



# Decoding the Architecture of the Varicella-Zoster Virus Transcriptome

 Shirley E. Braspenning,<sup>a</sup>  Tomohiko Sadaoka,<sup>b</sup>  Judith Breuer,<sup>c</sup>  Georges M. G. M. Verjans,<sup>a</sup>  Werner J. D. Ouwendijk,<sup>a</sup>  
 Daniel P. Depledge<sup>d</sup>

<sup>a</sup>Department of Viroscience, Erasmus MC, Rotterdam, Netherlands

<sup>b</sup>Division of Clinical Virology, Center for Infectious Diseases, Kobe University Graduate School of Medicine, Kobe, Japan

<sup>c</sup>Department of Infection and Immunity, University College London, London, United Kingdom

<sup>d</sup>Department of Medicine, New York University School of Medicine, New York, New York, USA

Werner J. D. Ouwendijk and Daniel P. Depledge are co-senior authors.

**ABSTRACT** Varicella-zoster virus (VZV), a double-stranded DNA virus, causes varicella, establishes lifelong latency in ganglionic neurons, and reactivates later in life to cause herpes zoster, commonly associated with chronic pain. The VZV genome is densely packed and produces multitudes of overlapping transcripts deriving from both strands. While 71 distinct open reading frames (ORFs) have thus far been experimentally defined, the full coding potential of VZV remains unknown. Here, we integrated multiple short-read RNA sequencing approaches with long-read direct RNA sequencing on RNA isolated from VZV-infected cells to provide a comprehensive reannotation of the lytic VZV transcriptome architecture. Through precise mapping of transcription start sites, splice junctions, and polyadenylation sites, we identified 136 distinct polyadenylated VZV RNAs that encode canonical ORFs, noncanonical ORFs, and ORF fusions, as well as putative noncoding RNAs (ncRNAs). Furthermore, we determined the kinetic class of all VZV transcripts and observed, unexpectedly, that transcripts encoding the ORF62 protein, previously designated *Immediate-Early*, were expressed with *Late* kinetics. Our work showcases the complexity of the VZV transcriptome and provides a comprehensive resource that will facilitate future functional studies of coding RNAs, ncRNAs, and the biological mechanisms underlying the regulation of viral transcription and translation during lytic VZV infection.

**IMPORTANCE** Transcription from herpesviral genomes, executed by the host RNA polymerase II and regulated by viral proteins, results in coordinated viral gene expression to efficiently produce infectious progeny. However, the complete coding potential and regulation of viral gene expression remain ill-defined for the human alphaherpesvirus varicella-zoster virus (VZV), causative agent of both varicella and herpes zoster. Here, we present a comprehensive overview of the VZV transcriptome and the kinetic class of all identified viral transcripts, using two virus strains and two biologically relevant cell types. Additionally, our data provide an overview of how VZV diversifies its transcription from one of the smallest herpesviral genomes. Unexpectedly, the transcript encoding the major viral transactivator protein (pORF62) was expressed with *Late* kinetics, whereas orthologous transcripts in other alphaherpesviruses are typically expressed during the immediate early phase. Therefore, our work both establishes the architecture of the VZV transcriptome and provides insight into regulation of alphaherpesvirus gene expression.

**KEYWORDS** direct RNA-Seq, ORF62, alphaherpesvirus, kinetics, transcriptome, varicella-zoster virus

**Citation** Braspenning SE, Sadaoka T, Breuer J, Verjans GMGM, Ouwendijk WJD, Depledge DP. 2020. Decoding the architecture of the varicella-zoster virus transcriptome. *mBio* 11:e01568-20. <https://doi.org/10.1128/mBio.01568-20>.

**Editor** Thomas Shenk, Princeton University

**Copyright** © 2020 Braspenning et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Werner J. D. Ouwendijk, [w.ouwendijk@erasmusmc.nl](mailto:w.ouwendijk@erasmusmc.nl), or Daniel P. Depledge, [daniel.depledge@nyulangone.org](mailto:daniel.depledge@nyulangone.org).

**Received** 12 June 2020

**Accepted** 31 August 2020

**Published** 6 October 2020

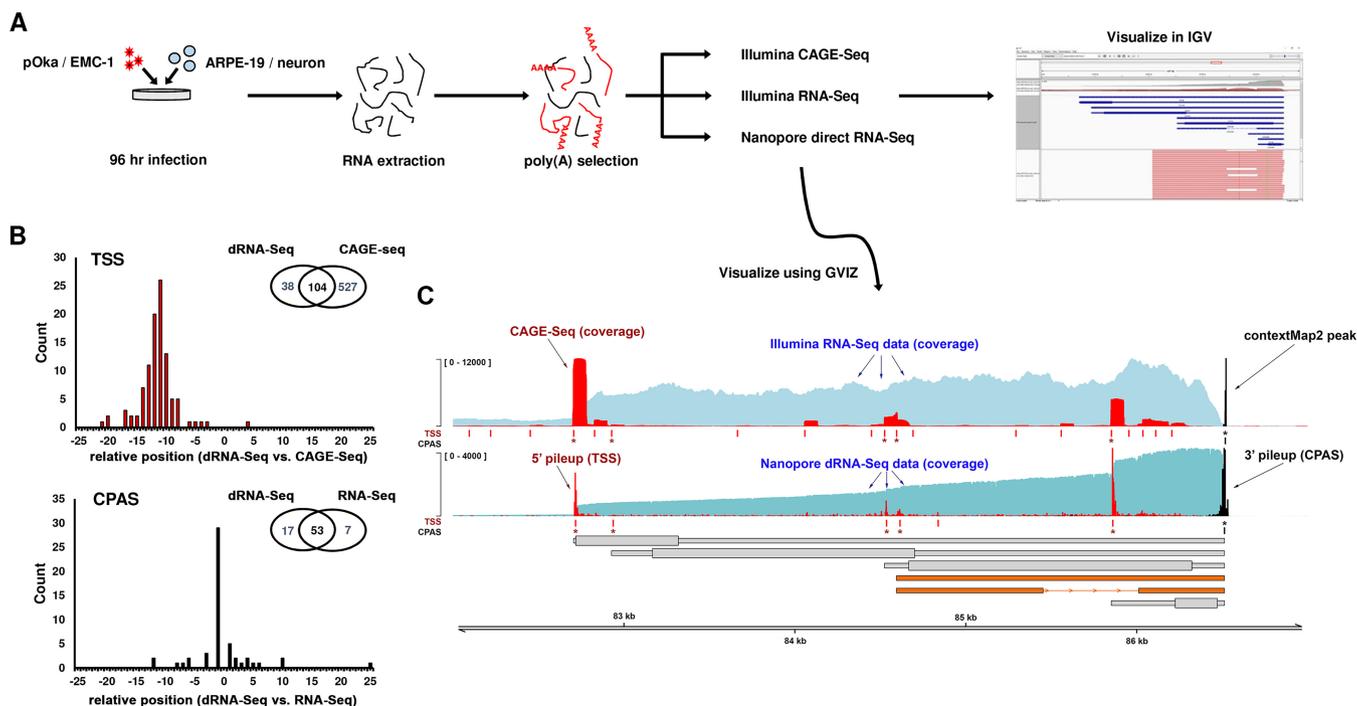
Varicella-zoster virus (VZV) is a ubiquitous human alphaherpesvirus and causative agent of both varicella (chickenpox) and herpes zoster (HZ or shingles) (1). Varicella results from primary VZV infection and leads to the establishment of a lifelong latent infection in sensory neurons of the trigeminal and dorsal root ganglia (2, 3). In one-third of infected individuals, VZV reactivates from latency later in life to cause HZ (1). Whereas varicella is generally experienced as a benign childhood disease, HZ is frequently associated with difficult-to-treat chronic pain (postherpetic neuralgia) (4, 5). Despite the recent availability of the highly effective HZ subunit vaccine (Shingrix), the health and societal burden of HZ and its complications remain high due to adverse side effects, the high cost of the vaccines, and changing demographics (6).

The 125-kb double-stranded DNA (dsDNA) genome of VZV, first sequenced in 1986, encodes at least 71 unique open reading frames (ORFs) that are expressed during lytic infection (7, 8). The current annotation of the VZV genome largely relies on both *in silico* ORF predictions and homologous ORFs in the closely related human alphaherpesvirus herpes simplex virus 1 (HSV-1) (8). For most VZV ORFs, the boundaries of transcription are not accurately determined, meaning that transcription start sites and polyadenylation sites are poorly resolved. Moreover, *in silico* ORF prediction does not account for the possibility of spliced transcripts. Indeed, we recently applied ultradeep short-read RNA sequencing to define the latent viral transcriptome and discovered the spliced VZV latency-associated transcript (VLT) (9).

The compact nature of viral genomes, combined with their ability to encode overlapping RNAs, presents a significant challenge to studies that rely on interrogating genome sequences or that use viral mutants to probe the function(s) of viral proteins. Here, a single point mutation or frameshift may impact multiple distinct RNAs at the same time, a factor that may be further confounded by inaccurate or missing transcript annotations. The reannotation of the human cytomegalovirus (HCMV) transcriptional and translational landscape and subsequent refinements of HSV-1, Kaposi's sarcoma-associated herpesvirus (KSHV), and human herpesvirus 6 (HHV6) transcriptome architectures have all demonstrated that herpesviruses exhibit a complex transcriptional pattern of alternative splicing, opposing transcription, read-through transcription, fusion transcripts, 5' untranslated region (5' UTR) and 3' UTR variations, and previously unidentified noncoding RNAs (ncRNAs) (10–17). Indeed, recent cDNA-based long-read sequencing also indicated that the lytic VZV transcriptome is substantially more complex than previously recognized (18).

By analogy to other herpesviruses and limited experimental data (19, 20), VZV transcripts and their encoded proteins have been divided into three kinetic classes: immediate early (*IE*), early (*E*), and late (*L*). Expression of *E* and *L* transcripts is considered dependent on viral proteins of the preceding kinetic classes, while expression of *IE* transcripts occurs in the absence of viral protein synthesis (21). Prior studies have defined four VZV proteins encoded by ORF4, ORF61, ORF62, and ORF63 as being transcriptional regulators that initiate lytic transcript expression (22–26), whose corresponding transcripts have been classified as *IE* by analogy to their HSV-1 orthologues. *L* transcripts, such as  $_{lyt}$ VLT, the lytic isoform of VLT, are expressed either at very low levels prior to or exclusively after viral DNA replication has commenced (9). However, the species specificity and highly cell-associated nature of VZV *in vitro* have hampered detailed analysis of VZV transcription. Improved protocols to obtain cell-free VZV and mass spectrometry have provided some insight into the temporal pattern of viral protein expression (27) but lack sensitivity compared to RNA sequencing and do not provide information on viral transcription.

To address this, we have decoded the architecture of the lytic VZV transcriptome in both human epithelial cells and neurons while contrasting discrete VZV strains. We subsequently established the kinetic class of all lytic viral transcripts and integrated these results to provide a comprehensive overview of the complexity and structure of the lytic VZV transcriptome as a rich resource that will enhance future functional studies of VZV biology.



**FIG 1** Decoding the complexity of the lytic VZV transcriptome. (A) Experimental strategy: ARPE-19 cells and hESC-derived neurons were infected with VZV EMC-1 (clade 1) or pOka (clade 2) for 96 h. Total RNA was extracted, and poly(A) fraction was isolated for sequencing by Illumina CAGE-Seq, Illumina RNA-Seq, and/or Nanopore dRNA-Seq. Sequence data were aligned against the VZV strain Dumas reference genome (GenBank accession no. [NC\\_001348.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_001348.1)) and visualized using the Integrative Genomics Viewer (IGV) (32) and GVIZ (74). (B) Transcription start sites (TSS) as well as cleavage and polyadenylation sites (CPAS) were identified in Nanopore and Illumina data sets. Histograms show the distances observed between Nanopore and Illumina predictions, while inset Venn diagrams indicate the numbers of sites identified and their conservation between data sets. (C) Integration of Illumina RNA-Seq, Illumina CAGE-Seq, and Nanopore dRNA-Seq data sets. Coverage plots for Illumina RNA-Seq (light blue), CAGE-Seq (red), and Nanopore dRNA-Seq (teal) are integrated with pileup data that maps TSS (red) and CPAS (black). Rows denoted by TSS and CPAS indicate positions of TSS and CPAS identified using HOMER software (77) for Nanopore dRNA-Seq and Illumina CAGE-Seq data, and ContextMap2 software (71) for Illumina RNA-Seq data. TSS and CPAS that are conserved between sequencing technologies are indicated with asterisks, and only these are used to define the boundaries of newly annotated transcripts. RNA structures (gray) are inferred from these conserved sites. Wide and thin boxes indicate canonical coding sequence (CDS) domains and untranslated regions (UTRs), respectively. Newly identified RNAs (orange) are shown without predicted CDS domains.

## RESULTS

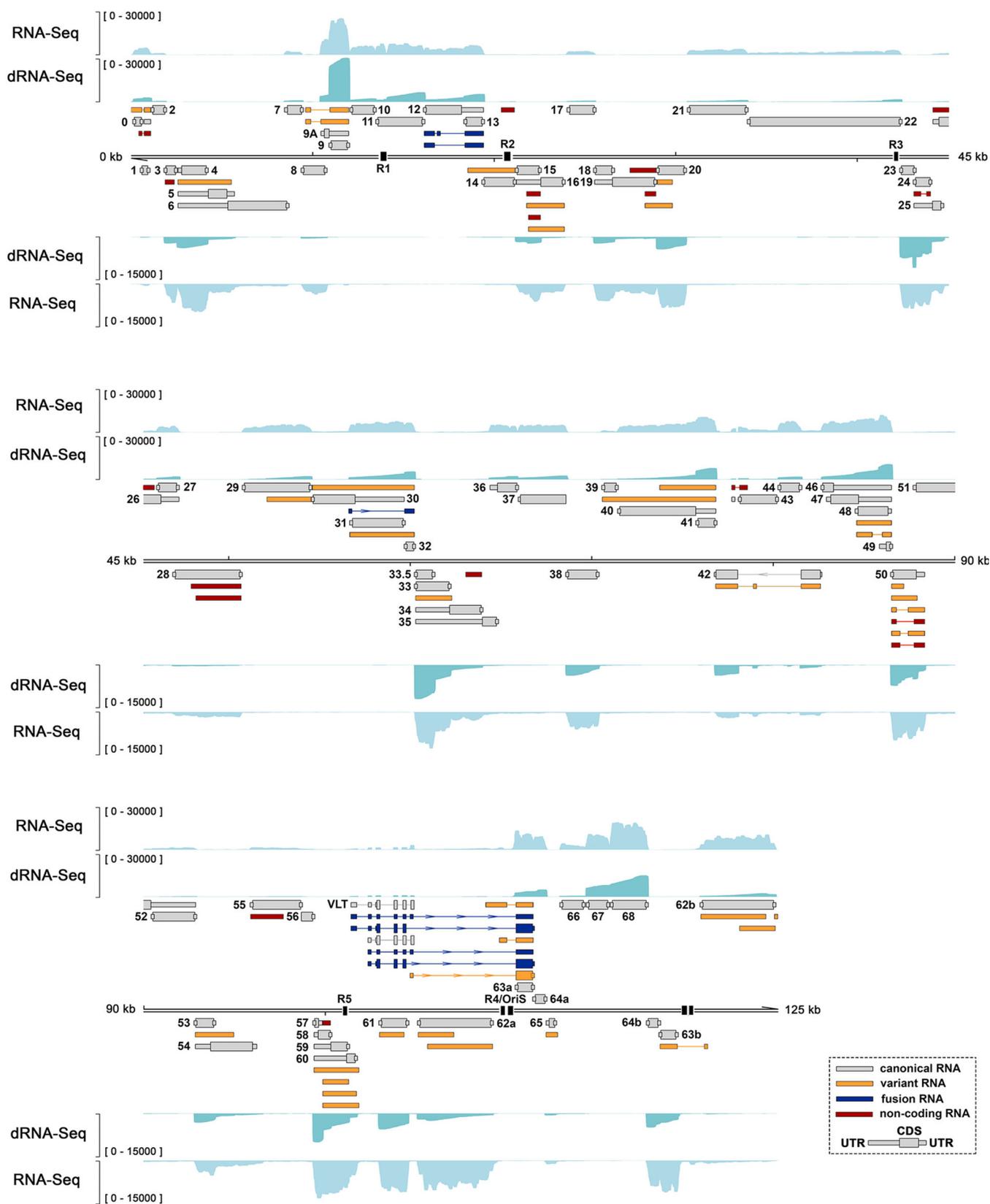
**Decoding the complexity of lytic VZV gene expression.** Standard methods for annotating viral transcriptomes require the integration of multiple types of Illumina RNA sequencing (RNA-Seq) data to identify transcription start sites (TSS), cleavage and polyadenylation sites (CPAS), splice sites, and transcript structures. The last is particularly challenging to infer using conventional short-read sequencing approaches (28). In contrast, direct RNA sequencing (dRNA-Seq) using Nanopore arrays offers the potential to capture all these distinct data points in a single sequencing run (10, 29). To examine the structure of the lytic VZV transcriptome, ARPE-19 cells were infected with the VZV pOka wild-type strain and total RNA was extracted at 96 h post-infection (hpi) (Fig. 1A). The 96-hpi time point was chosen to maximize the diversity of VZV transcripts likely to be present. Sequencing of the polyadenylated RNA fraction was performed using both RNA-Seq and dRNA-Seq (Fig. 1A). Whereas short reads generated by standard Illumina RNA-Seq are not amenable for accurate isoform reconstruction in complex reads, they provide higher sequencing depth and detection of CPAS and, crucially, enable splice-site correction of dRNA-Seq reads via the junction-polishing package in the software FLAIR (30). In contrast, dRNA-Seq can sequence full-length RNAs and provides critical information on the presence of discrete RNA isoforms with regions of overlap, while also allow mapping of TSS and CPAS. Finally, we performed Illumina Cap Analysis Gene Expression Sequencing (CAGE-Seq) (31) to map TSS by an orthologous approach (Fig. 1A).

TSS and CPAS estimates provided by dRNA-Seq data closely overlapped those derived from our Illumina approaches (Fig. 1B). TSS sensitivity was nearly 4-fold higher

in CAGE-Seq data sets than with dRNA-Seq, with effectively all TSS uniquely found by CAGE-Seq being low abundance—likely reflecting artifacts derived from RNA processing (i.e., recapping of cleaved RNA) or a generalized dysregulation of transcription initiation accompanying the late stages of a viral infection. A total of 104 TSS overlapped between the dRNA-Seq and CAGE-Seq data sets, all of which were the most abundant TSS in both data sets (see Table S1 in the supplemental material). Importantly, dRNA-Seq TSS estimates were located up to 20 nucleotides (nt) downstream (median 11 nt) of TSS derived via CAGE-Seq (Fig. 1B). This difference is best explained by the presence of low-quality ends of dRNA-Seq reads that are not aligned when using local alignment strategies. CPAS sensitivity was higher in dRNA-Seq than RNA-Seq data sets (70 versus 60 sites) with 53 sites detected by both approaches (Table S1). CPAS estimates provided by dRNA-Seq aligned closely with those derived from RNA-Seq data (median 1-nt difference, Fig. 1B) due to the 3'→5' direction of dRNA-Seq. Finally, we reconstructed the VZV transcriptome using TSS and CPAS to define transcript structures followed by visual confirmation of read data to identify splice sites, define alternatively spliced transcripts, and examine read-through transcription (Fig. 1C). For a given transcript to be included in our final annotation, we required the specified TSS and CPAS to be present in both CAGE-Seq and dRNA-Seq data sets (Table S1). We further required that the transcript could be verified by eye through inspection of sequence read alignments in BED files (available at <https://github.com/DepledgeLab/vzv-2.0>) using IGV (32) and that splice junctions (Table S2) were reported in both the RNA-Seq and dRNA-Seq data sets. This is exemplified in Fig. 1C where five TSS (indicated by asterisks) were identified by both dRNA-Seq and CAGE-Seq while only a single CPAS was identified in the RNA-Seq and dRNA-Seq data. This leads us to initially annotating five distinct transcripts (each with a unique TSS and common CPAS), while further inspection of the sequence read alignments in IGV confirmed the presence of both an unspliced and spliced transcript sharing the same TSS and CPAS (shown in orange in Fig. 1C), resulting in six distinct transcripts in total.

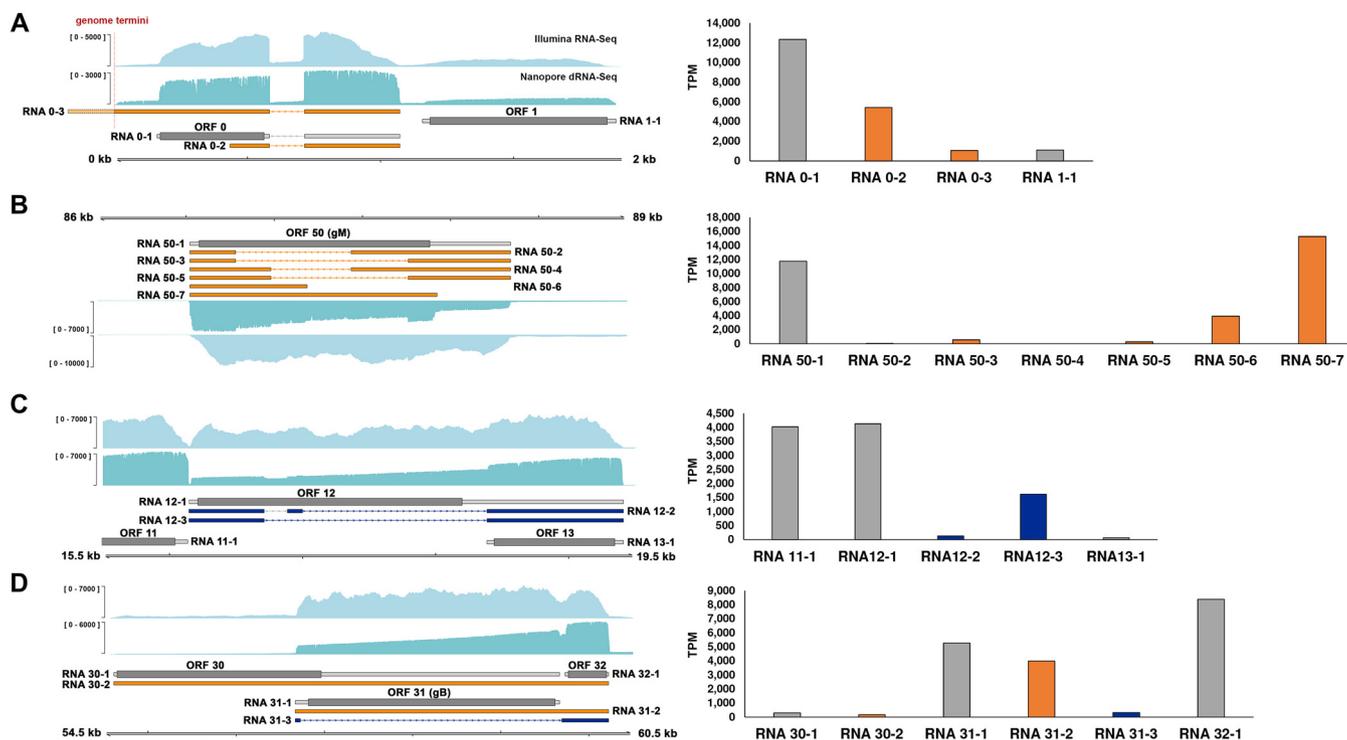
**Reannotation of the VZV transcriptome reveals alternative transcript isoforms and putative noncoding RNAs.** In VZV pOka-infected ARPE-19 cells, we identified 136 distinct VZV RNAs that were readily detectable at 96 hpi. Along with defining the UTRs of 96 RNAs encoding the 71 canonical VZV ORFs, we also identified 40 additional RNAs (Fig. 2 and Table S3). To reduce confusion, we numbered all RNAs according to the respective canonical ORFs and delineated transcript isoforms encoding at least part of the same ORF by number. For instance, three transcripts are transcribed from the ORF0 locus and these are here referred to as VZV RNA 0-1, 0-2, and 0-3. The identified RNAs included transcripts encoding 5' extended ORFs, 5' truncated ORFs, 3' extended ORFs, 3' truncated ORFs, internally spliced variants, and putative noncoding RNAs (ncRNAs). Importantly, our study confirmed several previously described RNA isoforms such as a genome-terminus-spanning RNA that encodes ORF0 (RNA 0-3) (33) and identified an additional 5' truncated ORF0 RNA of unknown function (RNA 0-2) (Fig. 3A). Similarly, extensive low-level internal splicing of ORF50 has previously been reported (34) (RNAs 50-1 to 50-5) and was similarly observed here, supplemented by two additional 5' truncated isoforms (RNAs 50-6 and 50-7) expressed at relatively high abundance (Fig. 3B). Examples of previously undocumented RNAs include two new spliced ORF9 transcript isoforms and two internally spliced transcript isoforms containing N-terminal ORF24 and ORF48 coding sequence (CDS) domains that are spliced into novel C-terminal domains (Fig. 2). Finally, we confirmed expression of transcripts encoding the novel ORF9 and ORF48 variants in VZV-infected ARPE-19 cells by RT-PCR (Fig. S1A and G and Fig. S1B and H).

Fusion transcripts combine sequences from two or more distinct canonical viral transcripts and most likely result from transcription termination occurring at an alternative CPAS downstream of the canonical CPAS, followed by internal splicing (10). The resulting fusion transcripts are predicted to encode new proteins that contain fused domains from two or more distinct protein products. We identified seven VZV fusion transcripts in total. Four of these contain distinct fusions of transcripts encoding pVLT and pORF63 (Figure 2, Table S3). Internal splicing of RNA 12-2 and RNA 12-3 transcripts



**FIG 2** Reannotation of the lytic VZV transcriptome reveals novel viral transcript isoforms, fusion transcripts, and putative noncoding RNAs. The reannotated lytic VZV transcriptome includes 71 transcripts encoding canonical ORFs with their UTRs defined (gray), 39 alternative isoforms of existing RNAs (orange), 7 fusion RNAs (dark blue), and 17 putative polyadenylated ncRNAs (red), the last predicted using CPC 2.0 (36). Wide and thin boxes indicate canonical CDS domains and UTRs, respectively. Absence of CDS regions indicates that the respective VZV RNA has an uncertain coding potential. Illumina RNA-Seq (light blue)

(Continued on next page)



**FIG 3** Examples of specific variant and fusion viral RNAs expressed during lytic VZV infection. (A to D) Examples of variant transcripts encoding ORF0 (A) and ORF50 (B) and novel fusion transcripts encoding (parts of) ORF12 and ORF13 (C) and ORF31 and ORF32 (D). Illumina RNA-Seq (light blue) and Nanopore dRNA-Seq (teal) coverage plots are derived from ARPE-19 cells lytically infected with VZV strain pOka for 96 h. Fusion RNAs are shown in dark blue, with variant RNAs shown in orange and canonical RNAs in gray. The dashed red line in panel A indicates the position of the genome termini spanned by RNA 0 to 3. (A to D) Genome coordinates are shown, while strand is indicated by placement of the genome track (below, top strand; above, bottom strand). y axis values indicate the maximum read depth of that track. For graphs shown on the right, the y axis denotes transcript per million (TPM) counts.

yields two distinct splice variants that fuse parts of ORF12 and ORF13 CDS domains (Fig. 3C). Additionally, we observed transcripts that fused the 5' UTR of ORF31 transcripts to ORF32 transcripts, resulting in a transcript encoding pORF32 with an alternative 5' UTR (Fig. 3D). We confirmed expression of transcripts containing the ORF12-ORF13 and ORF31-ORF32 fusions in VZV-infected ARPE-19 cells by RT-PCR (Fig. S1C and I and Fig. S1D and J).

Additionally, we discovered two novel polyadenylated VZV transcripts: RNA 13.5-1 and RNA 43-2. RNA 13.5-1 is 634 nt long, encodes two putative CDS domains (88 and 55 amino acids [aa]), and is positioned antisense to the RNA 14-1 (encoding pORF14). Like RNA 14-1, RNA 13.5-1 stretches across the R2 reiterative region, a short-repeat region that exhibits length variations between viral strains and within viral populations and thus leads to length variations in the encoded transcripts (35) (Fig. 2 and Fig. S1E). We also identified a highly expressed 3' truncated RNA (RNA 43-2) that overlaps the 5' end of RNA 43-1 (encoding pORF43) (Fig. 2 and Fig. S1F). RNA 43-2 is a spliced 590-nt transcript that encodes only a short CDS domain (21 aa). Expression of both RNA 13.5-1 and RNA 43-2 in VZV-infected ARPE-19 cells was confirmed by RT-PCR (Fig. S1E and K and Fig. S1F and L)

Finally, we used an *in silico* approach to predict the coding potential of all 136 polyadenylated VZV RNAs (Table S3). The Coding Potential Calculator version 2 algorithm (CPC 2.0) (36) calculates the coding probability of a transcript based on its length, isoelectric point, and Fickett score of the longest CDS encoded. Of the 96 VZV RNAs

**FIG 2** Legend (Continued)

and Nanopore dRNA-Seq (teal) coverage plots are derived from ARPE-19 cells lytically infected with VZV strain pOka for 96 h. y axis values indicate the maximum read depth of that track. See also Fig. S1 and S2. Reiterative repeat regions R1 to R5 and both copies of the OriS are shown as black boxes embedded in the genome track.

encoding the 71 canonical ORFs, 89 were assigned a coding probability exceeding 90%, with only two canonical RNAs—encoding the two smallest VZV proteins, pORF49 (81 aa) (37) and pORF57 (71 aa) (38)—incorrectly predicted to be a noncoding transcript (Table S3). Of the 40 VZV RNAs encoding noncanonical products, 17 were predicted to be noncoding (Table S3), including two novel transcripts: RNA 13.5-1 (7%) and RNA 43-2 (13%). Notably, only RNA 43-2 was expressed to relatively high levels as the 37th most abundant transcript at 24 hpi and 41st most abundant at 96 hpi (out of 136 VZV transcripts).

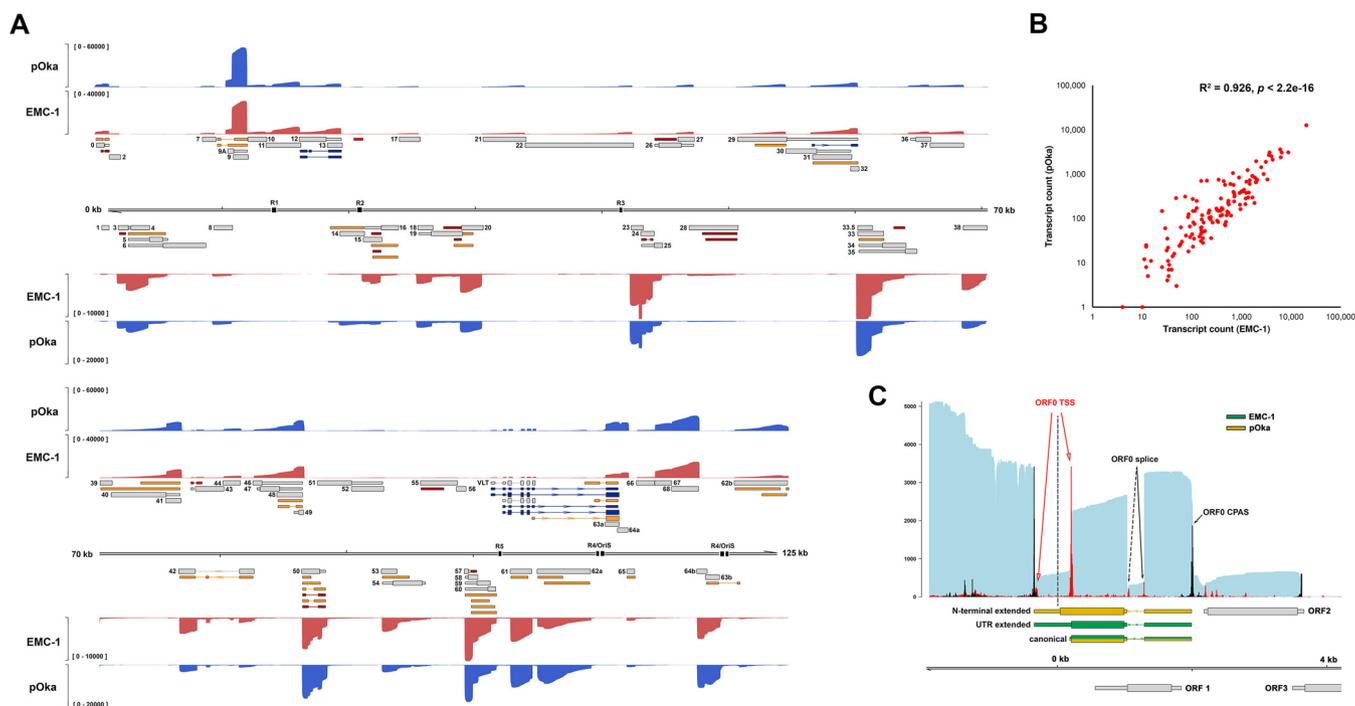
We next validated a selection of the reported transcript variants through comparative analysis of our sequencing-derived data against Northern blotting approaches used in prior studies. For instance, prior Northern blot analysis of the VZV RNA 50 locus (34) identified at least six discrete bands (2,100, 1,800, 1,400, 1,200, 1,000, and 900 nt), five of which corresponded to six transcripts observed in our data set (1,825, 1,400, 1,362, 1,162, 1,039, and 839 nt). We note that no annotated transcript in our data matched the observed 2,100-nt band, nor was any band observed that matched our 663-nt transcript variant. Similarly, Northern blotting of the VZV RNA 9 locus (39) demonstrated the presence of at least three bands, sized approximately 2,400, 1,600, and 1,100 nt which correspond with RNA 9-1 (1,030 nt), RNA 9A-1 (1,530 nt), and the unspliced precursors for RNA 9-2 and 9-3 (2,373 nt). Note that the presence of mature spliced RNA 9-2 (1,556 nt) and 9-3 (1,814 nt) was confounded by their size relative to RNA 9A-1. Finally, we performed Northern blotting to confirm the expression of the 698-nt RNA 64-1 and 1,653-nt polycistronic RNA 63-64, as well as a low-abundance  $\pm$ 400-nt RNA 64 variant that was detected, but not sufficiently evidenced for including in the annotation, in our data sets (Fig. S1M).

**The lytic VZV transcriptome is not influenced by viral strain or cell type.** VZV genome sequences are highly conserved (40), suggesting that strain-specific differences in coding capacity are likely minimal. To test this hypothesis, we infected ARPE-19 cells with VZV EMC-1 for 96 h and sequenced the poly(A) fraction of RNA by dRNA-Seq (Table S4). This enabled a comparative analysis of data sets obtained from ARPE-19 cells lytically infected with either VZV pOka or EMC-1 to determine whether either strain encodes unique transcripts (Fig. 4). No such transcripts were identified at the RNA level, although we note that nucleotide level changes may still impact encoded proteins—as is exemplified by the N-terminal extended pORF0 uniquely present in pOka (Fig. 4C).

As VZV is capable of infecting diverse cell types including epithelial cells and neurons, we also determined if the VZV transcriptome remains similar between VZV pOka-infected ARPE-19 cells and human embryonic stem cell (hESC)-derived neurons (41, 42). Again, no cell-type-specific novel VZV RNAs or variant VZV RNAs were identified (Fig. 5), indicating that observed differences in VZV RNA expression levels and infectivity in distinct cell types (41, 43) are not due to the presence of cell-type-specific VZV RNAs but are likely driven by host cell factors.

**Overexpression of RNA 43-2 does not impair VZV replication in epithelial cells.** Based on the high relative expression of RNA 43-2, its low protein coding potential, and genomic location, we hypothesized that it may function as an ncRNA involved in regulating the expression of the longer RNA 43-1 (encoding pORF43). RNA 43-1 encodes pORF43 and has previously been designated as an essential gene (44), which putatively encodes the capsid vertex component 1 and is postulated to be important for viral DNA encapsidation from the analogy to HSV UL17 (45). To test this hypothesis, we generated three stable ARPE-19 cell lines expressing RNA 43-2 or an empty vector as control (Fig. S2A and B) and analyzed VZV EMC-1 replication at 48 and 72 hpi by flow cytometry and plaque assay. We did not observe any significant differences in number of VZV-infected cells, number of plaques, or plaque sizes between RNA 43-2-expressing cells and vehicle control cells (Fig. S2C and D). Additionally, RNA 43-2 expression did not significantly reduce expression of RNA 43-1 in VZV-infected cells ( $P = 0.08$  by Student's  $t$  test) (Fig. S2E).

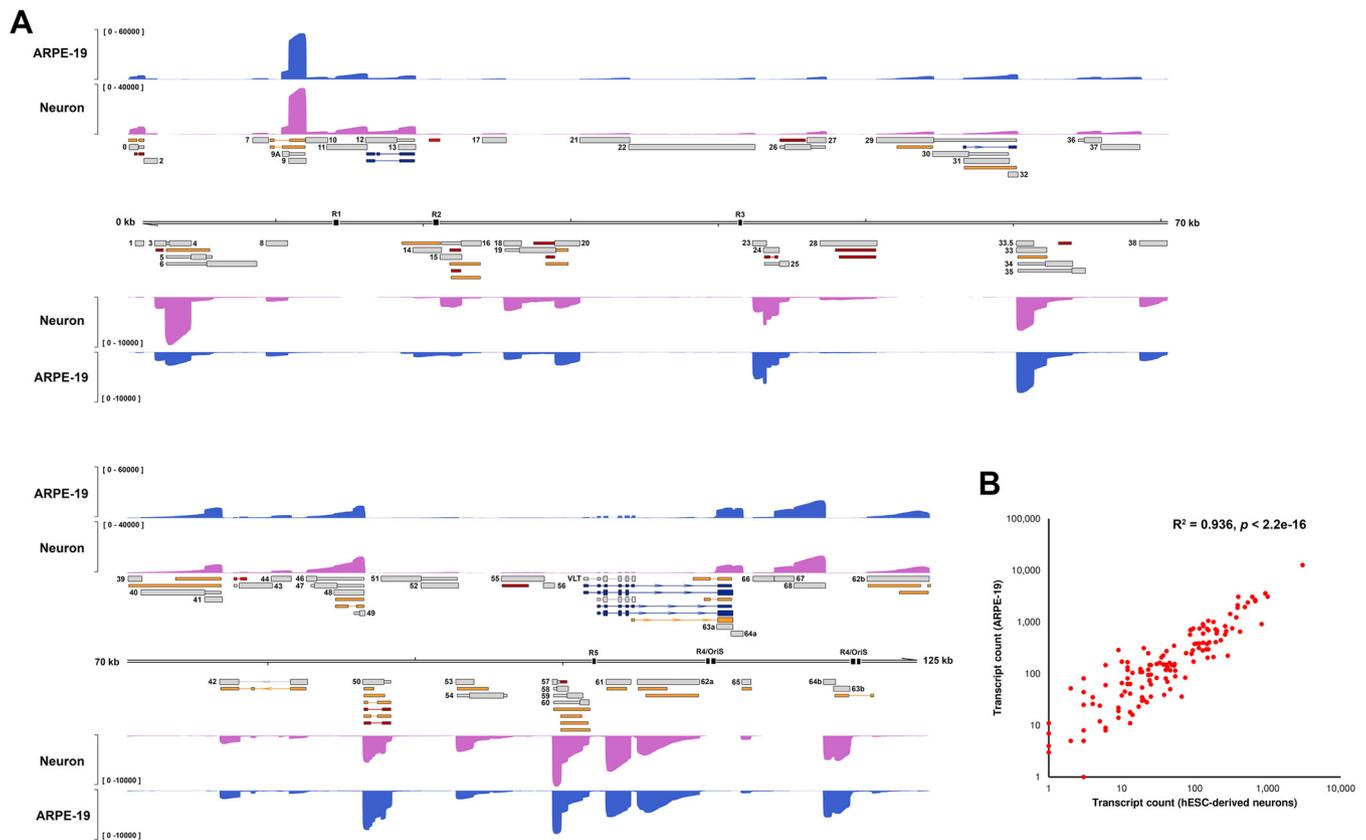
**Decoding the kinetic class of lytic VZV transcripts.** To determine the kinetic class of each lytic VZV RNA, cell-cycle-synchronized ARPE-19 cells were infected with cell-free



**FIG 4** VZV transcriptome diversity is consistent between viral strains. (A) Coverage plots derived from Nanopore dRNA-Seq of VZV strain pOka (blue)- and VZV strain EMC-1 (red)-infected ARPE-19 cells at 96 hpi. Note that while relative and absolute abundances of distinct VZV RNAs differ between strains, all RNAs included in our annotation are represented. y axis denotes coverage range. Canonical (gray), alternative (orange), fusion (dark blue), and putative (red) ncRNAs are shown with canonical CDS regions indicated by wide boxes and UTRs shown as thin boxes. The absence of a CDS region indicates that an RNA has uncertain coding potential. The reiterative repeat regions R1 to R5 and both copies of the OriS are shown as black boxes embedded in the genome track. (B) The scatterplot shows the correlation between VZV transcript abundance in VZV EMC-1-infected and VZV pOka-infected ARPE-19 cells. Pearson  $R^2$  and  $P$  value are indicated. (C) Identical TSS and CPAS were detected in ARPE-19 cells infected with cell-free VZV pOka (VZV clade 2; yellow) or EMC-1 (VZV clade 1; green) at 96 hpi, yielding two ORF0 RNA isoforms. However, due to nucleotide variability between the two strains, VZV pOka RNA 0-1 includes a larger CDS and can encode an alternative ORF0 protein variant, compared to VZV EMC-1. Nanopore dRNA-Seq (teal) coverage plots are shown. y axis values indicate the maximum read depth of that specific track. CDS regions and UTRs are shown as wide and thin boxes, respectively.

VZV EMC-1 and cultured for 12 or 24 h in the presence or absence of actinomycin D (ActD; transcription inhibitor), cycloheximide (CHX; translation inhibitor), or phosphonoacetic acid (PAA; inhibitor of the viral DNA polymerase), and viral RNAs were subsequently profiled using dRNA-Seq and RNA-Seq (Fig. 6 and Fig. S3). Transcription of *IE* RNAs is not dependent on *de novo* viral protein production, whereas transcription of *E* RNAs depends on *IE* proteins. *L* RNAs are further subclassified into two kinetic classes, *Leaky-Late* (*LL*) and *True-Late* (*TL*); *LL* RNAs are expressed at very low levels before, and *TL* RNAs exclusively after, viral DNA replication has commenced. To rule out the possibility of contamination introduced through cell-free infections, we also sequenced samples treated with ActD to control for the potential presence of residual background transcripts in the virus preparations and confirmed only minimal amounts of VZV transcripts to be present (Fig. S3).

Given the very high sensitivity of our Illumina RNA-Seq and Nanopore dRNA-Seq analyses, many VZV transcripts were detected under all experimental conditions albeit at vastly different abundancies (Table S3). To classify VZV transcripts into distinct kinetic classes in an unbiased fashion, we determined the relative expression of each of the 136 VZV RNAs in CHX-treated, PAA-treated, or untreated samples and clustered transcripts by average correlation (Fig. 7). In total, 8 transcripts were classified as *IE* and 53 were classified as *E* RNAs, including the experimentally validated transcripts encoding pORF28 and pORF29 (46). A further 28 transcripts were classed as *LL* and 41 transcripts were classed as *TL*, the latter including both RNA 14-1 (pORF14) and VLT (pVLT), both of which have been experimentally confirmed previously (9, 47). Importantly, 4 transcripts were classified as *Transactivated/True-Late* (*TA/TL*)—a putative new designation applied to transcripts that show detectable yet limited expression in CHX-treated



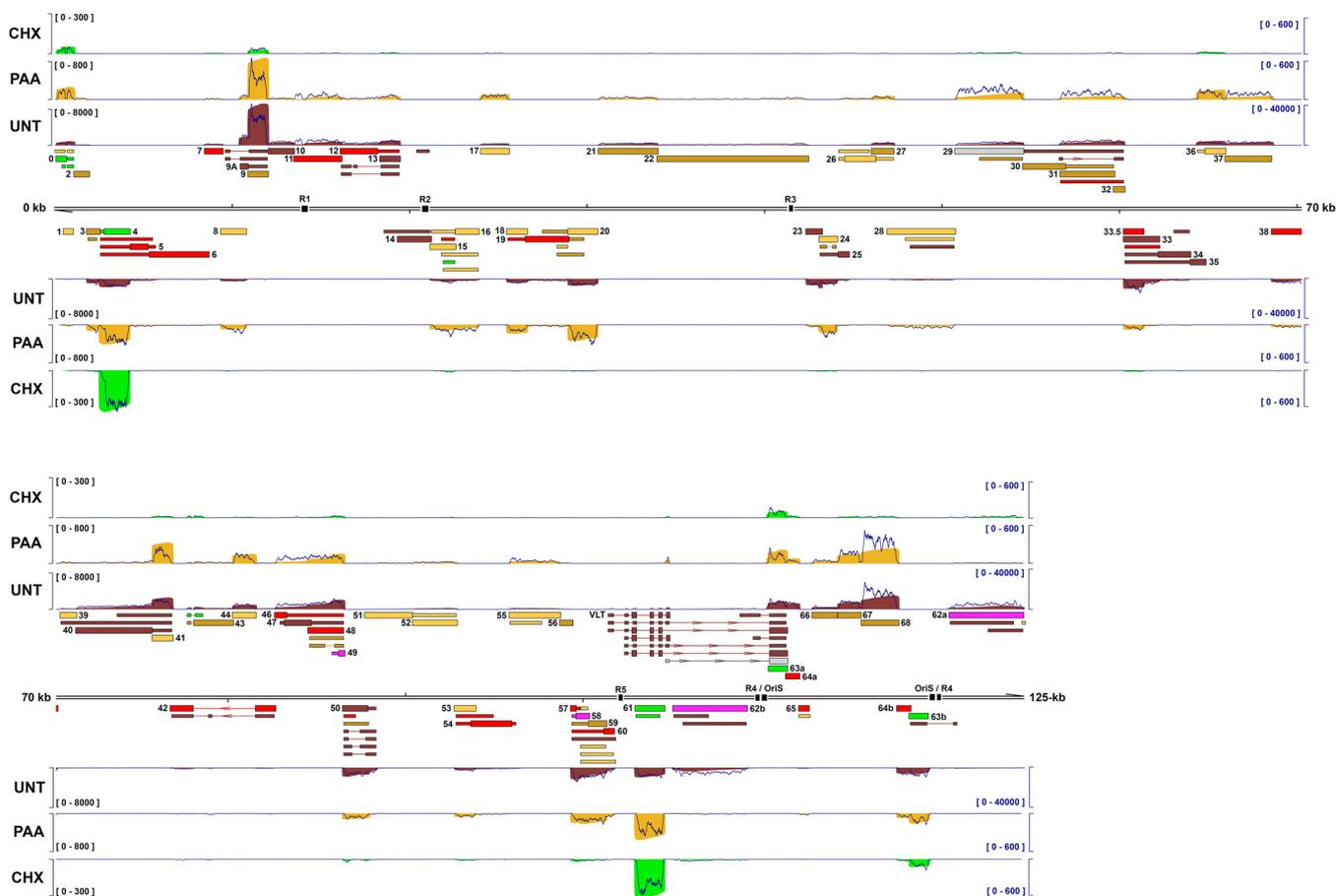
**FIG 5** VZV transcriptome diversity is consistent between infected cell types. (A) Coverage plots derived from Nanopore dRNA-Seq of VZV pOka-infected ARPE-19 cells (blue) and hESC-derived neurons (purple). RNA was collected at 96 hpi (ARPE-19 cells) or 144 hpi (neurons). Note that while relative and absolute abundances of distinct VZV RNAs differ between cell types, all RNAs included in our annotation are represented. y axis denotes coverage range. Canonical (gray), alternative (orange), fusion (dark blue), and putative (red) ncRNAs are shown with canonical CDS regions indicated by wide boxes and UTRs shown as thin boxes. The absence of a CDS region indicates that the respective VZV RNA has an uncertain coding potential. Reiterative repeat regions R1 to R5 and both copies of the OriS are shown as black boxes embedded in the genome track. (B) The scatterplot shows the correlation between VZV transcript abundance in hESC-derived neurons and ARPE-19 cells. Pearson  $R^2$  and  $P$  value are indicated.

samples, no or minimal expression in PAA-treated samples, and high expression in untreated samples at 24 hpi. Three VZV transcripts clustered closely with *IE* RNAs but showed variable expression patterns. One of these, RNA 0-1 (encoding pORF0), shares features with both *IE* and *E* transcripts and was provisionally classed as *IE* based on its relatively high abundance in CHX-treated samples (Table S3), whereas  $_{1\gamma T}$ VLT63-3 (encoding pORF63) appears to be transactivated during infection and then expressed with *E* kinetics (*TA/E*), and RNA 29-1 showed an expression pattern similar to *E* transcripts. Finally, we also note that *E* transcripts could be classified into two distinct subgroups termed *True-Early* and *Early-Late*, dependent on whether their expression was reduced or consistent in untreated samples compared to PAA-treated samples.

Collectively, our approach showed that two VZV transcripts (RNA 4-1 and RNA 61-1, encoding pORF4 and pORF61, respectively) were expressed at very high levels (accounting for 60% of all VZV transcripts in CHX-treated samples [Table S3]) in the absence of *de novo* protein production and, in agreement with prior studies (23, 25), were classified as *IE* transcripts (Fig. 6 and 7). Six additional transcripts—including both copies of RNA 63-1 (encoding pORF63)—were expressed to high levels in CHX-treated samples relative to other conditions and were also assigned *IE* status, albeit at much lower levels compared to RNA 4-1 and RNA 61-1. pORF4, pORF61, and pORF63 are known transcriptional activators of VZV and canonical *IE* transcripts (22, 23, 25, 48).

#### ORF62 transcripts are expressed with *Late* kinetics during lytic VZV infection.

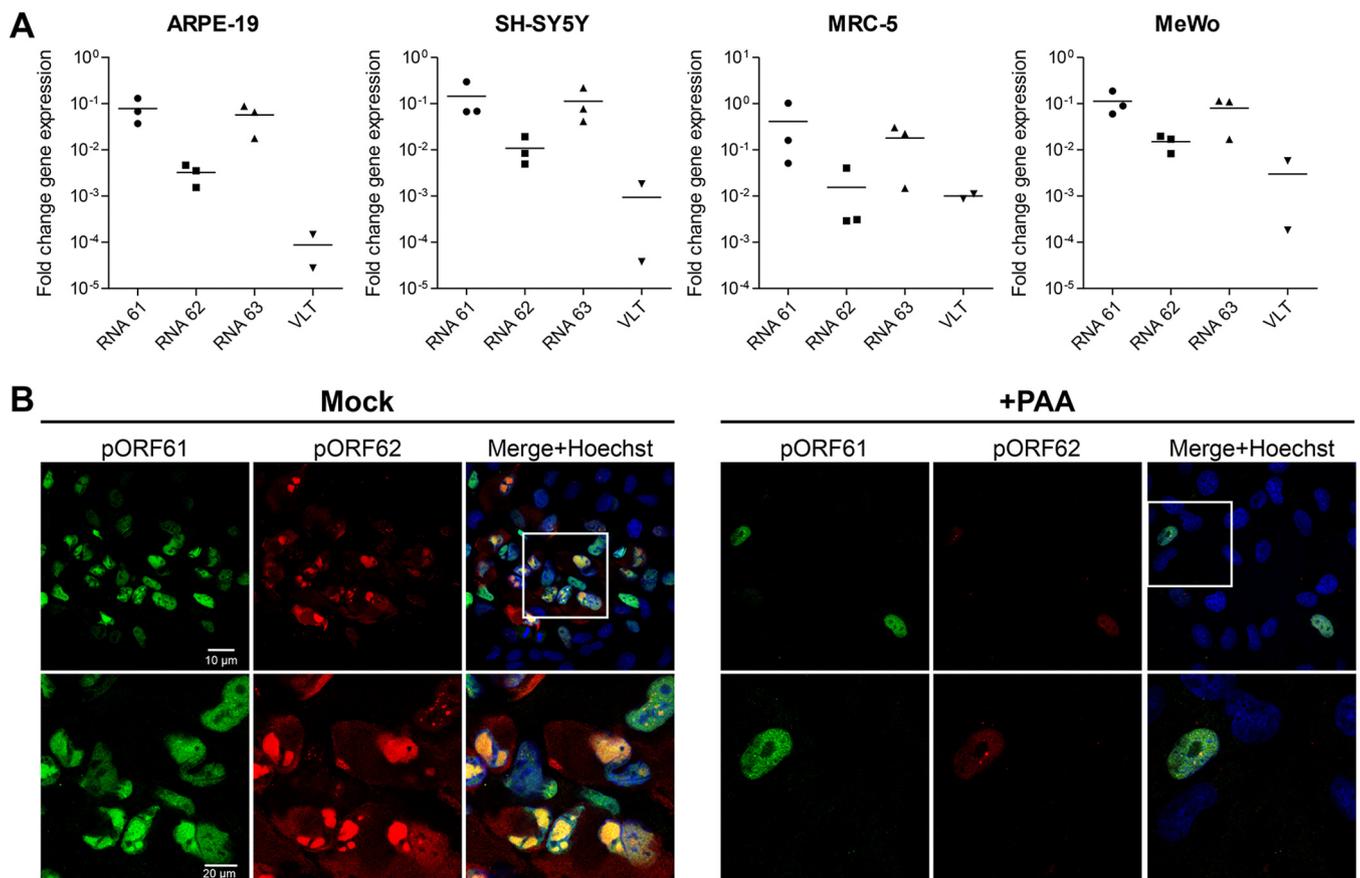
The VZV gene locus 62 is located within a short-repeat region that is duplicated and sits on either side of the Unique Short (US) region. The gene 62 duplications (62a and 62b)



**FIG 6** Decoding the kinetic class of VZV transcripts during lytic infection. Coverage plots derived from dRNA-Seq of ARPE-19 cells infected with cell-free VZV EMC-1 and treated with cycloheximide (CHX; green) for 12 h or phosphonoacetic acid (PAA; gold) for 24 h or untreated (UNT; red) for 24 h. Corresponding Illumina RNA-Seq coverage plots for each condition are shown as dark blue line plots. y axis denotes coverage range (left side, Nanopore dRNA-Seq; right side, Illumina RNA-Seq). Reannotated VZV genome is shown in the middle track. CDS regions and UTRs are shown as wide and thin boxes, respectively. Reiterative repeat regions R1 to R5 and both copies of the OriS are shown as black boxes embedded in the genome track. Transcripts annotated are colored according to kinetic class: *IE*, green; *TE*, yellow; *EL*, dark gold; *LL*, red; *TL*, dark red; *TA/TL*, purple; *Not classified*, gray. Canonical CDS regions are indicated by wide boxes with UTRs shown as thin boxes. Absence of a CDS region indicates that an RNA has an uncertain coding potential. Reiterative repeat regions R1 to R5 and both copies of the OriS are shown as black boxes embedded in the genome track.

are considered perfect in that even full-length sequence reads align against both copies with an equal match. VZV transcripts RNA 62-1 and RNA 62-2 encode, respectively, the major viral transcriptional activator protein, pORF62, and a predicted N-terminal truncated pORF62 variant. A third transcript (RNA 62-3, *TL*) shows a 3' truncation that disrupts the CDS through the absence of a stop codon. While this transcript is predicted to encode several putative micro-ORFs, it is not known whether any of these are translated. Surprisingly, our data indicate that following very limited transactivation during initial infection, expression of these transcripts is dependent on *de novo* (viral) protein synthesis and viral DNA replication, thereby classifying these RNA 62 transcripts as *TA/TL*. This contradicts the current classification of RNA62-1 (encoding pORF62) as an *IE* transcript, although we note that this classification was obtained by analogy to the function of its HSV-1 orthologue infected cell polypeptide 4 (ICP4) (49). To confirm and substantiate our findings, we analyzed the impact of viral DNA replication on pORF62-encoding RNA and protein expression in multiple VZV-susceptible cell types at 24 hpi. RT-qPCR analysis showed that PAA treatment was associated with an approximately 10-fold decrease in expression of *IE* transcripts encoding pORF61 and pORF63, consistent with the absence of VZV DNA replication and spread in culture, and a >10,000-fold decrease in expression of *TL*, viral DNA replication-dependent, RNA<sub>lyt</sub>VLT (Fig. 8A). Notably, expression of RNAs encoding pORF62 was more severely affected (~500-fold





**FIG 8** RNA 62 transcripts are expressed with *Late* kinetics during lytic VZV infection. Cell cycle-synchronized human ARPE-19 (epithelial), SH-SY5Y (neuroblastoma), MRC-5 (fibroblast), and MeWo (melanoma) cells were infected with cell-free VZV EMC-1 for 24 h in the presence or absence of PAA. (A) Fold change in expression of transcripts encoding pORF61, pORF62, and pORF63 in PAA-treated samples relative to untreated samples (calibrator) and normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) using the  $2^{-\Delta\Delta CT}$  method. In 1 of 3 experiments, PAA treatment reduced VLT expression to below detection limit. qPCR primer/probe set directed to the ORF62 CDS recognizes all three transcript variants (RNA 62-1, 62-2, and 62-3). (B) Representative images of lytically VZV-infected ARPE-19 cells stained by immunofluorescence for pORF61 (green) and pORF62 protein (red). Nuclei were counterstained with Hoechst 33342 (blue). Top row,  $\times 400$  magnification; bottom rows,  $\times 1,000$  magnification.

from the same locus and fusion RNAs that modify UTRs and CDS of multiple transcripts are produced. Given that 5' UTR sequences influence the translational efficiency of the downstream CDS (50), alternative UTR usage may provide the virus with an additional mechanism to regulate its protein expression throughout its infectious cycle. Collectively, this study defined 136 polyadenylated VZV RNAs that are expressed during lytic VZV infection, many of which are predicted to increase the diversity of the viral proteome.

Although the VZV genome is considered relatively stable, multiple strains currently cocirculate and recombine (40, 51), potentially influencing the viral transcriptome. However, our comparison of the transcriptional landscape of a VZV clade 1 (strain EMC-1) and a clade 2 (strain pOka) virus revealed that no strain-specific lytic transcript isoforms exist in VZV-infected ARPE-19 cells. However, interstrain differences in repeat variations could nevertheless impact transcription of RNA 11-1 (containing R1), RNA 13.5-1 (R2), RNA 14-1 (R2), RNA 22-1 (R3), RNA 63-2 (R4), RNA 63-3 (R4), and all RNA 59 and RNA 60 isoforms (R5) (35). Similarly, strain-specific polymorphisms may function to extend coding domains such as the N-terminal extended pORF0 (RNA 0-3) in VZV pOka in comparison with other VZV clades (Fig. 4C). Additionally, while previous studies suggested that the VZV transcriptome is generally similar across diverse cell types, none had sufficient resolution or used the methodologies required to disentangle transcript structures (9, 43, 52). Here, we demonstrated that identical transcript isoforms were detected during lytic VZV infection of human retinal pigment epithelial cells and hESC-derived neurons. Thus, while VZV strain-specific polymorphisms and/or cell type

may influence viral gene expression, their impact on the lytic VZV transcriptome structure appears to be small.

Many herpesviruses express ncRNAs during lytic and latent infections (53). Previously, we identified the putative dual-function polyadenylated VZV RNA VLT, which encodes a protein expressed during lytic infection but is also functional as an RNA, inhibiting ORF61 RNA expression in overexpression experiments (9). Here, we identified 17 additional polyadenylated VZV RNAs that are predicted to be noncoding. Diverse functions have been attributed to human ncRNAs, including the modification of antisense or overlapping transcription events (54, 55). The generation of functional VZV mutant viruses with impaired expression of identified putative ncRNAs is challenging due to overlap with other viral transcripts. Therefore, we have studied the function of VZV RNA 43-2 by means of RNA 43-2 overexpression followed by VZV superinfection. However, RNA 43-2 did not significantly reduce expression of the overlapping RNA 43-1 transcript nor influence VZV replication. These results most likely reflect the complexity of VZV transcript regulation during lytic infection, as putative ncRNA 43-2 is expressed earlier (*IE* kinetics) and at higher abundance than RNA 43-1 (*E* kinetics) (Fig. 6 and 7 and see also Fig. S2F and G in the supplemental material). Considering the multitude of ncRNAs expressed by other herpesviruses and their crucial roles during infection (53), delineating the functional importance of VZV ncRNAs should be considered a research goal.

Twenty-eight of 136 VZV transcripts are (multiply) spliced. Strikingly, the majority of splicing events occur in a hypercomplex region of the VZV genome encoding both pVLT and pORF61. We have previously shown this locus to be characterized by extensive alternative splicing (9) and here report the presence of multiple transcripts that variously encode a fusion protein of pVLT and pORF63 (pVLT-ORF63) or an N-terminal extended pORF63 (pORF63-N+). The coding potential of pVLT-ORF63 or pORF63-N+ and functional consequences of (and requirement for) these fusion transcripts during reactivation from latency are described in a separate study (78), while their functional role(s) during lytic infection is under investigation. Except for <sub>lyt</sub>VLT63-3, all spliced VZV RNAs used the canonical splice donor (GT) site. The splice donor sites were highly enriched for C/A (−3 position), A (−2), G (−1), and A (+1). All spliced VZV RNAs used the canonical splice acceptor site (AG), often flanked by C (−1) and G/A (+1). Consistent with the use of the cellular splicing machinery to process viral pre-mRNAs, VZV consensus splice donor and acceptor sites closely resemble those of the human transcriptome (56).

Herpesvirus transcripts are traditionally assigned kinetic classes based on their temporal expression pattern and dependence on *de novo* protein synthesis or viral DNA replication. Here, we provide an unbiased transcriptome-wide classification of VZV transcripts during lytic infection of epithelial cells and observed novel patterns of viral RNA expression, designated *TE*, *EL*, and *TA/TL*, in addition to traditional *IE*, *LL*, and *TL* kinetic classes. The data suggest that a subset of the relative expression of *TE* transcripts is reduced following viral DNA replication, whereas this is not the case for *EL* transcripts. Notably, *TE* VZV RNAs are enriched for transcripts encoding proteins involved in viral DNA replication (e.g., RNA 16 variants, RNA 28-1, and RNA 51-1 encoding the viral DNA polymerase processivity factor pORF16, DNA polymerase large subunit pORF28, and origin-binding protein pORF51, respectively), whereas *EL* RNAs encode a more diverse panel of VZV proteins also involved in virion formation and egress (e.g., RNA 31-1 and 37-1 encoding glycoproteins B and H, respectively). *TA/TL* RNAs were detectable at low levels in the absence of *de novo* protein production, but their abundant expression was highly dependent on viral DNA replication. We hypothesize that transcription of *TA/TL* transcripts in CHX-treated samples is due to low-level transactivation by viral tegument proteins delivered from incoming virions. Additionally, our kinetic classification identified that multiple transcripts originating from the same locus could either share the same kinetic class (e.g., RNAs 16-1, 16-2, and 16-3) or be of different kinetic classes (e.g., RNAs 9-1, 9-2, 9-3, and 9A-1) (Table S3 and Fig. S3). While the biological impetus for this is not clear, it seems likely that TSS and CPAS usage are dynamically regulated during lytic infection. For example, RNA 43-1 (*E*, pORF43) and the putative ncRNA 43-2 (*IE*) diverge in

CPAS usage and expression kinetics (Fig. 7 and Fig. S3). Similarly, multiple transcripts encoding pORF63 utilizing different TSS are expressed as different temporal classes and various abundancies, with canonical *IE* RNA 63-1 being most abundant, followed by *E* RNA<sub>lyt</sub> VLT63-3 and low quantities of two *TL* RNAs, 63-2 and 63-3 (Fig. 7 and Fig. S3).

Finally, our data provide novel insight into the expression of *IE* transcripts and the role of pORF62 during lytic VZV infection. The most abundantly expressed VZV *IE* transcripts, produced in the absence of new protein synthesis, encode pORF4 and pORF61, which are also the earliest proteins detected during lytic VZV infection (27). Interestingly, prior studies of the RNA 4-1 and RNA 61-1 promoter regions have shown that their efficient transactivation is dependent on pORF62 (57, 58), a major component of the viral tegument (59). However, our data indicate that abundant expression of VZV RNA 62-1 (pORF62), as well as RNAs 62-2 and 62-3, is dependent on viral DNA replication (*TL* and *TA/TL* kinetics), suggesting that tegument-derived pORF62 but not *de novo* pORF62 transactivates RNA 4-1 and RNA 61-1 expression, at least during establishment of the lytic infection cycle. Classification of RNA 62-1 (pORF62) as *TA/TL* is also supported by prior observations that a marked increase in pORF62 abundance occurs only after DNA replication has commenced (19). Importantly, our findings do not exclude any of the previously assigned functions of pORF62, most notably its function as a major transcriptional regulator of VZV transcripts (60–62). However, future studies aimed at better understanding the regulation of VZV transcript expression, and the distinct roles of newly produced RNA 62 isoforms, pORF62, and tegument-derived pORF62, are warranted.

In summary, this study describes the detailed analysis of the VZV transcriptome architecture and kinetic classification of viral transcripts in the context of lytic VZV infection. We provide these data as a comprehensive resource that will facilitate functional studies of coding RNAs and their protein products, the role of noncoding RNAs, and the regulation of VZV transcription and translation during lytic infection.

## MATERIALS AND METHODS

**Cells and viruses.** Human retinal pigmented epithelium ARPE-19 cells (American Type Culture Collection [ATCC] CRL-2302) were grown in a 1:1 (vol/vol) mixture of Dulbecco's Modified Eagle's medium (DMEM) (Lonza) and Ham's F-12 (Gibco) medium supplemented with 10% heat-inactivated fetal bovine serum (FBS; Lonza) and 0.6 mg/ml L-sodium glutamate (Lonza) or in DMEM/F-12 + GlutaMAX-1 (Thermo Fisher Scientific) supplemented with heat-inactivated 8% FBS (Sigma-Aldrich). Human neuroblastoma SH-SY5Y cells were grown in a 1:1 (vol/vol) mixture of Eagle's Minimum Essential Medium (EMEM) with Earle's balanced salt solution (EBSS) (Lonza) and Ham's F-12 (Gibco) medium supplemented with 15% FBS (Lonza), L-sodium glutamate, penicillin-streptomycin, nonessential amino acids (MP Biomedicals), and sodium-bicarbonate (Lonza). MRC-5 human embryonic lung fibroblasts and human melanoma MeWo cells were cultured in DMEM (Lonza), supplemented with 10% FBS, L-sodium glutamate, and penicillin-streptomycin. Human embryonic stem cell (hESC; H9)-derived neural stem cells (NSC) (Thermo Fisher Scientific) were cultured, propagated, and differentiated into neurons as described previously (63). Cell cultures were maintained at 37°C in a humidified CO<sub>2</sub> incubator. VZV strain pOka (parental Oka) was maintained in, and the cell-free virus was prepared from, ARPE-19 cells as described previously for MRC-5 cells (37). VZV strain EMC-1 is a low-passage-number clinical isolate and was cultured on ARPE-19 cells, and cell-free VZV was extracted as described previously (20, 64). Cell-free EMC-1 was freshly harvested on the day of use and pretreated with DNase I, RNase T<sub>1</sub>, and RNase A (all from Thermo Fisher Scientific) for 30 min at 37°C prior to infection.

**RNA extraction and cDNA synthesis.** ARPE-19 cells were infected with cell-associated VZV EMC-1 by cocultivation of uninfected and VZV EMC-1-infected ARPE-19 cells at an 8:1 cell ratio for 96 h. Alternatively, ARPE-19, SH-SY5Y, MRC-5, and MeWo cells were infected with cell-free VZV EMC-1 for indicated times. Cells were harvested in 1 ml TRIzol (Thermo Fisher Scientific), mixed with 200  $\mu$ l chloroform, and centrifuged for 15 min at 12,000  $\times$  g at 4°C. RNA was isolated from the aqueous phase using the RNeasy minikit (Qiagen) according to manufacturer's instructions, including on-column DNase I treatment, as described previously (65). RNA concentration and integrity were analyzed using a Nanodrop spectrophotometer (Thermo Fisher Scientific), and RNA was subjected to a second round of DNase treatment using the Turbo DNA-free kit (Ambion) according to manufacturer's instructions. For cDNA synthesis, a maximum of 5  $\mu$ g RNA was reverse transcribed using Superscript IV reverse transcriptase and oligo(dT) primers (Thermo Fisher Scientific) (RT<sup>+</sup>). As control, the same reaction was performed without reverse transcriptase (RT<sup>-</sup>). Alternatively, RNA was isolated using the FavorPrep blood/cultured cell total RNA minikit (Favorgen Biotech) in combination with the NucleoSpin RNA/DNA buffer set (Macherey-Nagel). DNA was first eluted from the column in 100  $\mu$ l DNA elution buffer, subsequently the column was treated with recombinant DNase I (20 units/100  $\mu$ l; Roche Diagnostics) for 30 min at 37°C, and finally RNA was eluted in 50  $\mu$ l nuclease-free water. RNA was directly treated with Baseline-Zero

DNase (2.5 units/50  $\mu$ l; Epicentre) for 30 min at 37°C. cDNA was synthesized with 12  $\mu$ l of RNA and anchored oligo(dT)<sub>18</sub> primer in a 20- $\mu$ l reaction mixture using the Transcriptor first-strand cDNA synthesis kit at 55°C for 30 min for reverse transcriptase reaction (Roche Diagnostics).

**PCR and sequence analysis.** PCR was performed on RT<sup>+</sup> and RT<sup>-</sup> cDNA reaction mixtures using AmpliTaq Gold DNA polymerase (Thermo Fisher Scientific) and primer pairs corresponding to each newly identified VZV transcript (see Table S5 in the supplemental material). Primers were directed to the predicted 5' and 3' ends of each transcript so that newly identified transcripts were completely amplified from the 5' to the 3' end. For ORF9 variants and ORF48, additional reverse primers were used to confirm splice junctions (Table S5). PCR amplification was performed as follows: initial denaturation at 95°C for 10 min, followed by 40 cycles of alternating denaturation (30 s, 95°C), primer annealing (30 s at the appropriate temperature) (Table S5), and subsequently primer extension (1 min/1,000 bp, 72°C) (Table S5). A final extension step of 10 min at 72°C was included. To amplify ORF13.5, each deoxynucleoside triphosphate (dNTP), including equimolar amounts of dGTP and 7-deaza-GTP (New England Biolabs), was used at a concentration of 200  $\mu$ M (66). PCR amplification of ORF13.5 was performed as follows: initial touchup PCR from 58°C to 70°C using a transcript-specific forward primer and an anchored primer on the poly(A) tail. Subsequently, seminested PCR was performed using the same forward primer and a reverse primer within the transcript using standard PCR protocol. Amplicons were purified from gel using the QIAquick gel extraction kit (Qiagen) and sequenced using the BigDye v3.1 cycle sequencing kit (Applied Biosciences) with corresponding forward and reverse primers on the ABI Prism 3130 XL genetic analyzer.

**Plasmid construction and generation of stable cell lines.** The RNA 43-2 transcript sequence (77775 to 78619 excluding the intron at 77869 to 78149; strain Dumas, NC\_001348.1) was amplified using cDNA from VZV EMC-1-infected ARPE-19 cells and primers NheI\_ORF43-5\_Fw and XhoI\_ORF43.5\_Rv (Table S5). Amplicon was digested with NheI and XhoI and cloned into pcDNA3.1. Three independent batches of ARPE-19 cells were transfected with either pcDNA3.1/empty or pcDNA3.1/RNA 43-2 using polyethylenimine (PEI). After 2 days, cells were incubated with 1 mg/ml Geneticin and cultured for at least 3 weeks to select for transfected cells. Subsequently, DNA was isolated using the QIAamp DNA minikit according to manufacturer's instructions and the presence of RNA 43-2 DNA was confirmed by PCR. Next, RNA was isolated as described above, and expression of RNA 43-2 was confirmed by RT-PCR.

**Flow cytometry.** ARPE-19 cells stably transfected with pcDNA3.1/empty or pcDNA3.1/RNA 43-2 were plated 1 day prior to infection in a 48-well plate. Cells were infected with cell-free VZV EMC-1 (multiplicity of infection [MOI] = 0.01), harvested at 48 hpi or 72 hpi, fixed and permeabilized with BD Cytotfix/Cytoperm, stained for VZV glycoprotein E (gE) (MAB8612; Millipore) in BD PermWash, and labeled with secondary allophycocyanin (APC)-conjugated goat anti-mouse Ig antibody (BD Biosciences). The frequency of VZV-infected (i.e., APC-positive) cells was measured on a BD FACSLyric flow cytometer and analyzed using FlowJo software (BD Biosciences).

**Plaque assay.** Confluent monolayers of stable pcDNA3.1-empty or pcDNA3.1-RNA 43-2 cells grown in a 12-well plate were infected with 2,000 PFU/well VZV EMC-1. At 72 h postinfection, plates were washed and fixed with 4% paraformaldehyde (PFA) in PBS. Subsequently plates were permeabilized using 0.1% Triton X-100 in PBS for 10 min, blocked with 5% normal goat serum in PBS-0.05% Tween 20 (PBS-T) for 30 min, and incubated with mouse anti-VZV gE antibody (MAB8612) diluted in PBS-T containing 0.1% bovine serum albumin (BSA) for 1 h at room temperature. Cells were washed with PBS-T, incubated for 1 h with polyclonal rabbit anti-mouse Ig antibody (Dako) in PBS-T + 0.1% BSA, washed, and stained with Alexa Fluor 488 (AF488)-conjugated goat anti-rabbit Ig (H+L) antibody (Thermo Fisher) in PBS-T. Plates were measured using the Immunospot S6 Ultimate UV image analyzer, and plaque size was determined using Immunospot software (Cellular Technology Limited).

**Kinetic class of VZV transcripts.** ARPE-19 cells were synchronized in the cell cycle using a double thymidine block approach (67). Briefly, ARPE-19 cells were seeded at semiconfluence in 12-well plates, and on the next day medium was replaced with growth medium containing 2 mM thymidine (Sigma-Aldrich). After 24 h, medium was replaced with normal growth medium for 8 h, after which medium was changed back to thymidine-containing medium. Thirty minutes prior to infection, cells were released from thymidine by replacing the medium with regular culture medium. Cells were infected with freshly harvested cell-free VZV EMC-1 using spin-inoculation for 15 min at 1,000  $\times$  g (MOI after spin-inoculation = 0.1 to 0.2). Cells were incubated for 45 min at 37°C, after which the inoculum was replaced with fresh medium, medium with 10  $\mu$ g/ml actinomycin D (ActD; Sigma-Aldrich), medium with 50  $\mu$ g/ml cycloheximide (CHX; C4859; Sigma-Aldrich), or medium with 400  $\mu$ g/ml phosphonoacetic acid (PAA; Sigma-Aldrich). Infected cells were harvested in 500  $\mu$ l TRIzol at 12 hpi (ActD and CHX) or 24 hpi (untreated and PAA) for RNA extraction. Alternatively, ARPE-19 cells, MRC-5 cells ( $1 \times 10^5$  cells/well on a 24-well plate), and hESC-derived neurons ( $1 \times 10^5$  cells/well on a 24-well plate as NSC and differentiated for 18 days) were infected with cell-free VZV pOka (20  $\mu$ l of  $4 \times 10^4$  PFU/ml) in the presence or absence of phosphonoformic acid (200  $\mu$ g/ml) (Sigma-Aldrich) for 1 h, the inoculum was replaced with fresh medium with phosphonoformic acid (200  $\mu$ g/ml), and cultures were maintained for 24 h.

**Quantitative PCR analysis.** Quantitative TaqMan real-time PCR (qPCR) was performed in duplicate on RT<sup>-</sup> and RT<sup>+</sup> cDNA using 4 $\times$  TaqMan Fast Advanced master mix (Applied Biosystems) on a 7500 TaqMan PCR system. Primer-probe sets directed to CDS regions of RNAs 61, 62, and 63 and VLT have been described previously (9, 68), and those directed to ORF43 are described in Table S5 in the supplemental material. Alternatively, cDNAs were subjected to qPCR using KOD SYBR qPCR mix (Toyobo) in the StepOnePlus real-time PCR system (Thermo Fisher Scientific) (1  $\mu$ l of cDNA per 10- $\mu$ l reaction mixture). All primer sets used for SYBR green chemistry (Table S5) were first confirmed for the amplification rate (98 to 100%) using 10 to  $10^6$  copies (10-fold dilution) of pOka-BAC genome or VLT plasmid (9) and the lack of nonspecific amplification using water. The qPCR program is as follows; 95°C

for 2 min (1 cycle), 95°C for 10 s and 60°C for 15 s (40 cycles), and 60 to 95°C for a dissociation curve analysis. Data are presented as relative VZV level to cellular beta-actin defined as  $2^{-C_T}$  (threshold cycle [ $C_T$ ] value VZV transcript  $- C_T$  value beta-actin).

**Northern blot analysis.** Biotinylated probe directed to VZV ORF64 (nucleotides 111646 to 111936 in reference strain Dumas, NC\_001348.1) was generated using the MAXIscript kit (Ambion; Thermo Fisher Scientific) according to the manufacturer's instructions. Northern blot analysis was performed using the NorthernMax-Gly kit (Ambion). In brief, 7.5  $\mu$ g total RNA extracted from mock- and VZV EMC-1-infected ARPE-19 cells was separated by electrophoresis on a 1% agarose gel and transferred to a nylon membrane (Brightstart-Plus; Ambion). Membranes were incubated for 15 min at 80°C to cross-link RNA, prehybridized in ULTRAhyb buffer for 30 min at 68°C, and hybridized with biotinylated ORF64-specific probe (0.1 nM diluted in ULTRAhyb buffer) overnight at 68°C. Probe binding was visualized using the AP chemiluminescent blotting kit (KPL) and ChemiDoc MP imaging system (Bio-Rad).

**Immunofluorescence staining.** ARPE-19 cells were plated on glass coverslips in 24-well plates 1 day prior to infection. Cells were inoculated with freshly harvested cell-free VZV EMC-1 in medium with or without 400  $\mu$ g/ml PAA and incubated for 24 h. Infected cells were fixed with 4% PFA, permeabilized for 10 min with 0.1% Triton X-100 in PBS, blocked with 5% goat serum diluted in 0.2% gelatin-PBS solution, and incubated on 30- $\mu$ l 0.2% gelatin-PBS droplets containing primary antibody overnight at 4°C. The following primary antibodies were used: anti-pORF61 antibody (1:1,000, gift from P. Kinchington) and monoclonal mouse anti-pORF62 antibody (1:200) (20). Cells were washed 3 times with 0.2% gelatin-PBS and incubated for 1 h at room temperature with secondary antibodies diluted in 0.2% gelatin-PBS. The following secondary antibodies were used: AF488-conjugated goat anti-rabbit Ig (H+L) antibody (1:500; Thermo Fisher) and AF594-conjugated goat anti-mouse Ig (H+L) antibody (1:500; Thermo Fisher). Cells were washed once with 0.2% gelatin-PBS, washed once with PBS, incubated with a 1:1,000 dilution of Hoechst 33342 (Life Technologies, 20 mM) in PBS for 5 min, washed with PBS, and mounted using Prolong Gold antifade mounting medium (Thermo Fisher). Stained cells were analyzed using a Zeiss LSM 700 confocal laser scanning microscope (Zeiss) with a magnification of  $\times 400$  or  $\times 1,000$ . Photoshop CC 2019 software (Adobe) was used to adjust brightness and contrast.

**Illumina RNA sequencing and analysis.** Stranded RNA libraries were prepared from poly(A)-selected RNA using the NEBNext Ultra II Directional RNA library prep kit for Illumina (New England Biolabs) and sequenced using a NextSeq 550. Sequence reads were trimmed using TrimGalore ([https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)) (–paired –length 30 –quality 30) and aligned against the VZV reference genome (strain Dumas, NC\_001348.1) using BBMAP (<https://sourceforge.net/projects/bbmap/>) with postalignment processing performed using SAMtools (69) and BEDTools (70) to generate BEDGRAPH and BED12 files. Candidate CPAS were identified using ContextMap2 (71) (–aligner\_name bowtie –polyA –strand-specific) with sequence reads aligned to VZV Dumas under default parameters by BWA (72).

**Cap Analysis Gene Expression (CAGE) sequencing (CAGE-Seq) and analysis.** Two biological replicates of total RNA were extracted from ARPE-19 cells infected with VZV pOka for 96 h and CAGE-Seq libraries prepared by DNAform (Yokohama, Japan) and subsequently sequenced using an Illumina NextSeq 550, as previously described (31). Resulting sequence reads were trimmed (–length 30 –q 30 –clip\_R1 1) using TrimGalore prior to alignment against the VZV Dumas genome using BBMAP. Postalignment processing was performed using SAMtools and BEDTools. TSS were identified using the HOMER findPeaks module (–style tss –localSize 100 –size 15). Only TSS present in both biological replicates were retained for analysis.

**Nanopore direct RNA sequencing.** For each biological sample, up to 1,000 ng of poly(A) RNA was isolated from up to 50  $\mu$ g of total RNA using the Dynabeads mRNA purification kit (Invitrogen; catalog no. 61006). Isolated poly(A) RNA was subsequently spiked with 0.3  $\mu$ l of a synthetic Enolase 2 (ENO2) calibration RNA (Oxford Nanopore Technologies Ltd.), and dRNA-Seq libraries were prepared as described previously (10). Sequencing was performed on a MinION Mk1b using R9.4.1 (rev D) flow cells (Oxford Nanopore Technologies Ltd.) for 18 to 44 h (one library per flowcell) and yielded between 720,000 and 1,290,000 reads per data set (Table S4). Raw fast5 data sets were then basecalled using Guppy v3.2.2 (–f FLO-MIN106 –k SQK-RNA002) with only reads in the pass folder used for subsequent analyses. Sequence reads were aligned against the VZV Dumas genome, using MiniMap2 (73) (–ax splice –k14 –uf –secondary=no), with subsequent parsing through SAMtools and BEDTools. Here, sequence reads were filtered to retain only primary alignments (alignment flag 0 [top strand] or 16 [bottom strand]).

**Splice junction correction in dRNA-Seq alignments.** Illumina-assisted correction of splice junctions in RNA-Seq data was performed using FLAIR v1.3 (30) in a stranded manner. Briefly, Illumina reads aligning to the VZV Dumas genome were split according to orientation and mapping strand (–f83 & –f163 [forward] and –f99 & –f147 [reverse]) and used to produce strand-specific junction files that were filtered to remove junctions supported by less than 50 Illumina reads. Direct RNA-Seq reads were similarly aligned to the adenovirus type 5 (Ad5) genome and separated according to orientation (–F4095 [forward] and –f16 [reverse]) prior to correction using the FLAIR correct module (default parameters). FLAIR-corrected alignments were used for all subsequent downstream analyses.

**TSS and CPAS identification in dRNA-Seq data.** TSS and CPAS were identified by parsing SAM files to BED12 files in a strand-specific manner using BEDTools and then truncating each aligned sequence read to its 5' or 3' termini for TSS and CPAS identification, respectively. Peak regions containing TSS and CPAS were identified using the HOMER findpeaks module (–o auto –style tss) using a –localSize of 100 and 500 and –size of 15 and 50 for TSS and CPAS, respectively. TSS peaks were compared against Illumina annotated splice sites to identify and remove peak artifacts derived from local alignment errors around

splice junctions. Thirty-eight putative TSS identified in the dRNA-Seq data set alone were flagged as artifacts and removed. Each of these TSS mapped precisely to a splice acceptor within a spliced RNA, and closer inspection of the reads showed these to result from local alignment processes (10).

**Generating RNA abundance counts from dRNA-Seq data.** Using the updated VZV strain Dumas annotation presented here, we generated a transcriptome database by parsing our GFF3 file to a BED12 file using the `gff3ToGenePred` and `genePredtoBED` functions within UCSCutils (<https://github.com/itsvenu/UCSC-Utils-Download>) and subsequently extracting a fasta sequence for each annotated RNA using the `getfasta` function within BEDTools. dRNA-Seq reads were then aligned against the transcriptome database using parameters optimized for transcriptome-level alignment (`minimap2 -ax map-ont -p 0.99`). RNA abundance counts were generated by counting alignments against a given RNA only if the alignment 5' end was located within the first 50 nt of the RNA and the alignment was not marked as supplementary. Transcript per million (TPM) counts (Fig. 3) were generated by dividing the RNA abundance count for a given transcript by the total number of sequence reads (one read = one RNA) present in the data set and subsequently multiplying by 1 million. Note that where raw abundance counts were <3, these counts were zeroed out to minimize possible artifact introduced by low-level noise.

**In silico prediction of coding potential.** CPC 2.0 (36) was used to examine the coding potential of all VZV RNAs defined in this study (Table S2). Note that RNAs were excluded from CPC 2.0 analysis and defined as putatively noncoding if no proteins greater than 50 amino acids in length were encoded.

**Data visualization.** Figures associated with this study were generated using the R packages `Gviz` (74) and `GenomicRanges` (75). Heatmaps were generated by uploading TPM counts (Table S3) to `ClustVis` (76) with grouping of transcripts performed using the average correlation function.

**Statistical analysis.** Statistical analyses were performed where appropriate using `GraphPad Prism 5` software with specific tests indicated in the figure legends.

**Data availability.** All sequencing data sets associated with this study are available via the European Nucleotide Archive under the accession number [PRJEB38829](https://www.ebi.ac.uk/ena/record/PRJEB38829). Analyzed data sets generated as part of this study, including a database of transcripts, BED12 alignment files, and GFF3 files describing our VZV annotation, are freely available at <https://github.com/DepledgeLab/vzv-2.0>.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, TIF file, 1 MB.

**FIG S2**, TIF file, 1.2 MB.

**FIG S3**, TIF file, 0.6 MB.

**FIG S4**, TIF file, 1.5 MB.

**TABLE S1**, XLSX file, 0.05 MB.

**TABLE S2**, XLSX file, 0.01 MB.

**TABLE S3**, XLSX file, 0.04 MB.

**TABLE S4**, XLSX file, 0.01 MB.

**TABLE S5**, XLSX file, 0.02 MB.

## ACKNOWLEDGMENTS

We extend special thanks to Ian Mohr (New York University School of Medicine) for support of D.P.D. in part through National Institutes of Health (NIH) grants R01-AI073898 and R01-GM056927. S.E.B. was in part supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) project 022.005.032. This work was supported in part by the Takeda Science Foundation, Daiichi Sankyo Foundation of Life Science, Japan Society for the Promotion of Science (JSPS KAKENHI JP17K008858, JP16H06429, and JP16K21723) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT KAKENHI JP17H05816) (T.S.). J.B. acknowledges support from the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

The computational requirements for this work were supported in part by the NYU Langone High Performance Computing (HPC) Core's resources and personnel.

We declare no competing interests, financial or otherwise.

## REFERENCES

- Gershon AA, Breuer J, Cohen JI, Cohrs RJ, Gershon MD, Gildea D, Grose C, Hambleton S, Kennedy PGE, Oxman MN, Seward JF, Yamanishi K. 2015. Varicella zoster virus infection. *Nat Rev Dis Primers* 1:15016. <https://doi.org/10.1038/nrdp.2015.16>.
- Gilden DH, Vafai A, Shtram Y, Becker Y, Devlin M, Wellish M. 1983. Varicella-zoster virus DNA in human sensory ganglia. *Nature* 306:478–480. <https://doi.org/10.1038/306478a0>.
- Depledge DP, Sadaoka T, Ouwendijk WJD. 2018. Molecular aspects of varicella-zoster virus latency. *Viruses* 10:349. <https://doi.org/10.3390/v10070349>.
- Gilden DH, Kleinschmidt-DeMasters BK, LaGuardia JJ, Mahalingam R, Cohrs RJ. 2000. Neurologic complications of the reactivation of varicella-zoster virus. *N Engl J Med* 342:635–645. <https://doi.org/10.1056/NEJM200003023420906>.

5. Johnson RW, Rice ASC. 2014. Postherpetic neuralgia. *N Engl J Med* 371:1526–1533. <https://doi.org/10.1056/NEJMcp1403062>.
6. Gater A, Uhart M, McCool R, Pr eaud E. 2015. The humanistic, economic and societal burden of herpes zoster in Europe: a critical review. *BMC Public Health* 15:193. <https://doi.org/10.1186/s12889-015-1514-y>.
7. Davison AJ, Scott JE. 1986. The complete DNA sequence of varicella-zoster virus. *J Gen Virol* 67:1759–1816. <https://doi.org/10.1099/0022-1317-67-9-1759>.
8. Cohen JL. 2010. The varicella-zoster virus genome. *Curr Top Microbiol Immunol* 342:1–14. [https://doi.org/10.1007/82\\_2010\\_10](https://doi.org/10.1007/82_2010_10).
9. Depledge DP, Ouwendijk WJD, Sadaoka T, Braspenning SE, Mori Y, Cohrs RJ, Verjans GMGM, Breuer J. 2018. A spliced latency-associated VZV transcript maps antisense to the viral transactivator gene. *Nat Commun* 9:1167. <https://doi.org/10.1038/s41467-018-03569-2>.
10. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, Placantonakis DG, Mohr I, Wilson AC. 2019. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat Commun* 10:754. <https://doi.org/10.1038/s41467-019-08734-9>.
11. Stern-Ginossar N, Weisburd B, Michalski A, Le VTK, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Wilson AC. 2019. Decoding human cytomegalovirus. *Science* 338:1088–1093. <https://doi.org/10.1126/science.1227919>.
12. O'Grady T, Wang X, H oner z, Benstrup K, Baddoo M, Concha M, Flemington EK. 2016. Global transcript structure resolution of high gene density genomes through multi-platform data integration. *Nucleic Acids Res* 44:e145. <https://doi.org/10.1093/nar/gkw629>.
13. Arias C, Weisburd B, Stern-Ginossar N, Mercier A, Madrid AS, Bellare P, Holdorf M, Weissman JS, Ganem D. 2014. KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog* 10:e1003847. <https://doi.org/10.1371/journal.ppat.1003847>.
14. Finkel Y, Schmiedel D, Tai-Schmiedel J, Nachshon A, Winkler R, Dobesova M, Schwartz M, Mandelboim O, Stern-Ginossar N. 2020. Comprehensive annotations of human herpesvirus 6A and 6B genomes reveal novel and conserved genomic features. *Elife* 9:e50960. <https://doi.org/10.7554/eLife.50960>.
15. Bencun M, Klinke O, Hotz-Wagenblatt A, Klaus S, Tsai MH, Poirey R, Delecluse HJ. 2018. Translational profiling of B cells infected with the Epstein-Barr virus reveals 5 leader ribosome recruitment through upstream open reading frames. *Nucleic Acids Res* 46:2802–2819. <https://doi.org/10.1093/nar/gky129>.
16. O'Grady T, Teswick A, Hoffman BA, Wang Y, Medina EM, Kara M, van Dyk LF, Flemington EK, Tibbetts SA. 2019. Genome-wide transcript structure resolution reveals abundant alternate isoform usage from murine gammaherpesvirus 68. *Cell Rep* 27:3988–4002.e5. <https://doi.org/10.1016/j.celrep.2019.05.086>.
17. Whisnant AW, J urges CS, Hennig T, Wyler E, Prusty B, Rutkowski AJ, L'hernault A, Djakovic L, G obel M, D oring K, Menegatti J, Antrobus R, Matheson NJ, K unzig FWH, Mastrobuoni G, Bielow C, Kempa S, Liang C, Dandekar T, Zimmer R, Landthaler M, Gr asser F, Lehner PJ, Friedel CC, Erhard F, D olken L. 2020. Integrative functional genomics decodes herpes simplex virus 1. *Nature Communications* 11:2038. <https://doi.org/10.1038/s41467-020-15992-5>.
18. Prazs ak I, Moldov an N, Bal azs Z, Tomb acz D, Megyeri K, Sz ucs A, Csabai Z, Boldogk oi Z. 2018. Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics* 19:873. <https://doi.org/10.1186/s12864-018-5267-8>.
19. Reichelt M, Brady J, Arvin AM. 2009. 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* 83:3904–3918. <https://doi.org/10.1128/JVI.02137-08>.
20. Lenac Rovis T, Bailer SM, Pothineni VR, Ouwendijk WJD, Simic H, Babic M, Miklic K, Malic S, Verweij MC, Baiker A, Gonzalez O, von Brunn A, Zimmer R, Fruh K, Verjans GMGM, Jonjic S, Haas J. 2013. Comprehensive analysis of varicella-zoster virus proteins using a new monoclonal antibody collection. *J Virol* 87:6943–6954. <https://doi.org/10.1128/JVI.00407-13>.
21. Honess RW, Roizman B. 1974. Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *J Virol* 14:8–19. <https://doi.org/10.1128/JVI.14.1.8-19.1974>.
22. Kost RG, Kupinsky H, Straus SE. 1995. Varicella-zoster virus gene 63: transcript mapping and regulatory activity. *Virology* 209:218–224. <https://doi.org/10.1006/viro.1995.1246>.
23. Moriuchi H, Moriuchi M, Straus SE, Cohen JL. 1993. Varicella-zoster virus (VZV) open reading frame 61 protein transactivates VZV gene promoters and enhances the infectivity of VZV DNA. *J Virol* 67:4290–4295. <https://doi.org/10.1128/JVI.67.7.4290-4295.1993>.
24. Perera LP, Mosca JD, Ruyechan WT, Hayward GS, Straus SE, Hay J. 1993. A major transactivator of varicella-zoster virus, the immediate-early protein IE62, contains a potent N-terminal activation domain. *J Virol* 67:4474–4483. <https://doi.org/10.1128/JVI.67.8.4474-4483.1993>.
25. Moriuchi M, Moriuchi H, Straus SE, Cohen JL. 1994. Varicella-zoster virus (VZV) virion-associated transactivator open reading frame 62 protein enhances the infectivity of VZV DNA. *Virology* 200:297–300. <https://doi.org/10.1006/viro.1994.1190>.
26. Defechereux P, Melen L, Baudoux L, Merville-Louis MP, Rentier B, Piette J. 1993. Characterization of the regulatory functions of varicella-zoster virus open reading frame 4 gene product. *J Virol* 67:4379–4385. <https://doi.org/10.1128/JVI.67.7.4379-4385.1993>.
27. Ouwendijk WJD, Dekker L, Van den Ham H-J, Lenac Rovi s T, Haefner ES, Jonjic S, Haas J, Luidier TM, Verjans GMGM. 2020. Analysis of virus and host proteomes during productive HSV-1 and VZV infection in human epithelial cells. *Front Microbiol* 11:1179. <https://doi.org/10.3389/fmicb.2020.01179>.
28. Depledge DP, Mohr I, Wilson AC. 2018. Going the distance: optimizing RNA-Seq strategies for transcriptomic analysis of complex viral genomes. *J Virol* 93:e01342-18. <https://doi.org/10.1128/JVI.01342-18>.
29. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, Bruce M, Pantic N, Admassu T, James P, Warland A, Jordan M, Ciccone J, Serra S, Keenan J, Martin S, McNeill L, Wallace EJ, Jayasinghe L, Wright C, Blasco J, Young S, Brocklebank D, Juul S, Clarke J, Heron AJ, Turner DJ. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 15:201–206. <https://doi.org/10.1038/nmeth.4577>.
30. Tang AD, Soulette CM, van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, Brooks AN. 2020. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat Commun* 11:1438. <https://doi.org/10.1038/s41467-020-15171-6>.
31. Murata M, Nishiyori-Sueki H, Kojima-Ishiyama M, Carninci P, Hayashizaki Y, Itoh M. 2014. Detecting expressed genes using CAGE. *Methods Mol Biol* 1164:67–85. [https://doi.org/10.1007/978-1-4939-0805-9\\_7](https://doi.org/10.1007/978-1-4939-0805-9_7).
32. Thorvaldsd ottir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192. <https://doi.org/10.1093/bib/bbs017>.
33. Kemble GW, Annunziato P, Lungu O, Winter RE, Cha T-A, Silverstein SJ, Spaete RR. 2000. Open reading frame S/L of varicella-zoster virus encodes a cytoplasmic protein expressed in infected cells. *J Virol* 74:11311–11321. <https://doi.org/10.1128/jvi.74.23.11311-11321.2000>.
34. Sadaoka T, Yanagi T, Yamanishi K, Mori Y. 2010. Characterization of the Varicella-zoster virus ORF50 gene, which encodes glycoprotein M. *J Virol* 84:3488–3502. <https://doi.org/10.1128/JVI.01838-09>.
35. Jensen NJ, Depledge DP, Ng TFF, Leung J, Quinlivan M, Radford KW, Folster J, Tseng H-F, LaRussa P, Jacobsen SJ, Breuer J, Schmid DS. 2020. Analysis of the reiteration regions (R1 to R5) of varicella-zoster virus. *Virology* 546:38–50. <https://doi.org/10.1016/j.virol.2020.03.008>.
36. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, Gao G. 2017. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res* 45:W12–W16. <https://doi.org/10.1093/nar/gkx428>.
37. Sadaoka T, Yoshii H, Imazawa T, Yamanishi K, Mori Y. 2007. Deletion in open reading frame 49 of varicella-zoster virus reduces virus growth in human malignant melanoma cells but not in human embryonic fibroblasts. *J Virol* 81:12654–12665. <https://doi.org/10.1128/JVI.01183-07>.
38. Cox E, Reddy S, Iofin I, Cohen JL. 1998. Varicella-zoster virus ORF57, unlike its pseudorabies virus UL3.5 homolog, is dispensable for viral replication in cell culture. *Virology* 250:205–209. <https://doi.org/10.1006/viro.1998.9349>.
39. Cohrs RJ, Hurlley MP, Gildeen DH. 2003. Array analysis of viral gene transcription during lytic infection of cells in tissue culture with varicella-zoster virus. *J Virol* 77:11718–11732. <https://doi.org/10.1128/jvi.77.21.11718-11732.2003>.
40. Norberg P, Depledge DP, Kundu S, Atkinson C, Brown J, Haque T, Hussaini Y, MacMahon E, Molyneaux P, Papaevangelou V, Sengupta N, Koay ESC, Tang JW, Underhill GS, Grahm A, Studahl M, Breuer J, Bergstr om T. 2015. Recombination of globally circulating varicella-zoster virus. *J Virol* 89:7133–7146. <https://doi.org/10.1128/JVI.00437-15>.
41. Sadaoka T, Schwartz CL, Rajbhandari L, Venkatesan A, Cohen JL. 2017. Human embryonic stem cell-derived neurons are highly permissive for

- varicella-zoster virus lytic infection. *J Virol* 92:e01108-17. <https://doi.org/10.1128/JVI.01108-17>.
42. Sadaoka T, Depledge DP, Rajbhandari L, Venkatesan A, Breuer J, Cohen JI. 2016. In vitro system using human neurons demonstrates that varicella-zoster vaccine virus is impaired for reactivation, but not latency. *Proc Natl Acad Sci U S A* 113:E2403–E2412. <https://doi.org/10.1073/pnas.1522575113>.
  43. Baird NL, Bowlin JL, Cohrs RJ, Gildea D, Jones KL. 2014. Comparison of varicella-zoster virus RNA sequences in human neurons and fibroblasts. *J Virol* 88:5877–5880. <https://doi.org/10.1128/JVI.00476-14>.
  44. Zhang Z, Selariu A, Warden C, Huang G, Huang Y, Zacccheus O, Cheng T, Xia N, Zhu H. 2010. Genome-wide mutagenesis reveals that ORF7 is a novel VZV skin-tropic factor. *PLoS Pathog* 6:e1000971. <https://doi.org/10.1371/journal.ppat.1000971>.
  45. Toropova K, Huffman JB, Homa FL, Conway JF. 2011. The herpes simplex virus 1 UL17 protein is the second constituent of the capsid vertex-specific component required for DNA packaging and retention. *J Virol* 85:7513–7522. <https://doi.org/10.1128/JVI.00837-11>.
  46. Yang M, Hay J, Ruyechan WT. 2004. The DNA element controlling expression of the varicella-zoster virus open reading frame 28 and 29 genes consists of two divergent unidirectional promoters which have a common USF site. *J Virol* 78:10939–10952. <https://doi.org/10.1128/JVI.78.20.10939-10952.2004>.
  47. Storlie J, Jackson W, Hutchinson J, Grose C. 2006. Delayed biosynthesis of varicella-zoster virus glycoprotein C: upregulation by hexamethylene bisacetamide and retinoic acid treatment of infected cells. *J Virol* 80:9544–9556. <https://doi.org/10.1128/JVI.00668-06>.
  48. Koshizuka T, Ota M, Yamanishi K, Mori Y. 2010. Characterization of varicella-zoster virus-encoded ORF0 gene—comparison of parental and vaccine strains. *Virology* 405:280–288. <https://doi.org/10.1016/j.virol.2010.06.016>.
  49. Felser JM, Kinchington PR, Inchauspe G, Straus SE, Ostrove JM. 1988. Cell lines containing varicella-zoster virus open reading frame 62 and expressing the “IE” 175 protein complement ICP4 mutants of herpes simplex virus type 1. *J Virol* 62:2076–2082. <https://doi.org/10.1128/JVI.62.6.2076-2082.1988>.
  50. Leppik K, Das R, Barna M. 2018. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* 19:158–174. <https://doi.org/10.1038/nrm.2017.103>.
  51. Tyler SD, Peters GA, Grose C, Severini A, Gray MJ, Upton C, Tipples GA. 2007. Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain variability with implications for attenuation and phenotypic differences. *Virology* 359:447–458. <https://doi.org/10.1016/j.virol.2006.09.037>.
  52. Jones M, Dry IR, Frampton D, Singh M, Kanda RK, Yee MB, Kellam P, Hollinshead M, Kinchington PR, O’Toole EA, Breuer J. 2014. RNA-seq analysis of host and viral gene expression highlights interaction between varicella zoster virus and keratinocyte differentiation. *PLoS Pathog* 10:e1003896. <https://doi.org/10.1371/journal.ppat.1003896>.
  53. Hancock MH, Skalsky RL. 2018. Roles of non-coding RNAs during herpesvirus infection. *Curr Top Microbiol Immunol* 419:243–280. [https://doi.org/10.1007/82\\_2017\\_31](https://doi.org/10.1007/82_2017_31).
  54. Saxena A, Carninci P. 2011. Long non-coding RNA modifies chromatin: epigenetic silencing by long non-coding RNAs. *Bioessays* 33:830–839. <https://doi.org/10.1002/bies.201100084>.
  55. Pelechano V, Steinmetz LM. 2013. Gene regulation by antisense transcription. *Nat Rev Genet* 14:880–893. <https://doi.org/10.1038/nrg3594>.
  56. Zhang C, Hastings ML, Krainer AR, Zhang MQ. 2007. Dual-specificity splice sites function alternatively as 5′ and 3′ splice sites. *Proc Natl Acad Sci U S A* 104:15028–15033. <https://doi.org/10.1073/pnas.0703773104>.
  57. Michael EJ, Kuck KM, Kinchington PR. 1998. Anatomy of the varicella-zoster virus open-reading frame 4 promoter. *J Infect Dis* 178:S27–S33. <https://doi.org/10.1086/514273>.
  58. Wang L, Sommer M, Rajamani J, Arvin AM. 2009. Regulation of the ORF61 promoter and ORF61 functions in varicella-zoster virus replication and pathogenesis. *J Virol* 83:7560–7572. <https://doi.org/10.1128/JVI.00118-09>.
  59. Kinchington PR, Hougland JK, Arvin AM, Ruyechan WT, Hay J. 1992. The varicella-zoster virus immediate-early protein IE62 is a major component of virus particles. *J Virol* 66:359–366. <https://doi.org/10.1128/JVI.66.1.359-366.1992>.
  60. Perera LP, Mosca JD, Sadeghi-Zadeh M, Ruyechan WT, Hay J. 1992. The varicella-zoster virus immediate early protein, IE62, can positively regulate its cognate promoter. *Virology* 191:346–354. [https://doi.org/10.1016/0042-6822\(92\)90197-w](https://doi.org/10.1016/0042-6822(92)90197-w).
  61. Yang M, Peng H, Hay J, Ruyechan WT. 2006. Promoter activation by the varicella-zoster virus major transactivator IE62 and the cellular transcription factor USF. *J Virol* 80:7339–7353. <https://doi.org/10.1128/JVI.00309-06>.
  62. Ruyechan WT, Peng H, Yang M, Hay J. 2003. Cellular factors and IE62 activation of VZV promoters. *J Med Virol* 70(Suppl 1):S90–S94. <https://doi.org/10.1002/jmv.10328>.
  63. Sadaoka T, Rajbhandari L, Shukla P, Jagdish B, Lee H, Lee G, Venkatesan A. 2020. Human stem cell derived sensory neurons are positioned to support varicella zoster virus latency. *bioRxiv*. <https://doi.org/10.1101/2020.01.24.919290>.
  64. Ouwendijk WJD, Geluk A, Smits SL, Getu S, Osterhaus ADME, Verjans GMGM. 2014. Functional characterization of ocular-derived human alpha herpesvirus cross-reactive CD4 T cells. *J Immunol* 192:3730–3739. <https://doi.org/10.4049/jimmunol.1302307>.
  65. Ouwendijk WJD, Mahalingam R, de Swart RL, Haagsmans BL, van Amerongen G, Getu S, Gildea D, Osterhaus ADME, Verjans GMGM. 2013. T-cell tropism of simian varicella virus during primary infection. *PLoS Pathog* 9:e1003368. <https://doi.org/10.1371/journal.ppat.1003368>.
  66. Maertzdorf J, Remeijer L, Van Der Lelij A, Buitenwerf J, Niesters HGM, Osterhaus ADME, Verjans GMGM. 1999. Amplification of reiterated sequences of herpes simplex virus type 1 (HSV-1) genome to discriminate between clinical HSV-1 isolates. *J Clin Microbiol* 37:3518–3523. <https://doi.org/10.1128/JCM.37.11.3518-3523.1999>.
  67. Ma HT, Poon RYC. 2017. Synchronization of HeLa cells. *Methods Mol Biol* 1524:189–201. [https://doi.org/10.1007/978-1-4939-6603-5\\_12](https://doi.org/10.1007/978-1-4939-6603-5_12).
  68. Ouwendijk WJD, Choe A, Nagel MA, Gildea D, Osterhaus ADME, Cohrs RJ, Verjans GMGM. 2012. Restricted varicella-zoster virus transcription in human trigeminal ganglia obtained soon after death. *J Virol* 86:10203–10206. <https://doi.org/10.1128/JVI.01331-12>.
  69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
  70. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
  71. Bonfert T, Kirner E, Csaba G, Zimmer R, Friedel CC. 2015. ContextMap 2: fast and accurate context-based RNA-seq mapping. *BMC Bioinformatics* 16:122. <https://doi.org/10.1186/s12859-015-0557-5>.
  72. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
  73. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
  74. Hahne F, Ivanek R. 2016. Visualizing genomic data using Gviz and bioconductor. *Methods Mol Biol* 1418:335–351. [https://doi.org/10.1007/978-1-4939-3578-9\\_16](https://doi.org/10.1007/978-1-4939-3578-9_16).
  75. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol* 9:e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>.
  76. Metsalu T, Vilo J. 2015. ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res* 43:W566–W570. <https://doi.org/10.1093/nar/gkv468>.
  77. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>.
  78. Ouwendijk WJD, Depledge DP, Rajbhandari L, Lenac Rovis T, Jonjic S, Breuer J, Venkatesan A, Verjans GMGM, Sadaoka T. 2020. Expression of varicella-zoster virus VLT-ORF63 fusion transcript induces broad viral gene expression during reactivation from neuronal latency. *bioRxiv*. <https://doi.org/10.1101/2020.09.11.294280>.