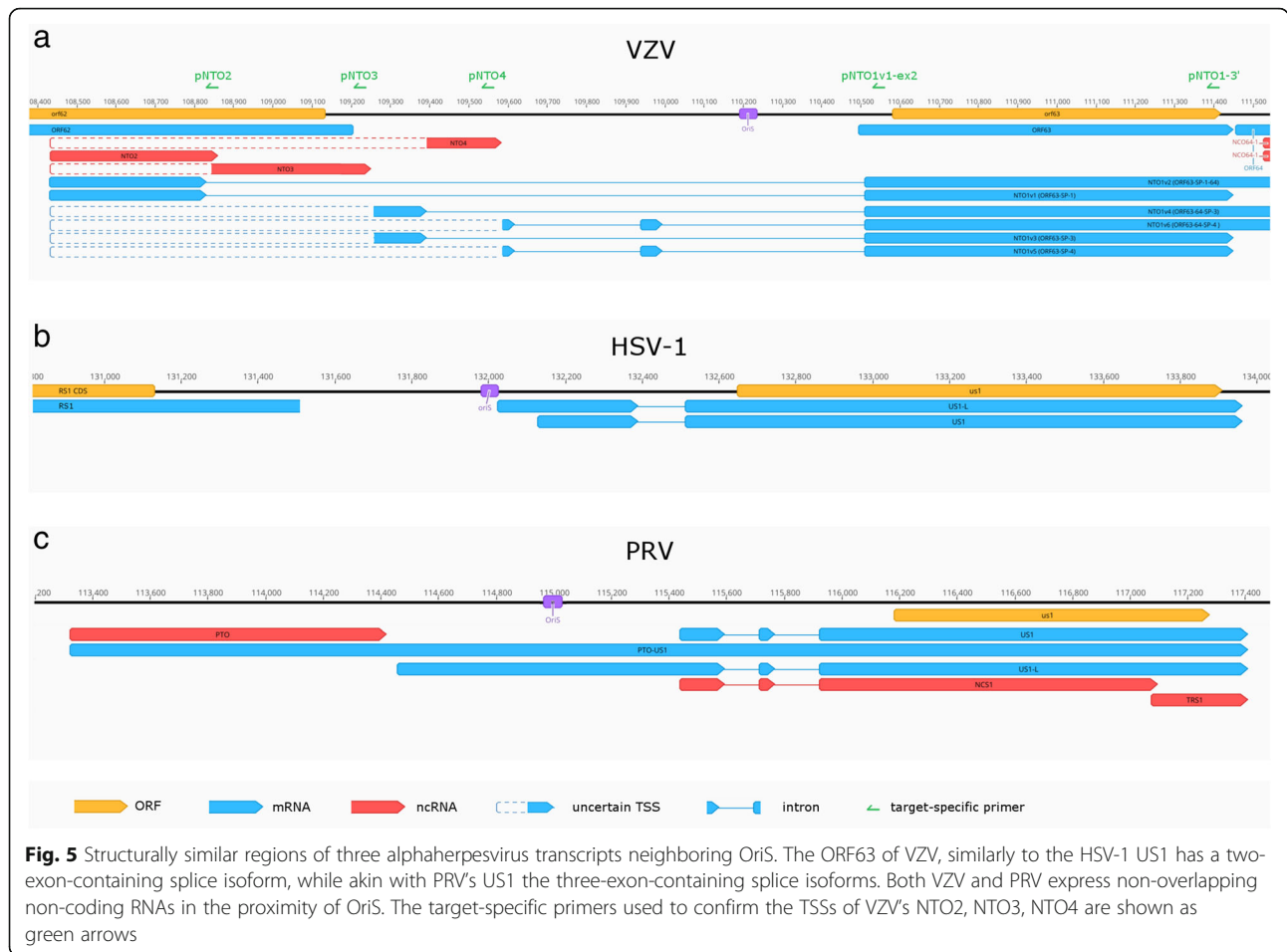
We detected all of the three ORF50 splice isoforms previously described [24], but ORF50C occurs in

relatively low abundance, below our acceptance threshold. We detected another splice variant in low abundance in the ORF50 cluster, and named it ORF50D (Table 4).

In four isoforms, splicing affects the sizes of the ORFs, all producing hypothetical proteins truncated near the N-terminal. Roviš and colleagues [63] have identified the majority of VZV proteins using monoclonal antibodies. In some cases, lower-size bands appeared in their Western blots, which they explained to be the result of putative post-translational modification, or proteolytic cleavage. We propose splicing as possible option for these nonspecific protein isoforms, listed in Table 5.

Near-replication-origin (nro)RNAs – A novel class of transcripts

In our earlier work, we reported [64] that PRV expresses transcripts located near the replication origins (Oris): the CTO family (including CTO-S, CTO-S-AT, CTO-M and CTO-L) at the OriL and the PTOs (PTO and PTO-US1) [18] at the OriS. Expression of nroRNAs has also been described in HSV-1 (OriS-RNA: [65]) and HCMV (OriLyt: [66]; RNA4.9: [67]). The VZV genome contains exclusively OriS, and lacks the replication origin at the UL segment. We identified nine nroRNAs starting in the proximity of VZV OriS. The NTO1v1, NTO1v3 and NTO1v5 are the spliced long TSS variants of ORF63 while the NTO1v2, NTO1v4 and NTO1v6 are the spliced long TSS variant of ORF64 (Fig. 5 panel a). NTO2 starts at the same TSS as NTO1v1 but terminates 30 bp downstream of its splice donor site. The 5' end of the reads of NTO3 and NTO4 are positioned downstream of NTO2. These transcripts were present in our cap-selected but not in the non-cap-selected sample presumably because of their low abundance. In order to

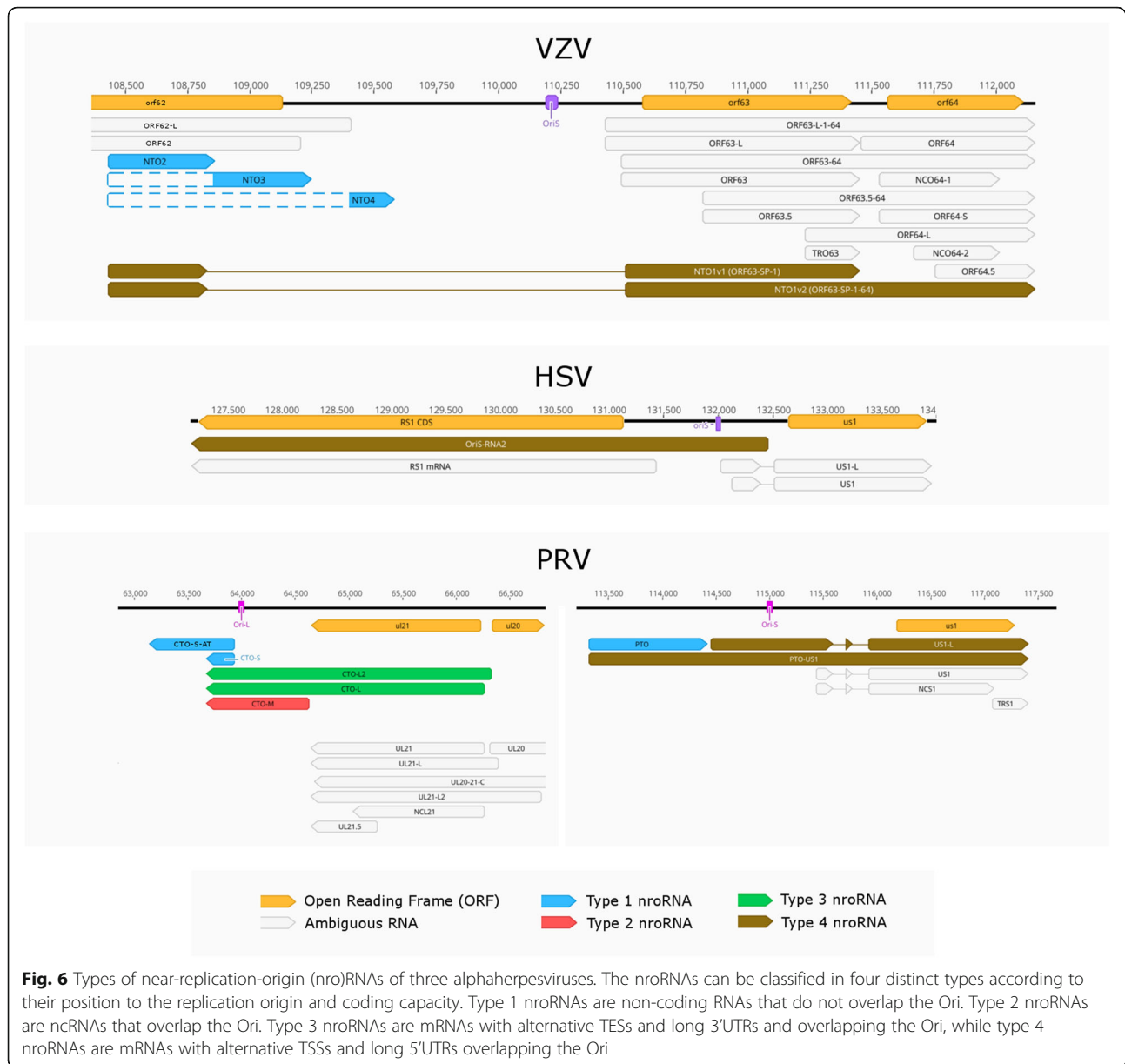


augment the detection of these transcripts, we used target-specific primers (Fig. 5 panel a.) for sequencing library preparation, followed by MinION cDNA sequencing. The starting positions detected in the cap-selected sample were not confirmed by the targeted sequencing of these transcripts, thus we hypothesize that NTO3 and NTO4 starts at the closest upstream confirmed TSS, that of NTO1v1. The arrangement of the NTO transcripts is similar to that observed in PRV, where the TSS of the PTO-US1 (positionally similar to the NTO1v1) is located at the same genetic locus as PTO (locationally similar to the NTO2, NTO3 and NTO4, but not homologous) (Fig. 5 panels a and c). Our in silico analysis revealed that the poly(A) cleavage site of NTO2 is located within a canonical CA element, with a GU-rich region starting 13 nt downstream from the TES. The NTO3 has a CTTAAA poly(A) signal [68] starting 20 nt upstream from its TES, with a cleavage site in a homopolymer C run and with a consensus GU-rich region starting 25 nt downstream from the TES. The NTO4 has an ATATAA poly(A) signal [69] starting 24 nt upstream the TES, the

cleavage site being marked by a consensus CA signal, followed by a GU-rich region 33 nt downstream. No enrichment of adenines was observed in the genomic region of the TES-s in any of the three NTO transcripts, which would be the source of strand switching or false priming. Based on these results, we can distinguish four types of nroRNAs: (1) ncRNAs that do not overlap the Ori (such as CTO-S, CTO-S-AT, PTO, as well as NTO2, NTO3 and NTO4); (2) ncRNAs that do overlap the Ori (such as CTO-M), (3) mRNA isoforms with very long alternative TES (such as CTO-L and CTO-L2); and (4) mRNA isoforms with very long TSS variant [such as PTO-US1, US1-L (PRV), OriS-RNA2 (HSV-1), and the now discovered NTO1 isoforms] (Fig. 6).

Polycistronic and complex transcripts

A major issue of SRS, as well as microarray and quantitative PCR approaches is that they have severe limitations in distinguishing between mono- and polycistronic transcripts. In contrast, LRS sequencing is suitable for making this distinction, and it is particularly superior in the



detection of low abundance multi-genic transcripts. In this work, we identified 33 novel multigenic RNAs, including 29 polycistronic and 4 complex transcripts. Complex transcripts are multigenic RNAs that contain one or more genes in opposite orientations. Antisense sequences on the complex transcripts are unable to encode proteins. We also detected ten complex transcripts in low abundance in the region of VLT, seven of which are co-terminal with ORF63 and three with ORF64. These transcripts overlap with several oppositely oriented coding sequences and are spliced in a similar manner as the VLT_{ly} isoforms (Fig. 7). In silico analysis detected an in-frame ORF incorporating the coding sequence of ORF63 (it has an upstream AUG). This results in VLT_{ly}-ORF63-C1,

VLT_{ly}-ORF63-C4, VLT_{ly}-ORF63-C5, VLT_{ly}-ORF63-C6 and VLT_{ly}-ORF63-64-C1 encoding hypothetical proteins whose N-terminal is longer with 88 amino acids (aa) than the one encoded by the *orf63* gene, while the VLT_{ly}-ORF63-C2; VLT_{ly}-ORF63-C3 and the VLT_{ly}-ORF63-64-C2 encoding hypothetical proteins with 179 aa longer than those coded by *orf63* (Fig. 7). The N-terminal overhangs of the putative proteins show no homology with any other known proteins in online databases.

Transcriptional overlaps

Transcripts can overlap each other in parallel (tandem or head-to-tail), convergent (tail-to-tail) or divergent (head-to-head) manner (Fig. 8). RNAs identified and

Table 4 Splice junction sites of the VZV transcriptome

Intron Start	Intron End	Strand	Intron Length	Donor Site	Acceptor Site	Accepted as novel	Intron depth	Transcript name	References
585	714	+	129	GT	AG		45,62%	ORF0-1-C-SP	[25]
5004	4155	-	849	GT	AG	✓	2,90%	ORF5-4-SP	
16,712	16,886	+	174	GT	AG	✓	5,04%	ORF12-13-SP	
43,826	43,505	-	321	GT	AG	✓	4,11%	ORF24-SP	
43,950	43,863	-	87	GT	AG	✓	1,40%	ORF24-SP-2	
61,790	63,564	+	1774	GT	AG	✓	2,40%	-	
77,870	78,148	+	278	GT	AG	✓	11,29%	ORF42-43-AS	
78,938	78,039	-	899	GT	AG	✓	38,88%	ORF42/ORF45-SP-2	
81,537	78,039	-	3498	GT	AG		100,00%	ORF42/45-SP-1	[6]
81,537	79,053	-	2484	GT	AG	✓	12,72%	ORF42/ORF45-SP-2	
87,759	86,782	-	977	GT	AG		4,15%	ORF50B	[24]
87,759	86,982	-	777	GT	AG		2,33%	ORF50A	[24]
87,436	86,782	-	654	GT	AG		0,62%	ORF50C	[24]
86,982	87,436	-	454	GT	AG		0,33%	ORF50D	
101,728	102,420	+	692	GT	AG		87,50%	VLT _{ly}	[9]
102,484	102,853	+	369	GT	AG		100,00%	VLT _{ly}	[9]
102,983	103,827	+	844	GT	AG		2,42%	VLT _{ly}	[9]
103,924	104,293	+	369	GT	AG		3,99%	VLT _{ly}	[9]
104,433	104,768	+	335	GT	AG		5,22%	VLT _{ly}	[9]
104,824	110,509	+	5685	GT	AG	✓	2,18%	ORF63-SP2-C	
108,830	110,509	+	1679	GC	AG	✓	11,12%	NTO1v1 (ORF63-SP-1)	
114,920	115,050	+	130	GT	AG	✓	1,04%	ORF67-SP	
121,067	119,388	-	1679	GC	AG	✓	10,65%	NTO1v1 (ORF63-SP-1)	

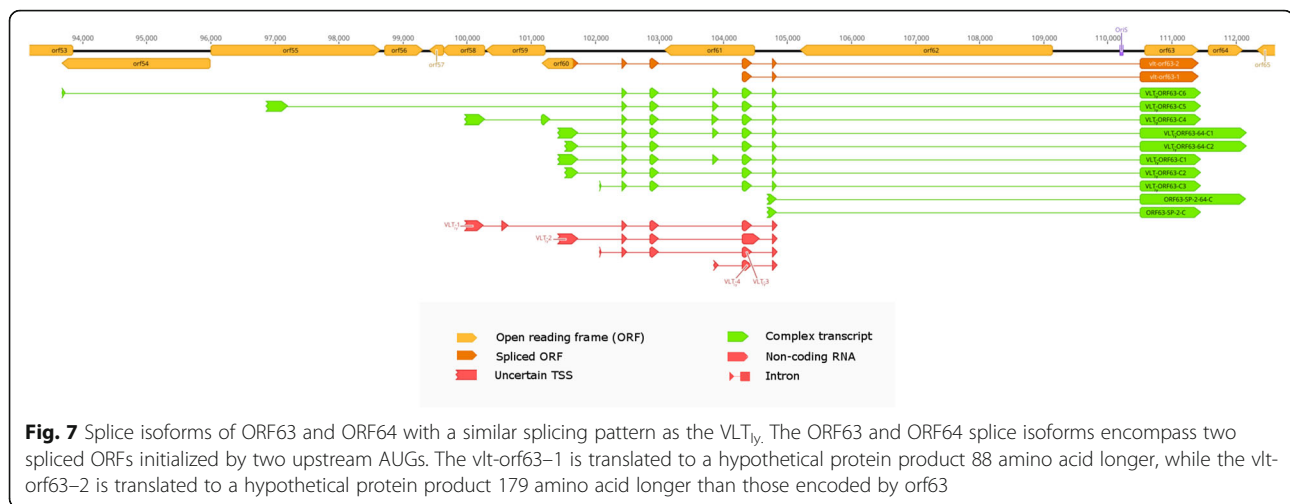
The splice junctions listed in this table passed our criteria of intron detection, with the exception of ORF50C, which is listed as a confirmation, and ORF50D, which is a novel combination of previously known splice sites. Other rare splice variants below the 1% spliced read limit are listed in Additional file 2: Table S6. The TSS and TES belonging to the transcript with splice sites in row 7 could not be determined with certainty

annotated with a certain TSS and TES in this work form a total of 481 overlaps, of which 15 are head-to-head, 450 are head-to-tail and 16 are tail-to-tail (Additional file 2: Table S8). The overlaps can be full or partial. Full overlaps can be formed between the RNA molecules encoded by polycistronic transcription units (Fig. 8 panel a.), between embedded and host mRNA molecules,

between mRNAs and 3' as well as 5' truncated transcripts, between the TSS and TES isoforms, etc. Partial overlaps can be formed between every transcript type. An overlap is 'hard' if two genes can only produce overlapping transcripts, or 'soft' if both overlapping and non-overlapping transcripts are formed. The soft overlap can be the result of alternative promoter usage (TSS

Table 5 The proteins and hypothetical proteins produced by spliced transcripts

Transcript	Host protein length (aa)	Spliced protein length (aa)	Intron position	Splice donor	Splice acceptor
ORF5-4-SP	ORF4: 452	452	5' UTR	4141	2783
ORF12-13-SP	ORF12: 661	233	in frame, 5'truncated	16,214	17,088
ORF24-SP	ORF24: 269	137	in frame	44,021	43,286
ORF42/45-SP-2	ORF42/45: 747	380	in frame, 5'truncated	82,593	78,969
ORF50D	ORF50: 435	-	-	-	-
NTO1v1 (ORF63-SP-1)	ORF63: 278	278	5' UTR	110,581	111,417
ORF63-SP-2-C	ORF63: 278	278	5' UTR	110,581	111,417
ORF67-SP	354	311	in frame, 5'truncated	114,496	115,559



isoforms) or transcriptional read-through (TES isoforms). An example for the latter case is the ORF17 and ORF18, which produce non-overlapping transcripts, but the TES isoform of ORF18 (ORF18-AT), with its additional 89 bps in the 3'UTR, overlaps with ORF17 in a tail-to-tail manner (Fig. 8 panel c).

Upstream ORFs

Using *in silico* methods, we detected 44 potential (u)ORFs on the 5'-UTRs of 81 VZV transcripts (Additional file 2: Table S9). Five of the longer TSS variants contain uORFs, while their shorter isoform does not. We have previously described this phenomenon in HCMV [52] and PRV [23]. The average size of an uORF was 54.9 ± 44.64 nt (mean \pm SD, median = 35 nt), while the average distance of the uORF stop codon from the protein coding ORF's AUG was 174.84 ± 143.58 nt (mean \pm SD, median = 110 nt). This space between the two uORF and the canonical ORF is enough for a potential reinitiation event. We identified Kozak consensus sequences in five uORFs (Additional file 2: Table S9).

RNA editing

The sequencing reads of NTO3 transcript show a very high frequency of A to G substitution, which is not present in the overlapping reads of other transcripts. We found that 58% of all substitutions are A \rightarrow G (Fig. 9 panel a and b) in the reads of NTO3, which is significantly higher than the 12.98% in the overlapping transcripts in the same region ($p < 0.0001$, Fisher's exact test) (Fig. 9 panel c), making 22.07% of all As of the transcript edited. The targeted sequencing dataset revealed a similar A \rightarrow G substitution pattern in the NTO3 overlapping reads as the cap-selected data. This suggest a hyper-editing event in NTO3. Using the MEME software suite [70] we could detect a slight enrichment of the GU dinucleotide preceding

the editing site resulting in a GUAG motif (the editing site is underlined) (Fig. 9 panel d).

Using the 'RNAstructure' software suite [58], we predicted the secondary structure with the lowest free energy of the NTO3 ssRNA (Additional file 3 a and b) and of the NTO3-ORF62-5' dsRNA hybrid, using both the unedited and edited forms of the asRNA (Additional file 3 c and d). When forming an intramolecular secondary structure, the unedited form has a higher free energy state than the edited form (-143.2 kcal/mol compared to -169.4 kcal/mol), which suggests that hyper-editing confers thermodynamic stability to the secondary structure of the RNA. Twelve of the 17 Is may aid in the formation of stem structures, while in the unedited RNA only four of the As in the same position have a complementing nucleotide. Contrarily, the secondary structure of the dsRNA formed by the edited NTO3 and the ORF62 is slightly less stable than that formed by the unedited form (-818.7 kcal/mol compared to -822.9 kcal/mol), but allows the formation of identical secondary structures.

Discussion

Until now, the VZV transcriptome have been analyzed by Northern blot, primer extension, microarray and Illumina sequencing [17, 71–78]. These techniques have generated useful data, but they have limitations to provide a comprehensive list of the VZV transcripts. In this work, we used the ONT MinION LRS technique for the investigation of the poly(A) + fraction of the VZV transcriptome. Our results increased the number of full length transcripts in VZV by one order of magnitude. We identified 108 novel transcripts, including novel mRNA molecules, monocistronic transcripts, UTR isoforms, as well as multigenic transcripts. Novel splice sites and splice variants were also detected. Additionally, we discovered five sncRNAs and 21 lncRNAs and

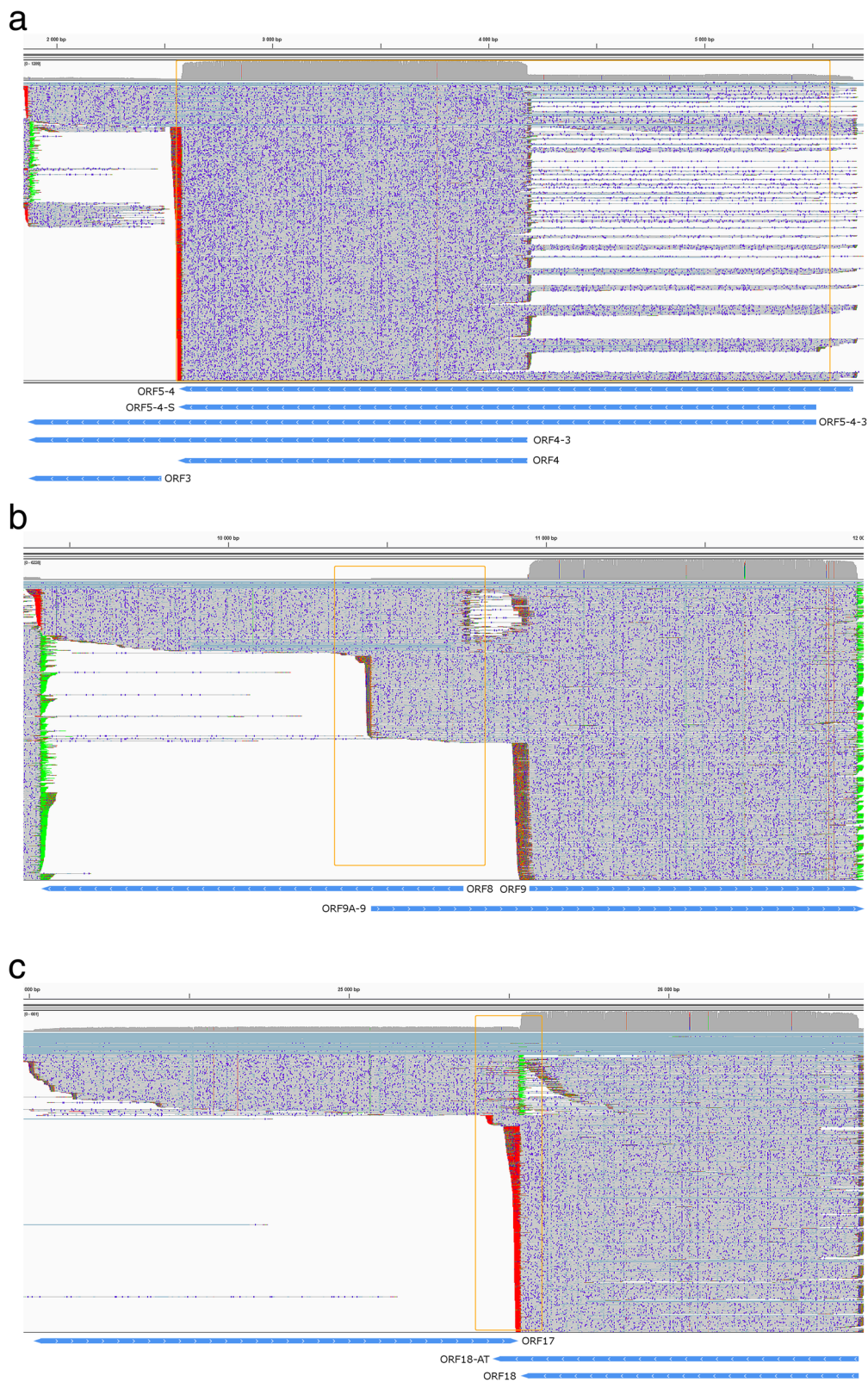


Fig. 8 Types of overlaps between VZV transcripts. Reads were visualized using IGV. Blue arrow-rectangles represent transcript annotations. The overlaps are framed in orange. **a.** Parallel overlaps of the ORF5–4-3 cluster. **b.** Divergent overlaps of the ORF8 and ORF9 and ORF9A-9 transcripts. **c.** Convergent overlaps of the ORF17 and ORF18-AT transcripts

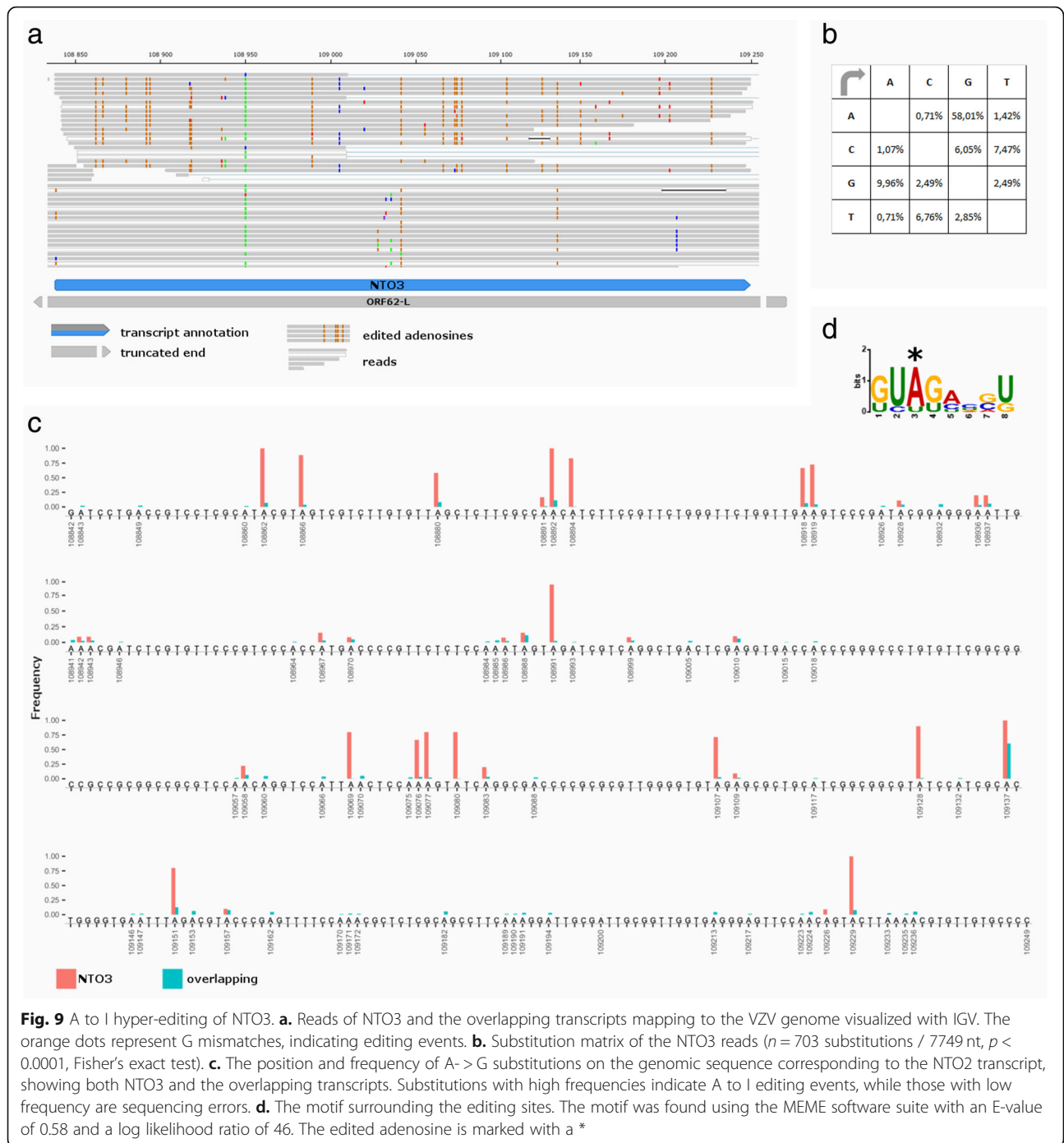


Fig. 9 A to I hyper-editing of NTO3. **a.** Reads of NTO3 and the overlapping transcripts mapping to the VZV genome visualized with IGV. The orange dots represent G mismatches, indicating editing events. **b.** Substitution matrix of the NTO3 reads ($n = 703$ substitutions / 7749 nt, $p < 0.0001$, Fisher's exact test). **c.** The position and frequency of A → G substitutions on the genomic sequence corresponding to the NTO2 transcript, showing both NTO3 and the overlapping transcripts. Substitutions with high frequencies indicate A to I editing events, while those with low frequency are sequencing errors. **d.** The motif surrounding the editing sites. The motif was found using the MEME software suite with an E-value of 0.58 and a log likelihood ratio of 46. The edited adenosine is marked with a *

annotated their TSSs and TESs with base pair precision. The precise annotation of the viral transcriptome is essential for understanding the genetic regulation of VZV. Different TSS isoforms can be regulated by alternative promoters, like for the *ul44* gene of HCMV, where two transcript isoforms initiate downstream of a canonical TATA promoter and are expressed at early time points of the viral infection, while the third promoter is non-canonical, and

the mRNA isoform starting downstream from it is transcribed at a later phase [79]. Furthermore, alternative TESs may influence RNA metabolism by allowing or forbidding the binding of micro (mi)RNAs to the viral RNA [80] or by facilitating or preventing deadenylation [81].

The enrichment of As and Us upstream of PASs in mammalian systems is well-established [82–84]. These homopolymer A stretches may also cause false priming

and template switching events during reverse transcription, which results in false TESs. In our work, we excluded these artefacts by the use of our analysis pipeline, and demonstrated the verity of the annotated TESs by the presence of the canonical GU-rich region in the +10 interval downstream of TESs [85].

Similar to other herpesviruses [22, 53], this study also revealed a complex meshwork of transcriptional read-throughs and overlaps. It has been known earlier that the polycistronic transcription units include co-terminal multigenic RNA molecules, which represent a large extent of parallel overlaps along the entire viral genome. We can raise the question as to whether the downstream sequences of multicistronic transcripts have any other functions besides their possible coding potential, or if they are mere random read-through products representing transcriptional noise. We can address the same question for the convergently and divergently overlapping RNA segments, and also for the alternative transcriptional overlaps. We have put forward a hypothesis that explains the potential role of this phenomenon which is based on transcriptional interaction between the RNA polymerase molecules at the overlapping regions. Since essentially every gene produces overlapping transcripts and therefore the pairwise interactions can spread along the genome thereby forming a transcriptional interference network (TIN), which alongside the promoter-transcription factor system determines the global gene expression pattern of the viral DNA [54].

Understanding the factors influencing the expression of *orf63* is of particular interest as its transcript is translated predominantly during latent infections of human ganglia [9, 86]. It has been previously shown that *orf63* has no trans-regulatory effect on the expression of *orf62*, which encodes the major immediate-early transactivator protein of VZV [87]. Nevertheless, deletion of *orf63* has demonstrated that it is indispensable for VZV replication, and its product has a protein interaction site with the translated *orf62* [86]. Other experiments have shown that deletion of *orf63* increases the expression of *orf62* [88], suggesting that there is a link between the regulation of the two genes. A possible explanation may be the transcriptional interference caused by the head-to-head overlaps between ORF62 and the isoforms of ORF63 (NTO1v1 and NTO1v2) in the late phase of the viral life cycle.

Polycistronic (especially bicistronic) transcripts are also common in eukaryotic organisms, but their generation is explained by trans-splicing and not by transcriptional read-through [89]. However, unprocessed transcripts are difficult to detect because of their short existence, therefore, it cannot be excluded that these transcripts are the result of a transcription read-through

mechanism. The predominant occurrence of adjacent genes in the chimeric transcripts suggests that transcriptional read-through followed by cis-splicing may be the case, that is, the existence of a large extent of transcriptional read-throughs may be not restricted to the herpesviruses but they may represent a general phenomenon. Furthermore, the antisense RNAs produced from their own promoters or by transcriptional read-throughs may hybridize with their sense counterparts thereby initiating RNA interference [90]. Very long complex transcripts form a distinct category among multigenic transcripts because of their oppositely oriented ORFs and increased size. Similarly to other herpesviruses [22, 53] they are present in very low abundance in VZV, however their existence is ambiguous, as they can be formed during reverse transcription by template switching [91].

It has been previously thought that the OriS of VZV lacks any overlapping transcripts [6]. Although Davison and colleagues (1986) hypothesized the existence of non-coding RNA in the intergenic region between *orf62* and *orf63* [6]. We detected several transcripts overlapping OriS, and a small non-coding RNA (the NCO3) between the before-mentioned two genes. This region of the viral transcriptome is structurally similar to the PRV transcriptome. In the vicinity of OriS, we identified two additional novel non-coding transcripts (NTO2 and NTO4) and two 5' elongated and spliced version of ORF63. These nroRNAs are supposed to play a role in the regulation of the viral replication [92] through the interplay between the transcriptional and replication apparatuses [54]. Specifically, the role of the transcription of these RNA molecules might be to interfere with the replication machinery, in order to force the replication fork to an unidirectional progression [64]. Furthermore, the intergenic region between *orf62* and *orf63* contains binding sites for T cell proteins including heat shock factor 1 (HSF-1), nuclear factor 1 (NF-1), NF- κ B/Rel, and octamer binding proteins (Oct-1), as well as cyclic AMP-responsive elements (CRE) [93]. The presence of NTO and ORF63 isoforms overlapping this region could interfere with the binding of these host cell proteins. Herpesviruses can evade host cell interferon production and T cell response by producing miRNAs derived from lncRNAs, or antisense RNAs [94–96]. Short noncoding RNAs have been recently documented in the *orf61–62–63* region of VZV [78]. The NTO and ORF63 isoforms detected in our study could be a possible source of these sncRNAs.

We detected a hyper-editing process in the novel NTO3 transcript. This phenomenon was previously observed in members of the herpesvirus family including alphaherpesviruses [97–99]. This process plays a crucial role in the cell's innate immunity [100], while it can be

hijacked by some viruses to evade inactivation [30]. Additionally, in hepatitis delta virus hyper-editing is indispensable for replication [31]. A to I editing decreases the affinity of the antisense transcript to the sense RNA, destabilizing their interaction, which may affect the binding of dsRNA enzymes like RNase III homologues [101]. Our *in-silico* analysis shows that despite elevating the level of free energy of the sense-antisense RNA hybrid, hyper-editing does not result in a change of the secondary structure, but in the formation of I-U wobble pairs which are significantly more resistant to possible Dicer cleavage inhibiting RNA interference [102]. It is also possible that I-U pairs play a role in the cleavage and degradation of ORF62 by a process mediated by the Tudor staphylococcal nuclease (Tudor-SN) [103, 104]. Further investigation is needed to elucidate the significance of hyper-editing on the NTO3. An evaluation of gene expression at different time points could shed light on the presence of the asRNA and on the amount of editing in different stages of the viral life cycle. Additionally, miRNA assays of the viral transcriptome in different time points of the infection could prove the effect of the RNAi on the dsRNA formed by the sense-antisense pair.

Splice events are thought to be rare in alphaherpesviruses, however, they appear to be underestimated in the light of LRS techniques [105]. In this work, we enriched the list of spliced transcripts of the lytic phase of the viral infection, and we identified the novel combinations of splice sites in ORF42/45. The scarcity of the novel splice isoforms implies further validation, using techniques capable of detecting rare transcripts like targeted direct RNA sequencing.

At least six VZV-transcripts have been shown to be expressed in latency [106, 107], however recent target enrichment SRS data suggests that VLT and ORF63 are the only two expressed transcripts, maintaining the dormant phase of VZV. In this work Depledge and coworkers showed that the VLT is spliced differently in the latent and lytic phase of the viral life cycle, and identified several length and splice isoforms during productive infection [9]. We confirmed the existence of four introns of this transcript in the longer, lytic forms of VLT. Additionally, we showed a splice isoform of ORF63 and ORF64 possessing the same introns as the VLT. This suggests that occasionally VLT, ORF63 and ORF64 can form a single transcriptional unit. We found that a novel splice isoform of ORF63 produces a long transcript, which has non-canonical splice donor site (GC, instead of GU). Another rare but distinct splice variant of ORF63-L has the same, non-canonical splice donor site. The reason for this could be simply the higher GC content of these regions. Another explanation may be that the splice

isoforms of ORF63 are recognized by host cell spliceosomes differentially, in order to perform varying splicing patterns during the life cycle of virus. Similar alternative splicing has been described in VLT, assuming distinct functions for it in lytic and latent phase [9, 105].

The uORFs are supposed to play a regulatory role in the translation of eukaryotic mRNAs. Our results suggests, that at least some of the uORFs could play a role in the production of N-terminally truncated proteins, while most of them have a large-enough space between their stop codons and the protein coding ORF's AUG for a reinitiation event, thus resulting in unaltered protein translation [108]. Further studies implying ribosome profiling and mRNA and protein expression studies are needed to determine the precise function of these uORFs. Nevertheless, the use of alternative promoters for producing TSS variants with or without uORFs may have a role in providing a differential control of translation at distinct stages of the viral lifecycle [22]. TSSs and TESs detected by former S1 nuclease and primer extension experiments agree with our results with base pair precision, demonstrating that our comprehensive LRS transcriptomic survey of VZV complemented with the proposed bioinformatics pipeline can provide a useful background to design other viral transcriptome reannotation analysis by ONT platform. It is also important to know which transcripts are affected by a genetic modification of the virus. The organization principles of the viral genome and the viral replication can only be understood with a detailed knowledge of the transcriptome.

Conclusions

This study substantially redefines the VZV transcriptome by identifying a large number of novel RNA molecules and transcript isoforms, as well as revealing a complex pattern of transcriptional overlaps. The discovery of novel TSS and TES isoforms can lead to a better understanding of the regulation of the viral gene expression through alternative promoters or through RNA turnover, whereas uORFs discovered in this work can modulate translation. The extensive transcriptional overlaps may indicate an interaction between the transcription machineries. Additionally, the discovery of nroRNAs suggests an interference between the replication and transcription apparatuses. Besides the significant advance in transcriptome annotation, these data, especially the data on novel putative mRNAs and ORFs, may also provide useful information to target transcript candidates for controlling this virus.

Additional files

Additional file 1: The procedure of finding novel transcriptional isoforms. The flowchart describes the pipeline used to detect novel TSSs and TES-s. (PDF 505 kb)

Additional file 2: The list of 5', 3' ends, TSSs, TESs, promoter motifs, polyadenylation signals, transcriptional overlaps and uORFs identified by this study. (XLSX 383 kb)

Additional file 3: The secondary structure of NTO3 (a. and b.) as well as the hybrid formed by the ORF62–5' fragment and the NTO3 (c and d). a. The secondary structure of the unedited NTO3 with a free energy of –143.2 kcal/mol. The adenines in the editing sites are colored in green. b. The secondary structure of the edited NTO3 with a free energy of –169.4 kcal/mol. The inosines in the editing sites are colored in orange. c. The secondary structure of the sense-antisense hybrid composed of the first 467 bases of ORF62 labeled ORF62–5' (gray) and the full sequence of NTO3 (blue), the latter is its unedited form. The free energy of the structure formed by the two molecules is –822.9 kcal/mol. d. The secondary structure of the sense-antisense hybrid composed of the first 467 bases of ORF62 labeled ORF62–5' (gray) and the full sequence of NTO3 (blue), the later in its edited form. The free energy of the structure formed by the two molecules is –818.7 kcal/mol. The position of IU base pairs is marked with black arrows. (TIFF 1530 kb)

Abbreviations

ADAR1: Adenosine deaminase acting on RNA type 1; AS: Antisense (transcript); AT: Alternative termination; C: Complex (transcript); HCMV: Human cytomegalovirus; HSV-1: Herpes simplex virus type 1; lncRNA: Long non-coding RNA; LRS: Long-read sequencing; miRNA: MicroRNA; ncRNA: Non-coding RNA; nroRNA: Near-replication-origin RNA; NTO: Near to (replication) origin; ONT: Oxford Nanopore Technologies; PAS: Poly(A) signal; PRV: Pseudorabies virus; RT: Reverse transcriptase; sncRNA: Short non-coding RNA; SRS: Short-read sequencing; TES: Transcription end site; tORF: Truncated open reading frame; -TR: 5'-truncated (transcript); TSS: Transcription start site; uORF: Upstream open reading frame; UTR: Untranslated region; VLT: VZV latency transcript; VLT_{ly}: Lytic form of VLT; VZV: Varicella Zoster Virus

Acknowledgements

Not applicable.

Author contributions

IP analyzed the data, participated in RNA-isolation sequence alignment and drafted and wrote the manuscript. NM participated in data analysis, carried out the statistical analysis, prepared the figures and revised the manuscript. ZBa carried out the analysis of the targeted sequencing. DT carried out the ONT MinION cDNA sequencing, participated in the design of the study, and took part in drafting the manuscript. AS participated in the sequence alignment and carried out the *in silico* analysis. ZC prepared the RNA, DNA and cDNA samples and participated in the MinION sequencing. KM carried out the virus infection and propagated the cells. ZB conceived, designed and coordinated the study and wrote the manuscript. All authors have read and approved the final version of the manuscript.

Funding

This work was supported by the Swiss-Hungarian Cooperation Programme [SH/7/2/8] and NKFIH OTKA [K128247] to ZB. The work was also supported by the Bolyai János Scholarship of the Hungarian Academy of Sciences to DT.

Availability of data and materials

The sequencing data and the transcriptome assembly have been uploaded to the European Nucleotide Archive under the project accession number PRJEB25401.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Medical Biology, Faculty of Medicine, University of Szeged, Szeged 6720, Hungary. ²Department of Medical Microbiology and Immunobiology, Faculty of Medicine, University of Szeged, Szeged 6720, Hungary.

Received: 4 September 2018 Accepted: 19 November 2018

Published online: 04 December 2018

References

- Kennedy PGE. Varicella-zoster virus latency in human ganglia. *Rev Med Virol.* 2002;12:327–34.
- Cohrs RJ, Gilden DH, Mahalingam R. Varicella zoster virus latency, neurological disease and experimental models: an update. *Front Biosci.* 2004;9:751–62.
- Arvin AM. Investigations of the pathogenesis of varicella zoster virus infection in the SCIDhu mouse model. *Herpesviridae.* 2006;13:75–80.
- Zerboni L, Sen N, Oliver SL, Arvin AM. Molecular mechanisms of varicella zoster virus pathogenesis. *Nat Rev Micro.* 2014;12:197–210.
- Maresova L, Pasięka TJ, Homan E, Gerday E, Grose C. Incorporation of three endocytosed varicella-zoster virus glycoproteins, gE, gH, and gB, into the Virion envelope. *J Virol.* 2005;79:997–1007.
- Davison AJ, Scott JE. The complete DNA sequence of varicella-zoster virus. *J Gen Virol.* 1986;67:1759–816.
- Tillieux SL, Halsey WS, Thomas ES, Voycik JJ, Sathe GM, Vassilev V. Complete DNA sequences of two oka strain varicella-zoster virus genomes. *J Virol.* 2008;82:11023–44.
- Kinchington PR, Leger AJS, Guedon J-MG, Hendricks RL. Herpes simplex virus and varicella zoster virus, the house guests who never leave. *Herpesviridae.* 2012;3:5.
- Depledge DP, Ouwendijk WJD, Sadaoka T, Braspenning SE, Mori Y, Cohrs RJ, et al. A spliced latency-associated VZV transcript maps antisense to the viral transactivator gene 61. *Nat Commun.* 2018;9:1167.
- Perera LP, Mosca JD, Sadeghi-Zadeh M, Ruyechan WT, Hay J. The varicella-zoster virus immediate early protein, IE62, can positively regulate its cognate promoter. *Virology.* 1992;191:346–54.
- Kinchington PR, Houghland JK, Arvin AM, Ruyechan WT, Hay J. The varicella-zoster virus immediate-early protein IE62 is a major component of virus particles. *J Virol.* 1992;66:359–66.
- Kinchington PR, Bookey D, Turse SE. The transcriptional regulatory proteins encoded by varicella-zoster virus open reading frames (ORFs) 4 and 63, but not ORF 61, are associated with purified virus particles. *J Virol.* 1995;69:4274–82.
- Moriuchi H, Moriuchi M, Straus SE, Cohen JL. Varicella-zoster virus open reading frame 10 protein, the herpes simplex virus VP16 homolog, transactivates herpesvirus immediate-early gene promoters. *J Virol.* 1993;67:2739–46.
- Che X, Zerboni L, Sommer MH, Arvin AM. Varicella-zoster virus open reading frame 10 is a virulence determinant in skin cells but not in T cells *in vivo*. *J Virol.* 2006;80:3238–48.
- Reichelt M, Brady J, Arvin AM. The replication cycle of varicella-zoster virus: analysis of the kinetics of viral protein expression, genome synthesis, and Virion assembly at the single-cell level. *J Virol.* 2009;83:3904–18.
- Khalil MI, Che X, Sung P, Sommer MH, Hay J, Arvin AM. Mutational analysis of varicella-zoster virus (VZV) immediate early protein (IE62) subdomains and their importance in viral replication. *Virology.* 2016;492:82–91.

17. Cohrs RJ, Hurley MP, Gilden DH. Array analysis of viral gene transcription during lytic infection of cells in tissue culture with varicella-zoster virus. *J Virol. American Society for Microbiology*. 2003;77:11718–32.
18. Tombácz D, Csabai Z, Oláh P, Balázs Z, Likó I, Zsigmond L, et al. Full-Length Isoform Sequencing Reveals Novel Transcripts and Substantial Transcriptional Overlaps in a Herpesvirus. *PLoS One Public Libr Sci*. 2016;11: 0162868.
19. Kronstad LM, Brulois KF, Jung JU, Glaunsinger BA. Dual short upstream open Reading frames control translation of a Herpesviral Polycistronic mRNA. *PLoS Pathog*. 2013;9:e1003156.
20. Talbot SJ, Weiss RA, Kellam P, Boshoff C. Transcriptional analysis of human herpesvirus-8 open reading frames 71, 72, 73, K14, and 74 in a primary effusion lymphoma cell line. *Virology*. 1999;257:84–94.
21. Griffiths A, Coen DM. An unusual internal ribosome entry site in the herpes simplex virus thymidine kinase gene. *Proc Natl Acad Sci U S A*. 2005;102: 9667–72.
22. Balázs Z, Tombácz D, Szűcs A, Csabai Z, Megyeri K, Petrov AN, et al. Long-read sequencing of human cytomegalovirus transcriptome reveals RNA isoforms carrying distinct coding potentials. *Sci Rep*. 2017;7:15989.
23. Moldován N, Tombácz D, Szűcs A, Csabai Z, Snyder M, Boldogkői Z. Multi-platform sequencing approach reveals a novel transcriptome profile in pseudorabies virus. *Front Microbiol*. 2018;8:2708.
24. Sadaoka T, Yanagi T, Yamanishi K, Mori Y. Characterization of the varicella-zoster virus ORF50 gene, which encodes glycoprotein M. *J Virol*. 2010;84: 3488–502.
25. Kemble GW, Annunziato P, Lungu O, Winter RE, Cha T-A, Silverstein SJ, et al. Open Reading frame S/L of varicella-zoster virus encodes a cytoplasmic protein expressed in infected cells. *J Virol*. 2000;74:11311–21.
26. Visalli RJ, Nicolosi DM, Irven KL, Goshorn B, Khan T, Visalli MA. The varicella-zoster virus DNA encapsidation genes: identification and characterization of the putative terminase subunits. *Virus Res*. 2007;129:200–11.
27. Walkley CR, Li JB. Rewriting the transcriptome: adenosine-to-inosine RNA editing by ADARs. *Genome Biol BioMed Central*. 2017;18:205.
28. Eggington JM, Greene T, Bass BL. Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun*. 2011;2:319.
29. Matthews MM, Thomas JM, Zheng Y, Tran K, Phelps KJ, Scott AI, et al. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol*. 2016;23:426–33.
30. Zahn RC, Schelp I, Utermöhlen O, von Laer D. A-to-G hypermutation in the genome of lymphocytic choriomeningitis virus. *J Virol. American Society for Microbiology*. 2007;81:457–64.
31. Wong SK, Lazinski DW. Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc Natl Acad Sci*. 2002;99:15118–23.
32. Kozak M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene Elsevier*. 2005;361:13–37.
33. Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. 1986;44: 283–92.
34. Pisarev AV, Kolupaeva VG, Pisareva VP, Merrick WC, Hellen CUT, Pestova TV. Specific functional interactions of nucleotides at key –3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev Cold Spring Harbor Laboratory Press*. 2006;20:624–36.
35. Luukkonen BG, Tan W, Schwartz S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *J Virol American Society for Microbiology (ASM)*. 1995;69:4086–94.
36. Kozak M. Constraints on reinitiation of translation in mammals. *Nucleic Acids Res Oxford University Press*. 2001;29:5226–32.
37. Chambers J, Angulo A, Amaratunga D, Guo H, Jiang Y, Wan JS, et al. DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. *J Virol*. 1999;73:5757–66.
38. Ebrahimi B, Dutia BM, Roberts KL, Garcia-Ramirez JJ, Dickinson P, Stewart JP, et al. Transcriptome profile of murine gammaherpesvirus-68 lytic infection. *J Gen Virol*. 2003;84:99–109.
39. Oláh P, Tombácz D, Póka N, Csabai Z, Prazsák I, Boldogkői Z. Characterization of pseudorabies virus transcriptome by Illumina sequencing. *BMC Microbiol*. 2015;15:130.
40. Baird NL, Bowlin JL, Cohrs RJ, Gilden D, Jones KL. Comparison of varicella-zoster virus RNA sequences in human neurons and fibroblasts. *J Virol*. 2014; 88:5877–80.
41. 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 Nat Publ Group*. 2017;8:16027.
42. 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;33:736–42.
43. Luo GX, Taylor J. Template switching by reverse transcriptase during DNA synthesis. *J Virol*. 1990;64:4321–8.
44. Cocquet J, Chong A, Zhang G, Veitia RA. Reverse transcriptase template switching and false alternative transcripts. *Genomics*. 2006;88:127–31.
45. Kuo RI, Tseng E, Eory L, Paton IR, Archibald AL, Burt DW. Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. *BMC Genomics*. 2017;18:323.
46. Moldován N, Szűcs 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;365:fny013.
47. 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;237:37–46.
48. Szűcs A, Moldován N, Tombácz D, Csabai Z, Snyder M, Boldogkői Z. Long-read sequencing reveals a GC pressure during the evolution of porcine endogenous retrovirus. *Genome Announc*. 2017;5:01040–17.
49. Moldován N, Tombácz D, Szűcs A, Csabai Z, Balázs Z, Kis E, et al. Third-generation Sequencing Reveals Extensive Polycistronism and Transcriptional Overlapping in a Baculovirus. *Sci Rep Nat Publ Group*. 2018;8:8604.
50. Tombacz D, Sharon D, Olah P, Csabai Z, Snyder M, Boldogkoi Z. Strain Kaplan of Pseudorabies Virus Genome Sequenced by PacBio Single-Molecule Real-Time Sequencing Technology. *Genome Announc*. 2014;2: e00628–14 e00628–14.
51. O’Grady T, Wang X, Höner zu Bentrup K, Baddoo M, concha M, Flemington EK. Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res*. 2016;44:e145.
52. Balázs Z, Tombácz D, Szűcs A, Snyder M, Boldogkői Z. Long-read sequencing of the human cytomegalovirus transcriptome with the Pacific Biosciences RSII platform. *Sci Data Nat Publ Group*. 2017;4:170194.
53. Tombácz D, Csabai Z, Szűcs A, Balázs Z, Moldován N, Sharon D, et al. Long-read isoform sequencing reveals a hidden complexity of the transcriptional landscape of herpes simplex virus type 1. *Front Microbiol*. 2017;8:1079.
54. Boldogkői Z. Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci. *Front Genet Frontiers*. 2012;3:122.
55. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics Oxford University Press*. 2005; 21:1859–75.
56. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform Oxford University Press*. 2013;14:178–92.
57. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
58. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*. 2010;11:129.
59. Kearsse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28: 1647–9.
60. Narang V, Sung W-K, Mittal A. Computational modeling of oligonucleotide positional densities for human promoter prediction. *Artif Intell Med Elsevier*. 2005;35:107–19.
61. Rock DL, Nesburn AB, Ghiasi H, Ong J, Lewis TL, Lokensgard JR, et al. Detection of latency-related viral RNAs in trigeminal ganglia of rabbits latently infected with herpes simplex virus type 1. *J Virol American Society for Microbiology (ASM)*. 1987;61:3820–6.
62. Jin L, Scherba G. Expression of the pseudorabies virus latency-associated transcript gene during productive infection of cultured cells. *J Virol American Society for Microbiology*. 1999;73:9781–8.
63. Lenac Roviš T, Bailer SM, Pothineni VR, Ouwendijk WJD, Šimić H, Babić M, et al. Comprehensive analysis of varicella-zoster virus proteins using a new monoclonal antibody collection. *J Virol*. 2013;87:6943–54.

64. Tombácz D, Csabai Z, Oláh P, Havelda Z, Sharon D, Snyder M, et al. Characterization of novel transcripts in pseudorabies virus. *Viruses Multidisciplinary Digital Publishing Institute (MDPI)*. 2015;7:2727–44.
65. Voss JH, Roizman B. Properties of two 5'-coterminal RNAs transcribed part way and across the S component origin of DNA synthesis of the herpes simplex virus 1 genome. *Proc Natl Acad Sci U S A National Academy of Sciences*. 1988;85:8454–8.
66. Huang L, Zhu Y, Anders DG. The variable 3' ends of a human cytomegalovirus oriLyt transcript (SRT) overlap an essential, conserved replicator element. *J Virol American Society for Microbiology Journals*. 1996;70:5272–81.
67. Kotenko SV, Saccani S, Izotova LS, Mirochnitchenko OV, Pestka S, Baluchova K, et al. Human cytomegalovirus harbors its own unique IL-10 homolog (cmvIL-10). *Proc Natl Acad Sci National Academy of Sciences*. 2000;97:1695–700.
68. Kainov YA, Aushev VN, Naumenko SA, Tchevkin EM, Bazykin GA. Complex selection on human polyadenylation signals revealed by polymorphism and divergence data. *Genome Biol Evol Oxford University Press*. 2016;8:1971–9.
69. Pinto PAB, Henriques T, Freitas MO, Martins T, Domingues RG, Wyrzykowska PS, et al. RNA polymerase II kinetics in polyadenylation signal selection. *EMBO J European Molecular Biology Organization*. 2011;30:2431–44.
70. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:W202–8.
71. Ostrove JM, Reinhold W, Fan CM, Zorn S, Hay J, Straus SE. Transcription mapping of the varicella-zoster virus genome. *J Virol*. 1985;56:600–6.
72. Kinchington PR, Vergnes JP, Defechereux P, Piette J, Turse SE. Transcriptional mapping of the varicella-zoster virus regulatory genes encoding open reading frames 4 and 63. *J Virol*. 1994;68:3571–81.
73. Kato T, Kitamura K, Hayakawa Y, Takahashi M, Kojima A, Sato S, et al. Transcription mapping of glycoprotein I (gpl) and gpIV of varicella-zoster virus and immunological analysis of the gpl produced in cells infected with the recombinant vaccinia virus. *Microbiol Immunol*. 1989;33:299–312.
74. Meier JL, Straus SE. Varicella-zoster virus DNA polymerase and major DNA-binding protein genes have overlapping divergent promoters. *J Virol*. 1993;67:7573–81.
75. Kennedy PGE, Grinfeld E, Craighan M, Vierlinger K, Roy D, Forster T, et al. Transcriptomic analysis of varicella-zoster virus infection using long oligonucleotide-based microarrays. *J Gen Virol*. 2005;86:2673–84.
76. Grinfeld E, Ross A, Forster T, Ghazal P, Kennedy PGE. Genome-wide reduction in transcriptomic profiles of varicella-zoster virus vaccine strains compared with parental Oka strain using long oligonucleotide microarrays. *Virus Genes*. 2009;38:19–29.
77. Cohrs RJ, Lee KS, Beach A, Sanford B, Baird NL, Como C, et al. Targeted Genome Sequencing Reveals Varicella-Zoster Virus Open Reading Frame 12 Deletion. *J Virol*. 2017;91:e01141–17.
78. Markus A, Golani L, Ojha NK, Borodiansky-Shteinberg T, Kinchington PR, Goldstein RS. Varicella-zoster virus expresses multiple small noncoding RNAs. *J Virol American Society for Microbiology*. 2017;91:e01710–7.
79. Isomura H, Stinski MF, Kudoh A, Murata T, Nakayama S, Sato Y, et al. Noncanonical TATA sequence in the UL44 late promoter of human cytomegalovirus is required for the accumulation of late viral transcripts. *J Virol American Society for Microbiology*. 2008;82:1638–46.
80. Barth S, Pfuhl T, Mamiani A, Ehses C, Roemer K, Kremmer E, et al. Epstein-Barr virus-encoded microRNA miR-BART2 down-regulates the viral DNA polymerase BALF5. *Nucleic Acids Res*. 2007;36:666–75.
81. Dickson AM, Anderson JR, Barnhart MD, Sokolowski KJ, Oko L, Opyrchal M, et al. Dephosphorylation of HuR protein during alphavirus infection is associated with HuR relocalization to the cytoplasm. *J Biol Chem American Society for Biochemistry and Molecular Biology*. 2012;287:36229–38.
82. Legendre M, Gautheret D. Sequence determinants in human polyadenylation site selection. *BMC Genomics BioMed Central*. 2003;4:7.
83. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res*. 2005;33:201–12.
84. Tian B, Graber JH. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA NIH Public Access*. 2012;3:385–96.
85. Hu J, Lutz CS, Wilusz J, Tian B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA Cold Spring Harbor Laboratory Press*. 2005;11:1485–93.
86. Sommer MH, Zagha E, Serrano OK, Ku CC, Zerboni L, Baiker A, et al. Mutational analysis of the repeated open reading frames, ORFs 63 and 70 and ORFs 64 and 69, of varicella-zoster virus. *J Virol*. 2001;75:8224–39.
87. Kost RG, Kupinsky H, Straus SE. Varicella-zoster virus gene 63: transcript mapping and regulatory activity. *Virology*. 1995;209:218–24.
88. Hoover SE, Cohrs RJ, Rangel ZG, Gilden DH, Munson P, Cohen JL. Downregulation of varicella-zoster virus (VZV) immediate-early ORF62 transcription by VZV ORF63 correlates with virus replication in vitro and with latency. *J Virol American Society for Microbiology (ASM)*. 2006;80:3459–68.
89. He Y, Yuan C, Chen L, Lei M, Zellmer L, Huang H, et al. Transcriptional-Readthrough RNAs Reflect the Phenomenon of "A Gene Contains Gene(s)" or "Gene(s) within a Gene" in the Human Genome, and Thus Are Not Chimeric RNAs. *Genes (Basel)*. 2018;9:40.
90. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. *Science*. 2005;309:1564–6.
91. Yuan C, Liu Y, Yang M, Liao DJ. New methods as alternative or corrective measures for the pitfalls and artifacts of reverse transcription and polymerase chain reactions (RT-PCR) in cloning chimeric or antisense-accompanied RNA. *RNA Biol*. 2013;10:957–67.
92. Huvet M, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, Arneodo A, et al. Human gene organization driven by the coordination of replication and transcription. *Genome Res*. 2007;17:1278–85.
93. Jones JO, Sommer M, Stamatis S, Arvin AM. Mutational analysis of the varicella-zoster virus ORF62/63 intergenic region. *J Virol American Society for Microbiology (ASM)*. 2006;80:3116–21.
94. Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, Xie M, et al. Viral noncoding RNAs: More surprises. *Genes Dev*. 2015;29:567–84.
95. Qiu L, Wang T, Tang Q, Li G, Wu P, Chen K. Long non-coding RNAs: regulators of viral infection and the interferon antiviral response. *Front Microbiol*. 2018;9:1621.
96. Meng XY, Luo Y, Anwar MN, Sun Y, Gao Y, Zhang H, et al. Long non-coding RNAs: emerging and versatile regulators in host-virus interactions. *Front Immunol*. 2017;8:1663.
97. Figueroa T, Boumart I, Coupeau D, Rasschaert D. Hyperediting by ADAR1 of a new herpesvirus lncRNA during the lytic phase of the oncogenic Marek's disease virus. *J Gen Virol*. 2016;97:2973–88.
98. Gandy SZ, Linnstaedt SD, Muralidhar S, Cashman KA, Rosenthal LJ, Casey JL. RNA editing of the human herpesvirus 8 kaposin transcript eliminates its transforming activity and is induced during lytic replication. *J Virol American Society for Microbiology (ASM)*. 2007;81:13544–51.
99. Iizasa H, Wulff B-E, Alla NR, Maragkakis M, Megraw M, Hatzigeorgiou A, et al. Editing of Epstein-Barr virus-encoded BART6 MicroRNAs controls their dicer targeting and consequently affects viral latency. *J Biol Chem*. 2010;285:33358–70.
100. Mannion NM, Greenwood SM, Young R, Cox S, Brindle J, Read D, et al. The RNA-editing enzyme ADAR1 controls innate immune responses to RNA. *Cell Rep*. 2014;9:1482–94.
101. Nishikura K. Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nat Rev Mol Cell Biol*. 2006;7:919–31.
102. Scadden AD, Smith CW. RNAi is antagonized by A-to-G hyper-editing. *EMBO Rep European Molecular Biology Organization*. 2001;2:1107–11.
103. Scadden AD, Smith CW. Specific cleavage of hyper-edited dsRNAs. *EMBO J*. 2001;20:4243–52.
104. Scadden AD. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nat Struct Mol Biol*. 2005;12:489–96.
105. Depledge DP, Sadaoka T, Ouwendijk WJD. Molecular aspects of varicella-zoster virus latency. *Viruses*. 2018;7:349.
106. Cohen JI, Cox E, Pesnicak L, Srinivas S, Krogmann T. The varicella-zoster virus open reading frame 63 latency-associated protein is critical for establishment of latency. *J Virol*. 2004;78:11833–40.
107. Nagel MA, Choe A, Traktinskiy I, Cordero-Cotter R, Gilden D, Cohrs RJ. Varicella-zoster virus transcriptome in latently infected human ganglia. *J Virol*. 2011;85:2276–87.
108. Poyry TAA, Kaminski A, Jackson RJ. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev*. 2004;18:62–75.