

# Reanalysis of single-cell RNA sequencing data does not support herpes simplex virus 1 latency in non-neuronal ganglionic cells in mice

Werner J. D. Ouwendijk,<sup>1</sup> Pavitra Roychoudhury,<sup>2,3</sup> Anthony L. Cunningham,<sup>4,5</sup> Keith R. Jerome,<sup>2,3</sup> David M. Koelle,<sup>2,3,6,7,8</sup> Paul R. Kinchington,<sup>9</sup> Ian Mohr,<sup>10</sup> Angus C. Wilson,<sup>10</sup> Georges G. M. G. M. Verjans,<sup>1</sup> Daniel P. Depledge<sup>10,11,12,13</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 11.

**ABSTRACT** Most individuals are latently infected with herpes simplex virus type 1 (HSV-1), and it is well-established that HSV-1 establishes latency in sensory neurons of peripheral ganglia. However, it was recently proposed that latent HSV-1 is also present in immune cells recovered from the ganglia of experimentally infected mice. Here, we reanalyzed the single-cell RNA sequencing (scRNA-Seq) data that formed the basis for that conclusion. Unexpectedly, off-target priming in 3' scRNA-Seq experiments enabled the detection of non-polyadenylated HSV-1 *latency-associated transcript (LAT)* intronic RNAs. However, *LAT* reads were near-exclusively detected in mixed populations of cells undergoing cell death. Specific loss of HSV-1 *LAT* and neuronal transcripts during quality control filtering indicated widespread destruction of neurons, supporting the presence of contaminating cell-free RNA in other cells following tissue processing. In conclusion, the reported detection of latent HSV-1 in non-neuronal cells is best explained using compromised scRNA-Seq datasets.

**IMPORTANCE** Most people are infected with herpes simplex virus type 1 (HSV-1) during their life. Once infected, the virus generally remains in a latent (silent) state, hiding within the neurons of peripheral ganglia. Periodic reactivation (reawakening) of the virus may cause fresh diseases such as cold sores. A recent study using single-cell RNA sequencing (scRNA-Seq) proposed that HSV-1 can also establish latency in the immune cells of mice, challenging existing dogma. We reanalyzed the data from that study and identified several flaws in the methodologies and analyses performed that invalidate the published conclusions. Specifically, we showed that the methodologies used resulted in widespread destruction of neurons which resulted in the presence of contaminants that confound the data analysis. We thus conclude that there remains little to no evidence for HSV-1 latency in immune cells.

**KEYWORDS** herpes simplex virus, latency, scRNA-Seq, immune cells, neurons, Latency-Associated Transcripts (LATs)

All herpesviruses share the ability to establish a lifelong latent infection in their hosts, which later facilitates virus reactivation and intermittent spread to naïve hosts. Herpesviruses belonging to the *Alphaherpesvirinae* subfamily, which includes the ubiquitous human pathogen herpes simplex virus type 1 (HSV-1), establish latency in neurons of the peripheral nervous system (1). The primary sites of HSV-1 latency are sensory neurons of the trigeminal ganglia (TG) and/or dorsal root ganglia (DRG) (2). Moreover, HSV-1 DNA also persists in neurons of other sensory and autonomic ganglia (3, 4), and the possible existence of a latent HSV-1 reservoir in the cornea has been a long-standing debate in the field (5, 6). Therefore, the development of single-cell RNA

**Editor** Felicia Goodrum, The University of Arizona, Tucson, Arizona, USA

Address correspondence to Georges G. M. G. M. Verjans, g.verjans@erasmusmc.nl, or Daniel P. Depledge, depledge.daniel@mh-hannover.de.

Werner J. D. Ouwendijk and Pavitra Roychoudhury contributed equally to this article. These authors are listed alphabetically by last name.

Georges G. M. G. M. Verjans and Daniel P. Depledge contributed equally to this article.

The authors declare no conflict of interest.

See the funding table on p. 11.

**Received** 28 November 2023

**Accepted** 14 February 2024

**Published** 6 March 2024

Copyright © 2024 Ouwendijk 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/).

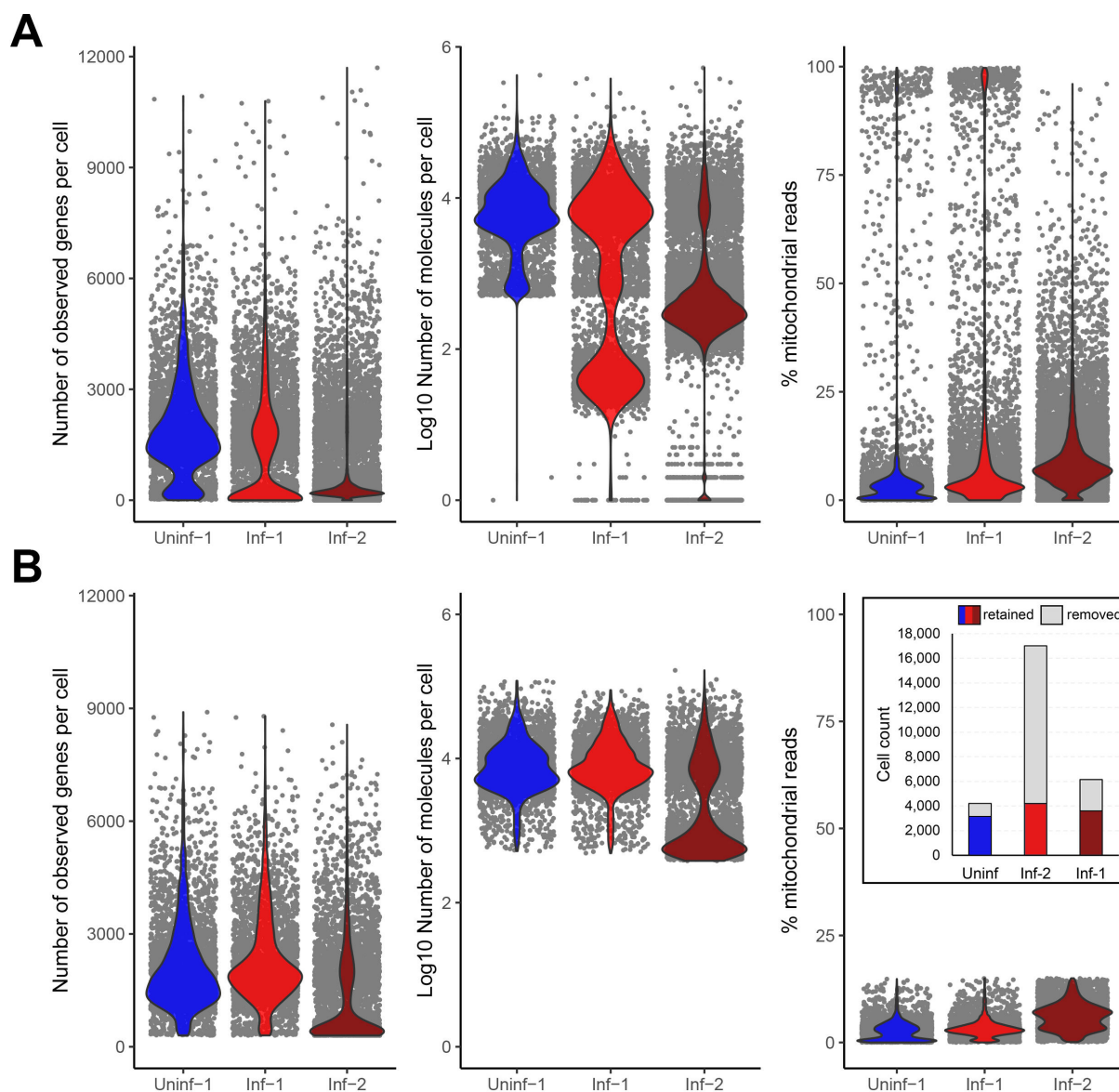
sequencing (scRNA-Seq) technologies provides a unique opportunity to study latency and reactivation in neuronal latency models (7), as well as to provide evidence for whether HSV latency is established in non-neuronal cells. A recent study by Wang et al. (8) addressed the latter and concluded that, in addition to neurons, HSV-1 establishes latency in immune cells that are present in the TG of HSV-1 experimentally infected mice. Here, we present a reanalysis of the scRNA-Seq data used by Wang et al. (8) and demonstrate significant problems with the quality of their scRNA-Seq datasets and conclude that these data cannot be used to support their conclusion.

During latency, HSV-1 gene expression is highly restricted and limited to the *latency-associated transcript (LAT)* and associated miRNAs (9–14). The primary *LAT* transcript is 8.3 kb in size, capped, and polyadenylated (15–18). Splicing produces stable 1.5 and 2 kb *LAT* intron lariats that accumulate to high levels in sensory neurons, while the highly unstable 6.3 kb spliced polyadenylated RNA is rapidly processed into viral microRNAs (19, 20) (for a comprehensive review of *LAT*, see (21)). Two major claims are reported by Wang et al.: (i) in addition to neurons, various types of immune cells recovered from TG of experimentally infected mice express HSV-1 *LAT*, and (ii) the presence of *LAT* in these cells indicates that HSV-1 can establish latency in non-neuronal cells present in the TG. The core data supporting these claims was obtained by drop-let-based scRNA-Seq analysis (10X Genomics platform) of TG from uninfected C57BL/6 mice (dataset: “Uninf-1”) and two biological replicate groups of C57BL/6 mice infected via the corneal route with  $2 \times 10^5$  plaque-forming units/eye of HSV-1 strain McKrae 35 days earlier (datasets: “Inf-1” and “Inf-2”). Each biological replicate was obtained by pooling the CD45<sup>POS</sup> leukocyte-enriched dissociated cells from paired left and right TG from 15 animals (i.e., 30 TG per biological replicate). To examine the claims by Wang et al. (8) in more detail, we aimed to reproduce their analyses. However, none of the (raw) scRNA-Seq datasets, barcode matrices, or analyses scripts are linked to the published article. We were later able to obtain the raw data from the study (SRA PRJNA937697, GEO GSE225839) via the handling editor of Science Advances. What follows is a reanalysis of the data presented by Wang et al. (8) using the same filtered barcode matrices that served as input for their scRNA-Seq analysis.

## RESULTS

### Quality control of scRNA-Seq datasets

Isolation of dissociated, single cells from organs requires mechanical and/or enzymatic tissue dissociation, typically followed by removal of dead cells and (if needed) further purification of cells of interest by magnetic bead- or flow cytometry-based cell sorting. Quality control (QC) filtering of the obtained scRNA-Seq datasets is therefore a critical first step (22). The Chromium Single Cell 3' v3.1 Reagent Kit (10X Genomics) used by Wang et al. (8) for library preparation is designed to capture polyadenylated RNAs and prime reverse transcription using a poly(T) primer that also includes the barcode and unique molecular index (UMI) sequences. QC filtering involves the identification and removal of doublets, as well as an assessment of cell viability in each of the samples. This latter is achieved by measuring, for each individual cell (i) the number of unique genes detected, (ii) the total number of RNA molecules (UMI) recovered and (iii) the proportion of reads derived from mitochondrial RNAs (23, 24) (Fig. 1). In the original matrix count files generated by Wang et al. (8), the dataset designated Uninf-1 had 1,822 distinct genes detected per cell and a median total RNA count of 6,879, while 83% cells had mitochondrial RNA fractions <15%. By contrast, the Inf-1 and Inf-2 datasets showed different results with a median of 1,979 and 558 distinct genes per cell, a median total RNA count of 7,148 and 929, and only 51% and 33% of cells having a mitochondrial RNA fraction <15% (Fig. 1A), respectively. These data indicate higher proportions of dead and dying cells within both the Inf-1 and Inf-2 datasets with Inf-2 particularly severely impacted. Next, we applied filters on mitochondrial RNA content and unique gene counts, according to the parameters described by Wang et al. (8), namely that cells were only retained if between 300 and 9,000 distinct genes were detected, and the



**FIG 1** Quality control of scRNA-Seq datasets derived from pools of HSV-1 latently-infected mouse trigeminal ganglia. scRNA-Seq datasets generated by Wang et al. (8) were obtained from uninfected (“Uninf-1”) and two biological replicates of HSV-1-infected C57BL/6 mice (“Inf-1” and “Inf-2”). Each replicate was obtained by pooling dissociated cells – composed of a 1:1 mixture of CD45-enriched cells and the original cell suspension – from left and right TG from 15 animals (30 ganglia). (A) Using the filtered barcode matrices generated by Wang et al. (8), the quality of each dataset was assessed by (left) the number of unique genes detected per cell, (middle) the total number of RNA molecules (UMI) recovered per cell and (right) the proportion of reads per cell derived from mitochondrial RNAs. (B) Quality control filtering of these datasets dramatically reduced the total number of cells available for analysis, indicative that many dead/dying cells were present in the original single-cell suspensions of Inf-1 and Inf-2. Filtering parameters removed cells with less than 300 or more than 9,000 distinct expressed genes, and cells for which more than 15% of reads derived from mitochondrial RNAs. Inset: Number of cells before and after filtering.

proportion of mitochondrial reads present was below 15%. At this stage, we observed large numbers of low-quality cells filtered out of each data set (Fig. 1B). This resulted in 3,608 cells for Uninf-1 (reduced from 4,206, a loss of 14%), 3,158 cells for Inf-1 (reduced from 6,155, a loss of 49%), and 5,660 cells for Inf-2 (reduced from 17,014, a loss of 67%).

### Clustering and annotation of single cells

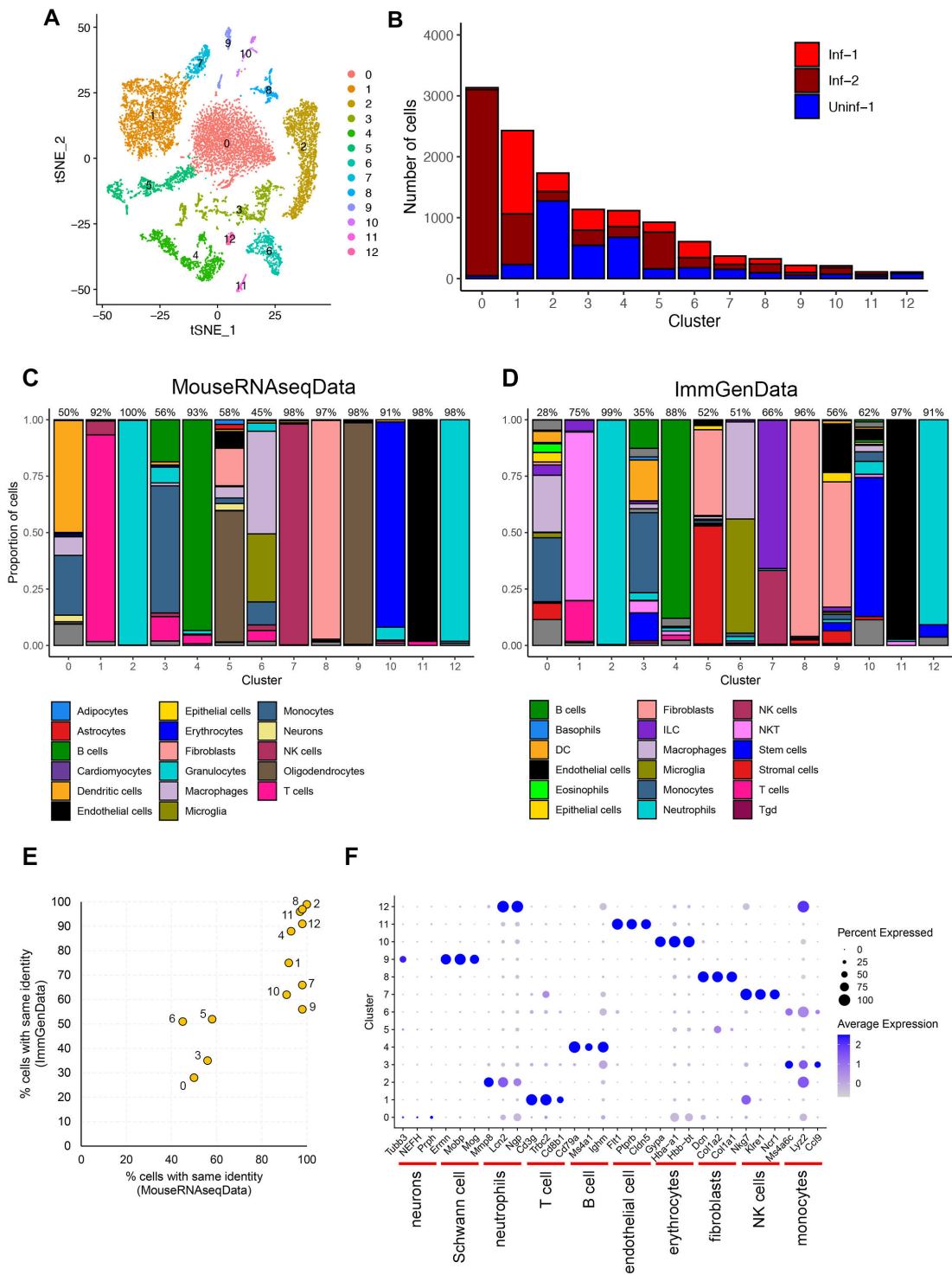
One of the most contentious components of any scRNA-Seq analyses is the reduction of multidimensional into simple two-dimensional figures through either t-distributed

stochastic neighbor embedding (t-SNE) or Uniform Manifold Approximation and Projection (UMAP) techniques. While this has been reviewed elsewhere (22), it is worth noting that significant care must be taken when trying to interpret these data. We performed integration and clustering in a similar manner to Wang et al. (8), matching as many parameters as possible (see Methods section) (Fig. 2A). A deeper analysis of the clusters revealed significant differences in the relative proportions of cell types present in each dataset with some clusters being almost entirely derived from a single dataset (e.g., Cluster 0 and to a lesser extent cluster 2, Fig. 2B).

A second challenging component of scRNA-Seq analyses is the process of assigning specific cellular identities to a given cluster. This is typically achieved by identifying distinct markers within a given cluster and comparing this to a well-described reference database of cell identities. Here, we used the same annotation tool and reference databases as Wang et al. (8) (SingleR (28), with (MouseRNAseqData (29) and ImmGen-Data (30) from the celldex package (<https://github.com/LTLA/celldex>) (Fig. 2C and D). This analysis (i) produced generally convergent results (Fig. 2E) and (ii) demonstrated that while most clusters could be identified with high confidence (i.e., more than 90% of cells present are predicted to have the same identity) other clusters are reported as mixtures of cell types (i.e., clusters 0, 3, 5, and 6) and assigning singular cell identities to these can only be done with low confidence (Fig. 2C and D). Subsequently, an analysis using representative markers for the cell types present in each cluster further confirmed the division between high confidence and low confidence cluster identities (Fig. 2F). We further analyzed Cluster 0 to better understand why no robust cellular identity could be assigned. Notably, this cluster derived almost entirely from the Inf-2 dataset (Fig. 2B), and when compared to other clusters, it was characterized by containing cells with high proportions of mitochondrial reads and low numbers of detectably expressed genes per cell (Fig. 3A and B). To test the hypothesis that most cells in this cluster were dying/dead, we summarized the expression of 40 cell death markers (31), and again when compared to other clusters, determined these to be predominantly expressed in the Inf-2 derived cells in Cluster 0 (Fig. 3C; Table S1). This is particularly relevant in the context of the original Wang et al. (8) analyses as the major conclusions in that study were derived from the analysis of cells in this cluster. Finally, we examined the possibility that ambiguous cell types and lack of clustering by cell type could be related to doublets present in the samples, or due to differences in integration methods between our analysis and Wang et al. While not performed by Wang et al., the identification and removal of doublets is considered an important quality control step during scRNA-Seq data processing. We used DoubletFinder (32) and identified just 847 doublets across all three datasets (representing 6.8% of total cells). Doublets were identified in multiple clusters, and importantly, very few were present in Cluster 0, representing only 1.2% of cells in that cluster (Fig. S1). We therefore concluded that doublets could not explain the issues with Cluster 0 and sample Inf-2. Secondly, while the integration of datasets is not strictly required and is mostly useful for multimodal datasets, we performed integration after doublet removal and found that, while this resulted in 12 clusters rather than 13, the overall cell type classifications did not change (Fig. S2). In conclusion, these data indicate that (i) Inf-1 and Inf-2 datasets are not valid biological replicates, (ii) clusters 0, 3, 5, and 6 cannot be assigned a specific cell identity with high confidence (Fig. 2), and (iii) Inf-2 derived cells in Cluster 0 are likely undergoing programmed cell death (Fig. 3).

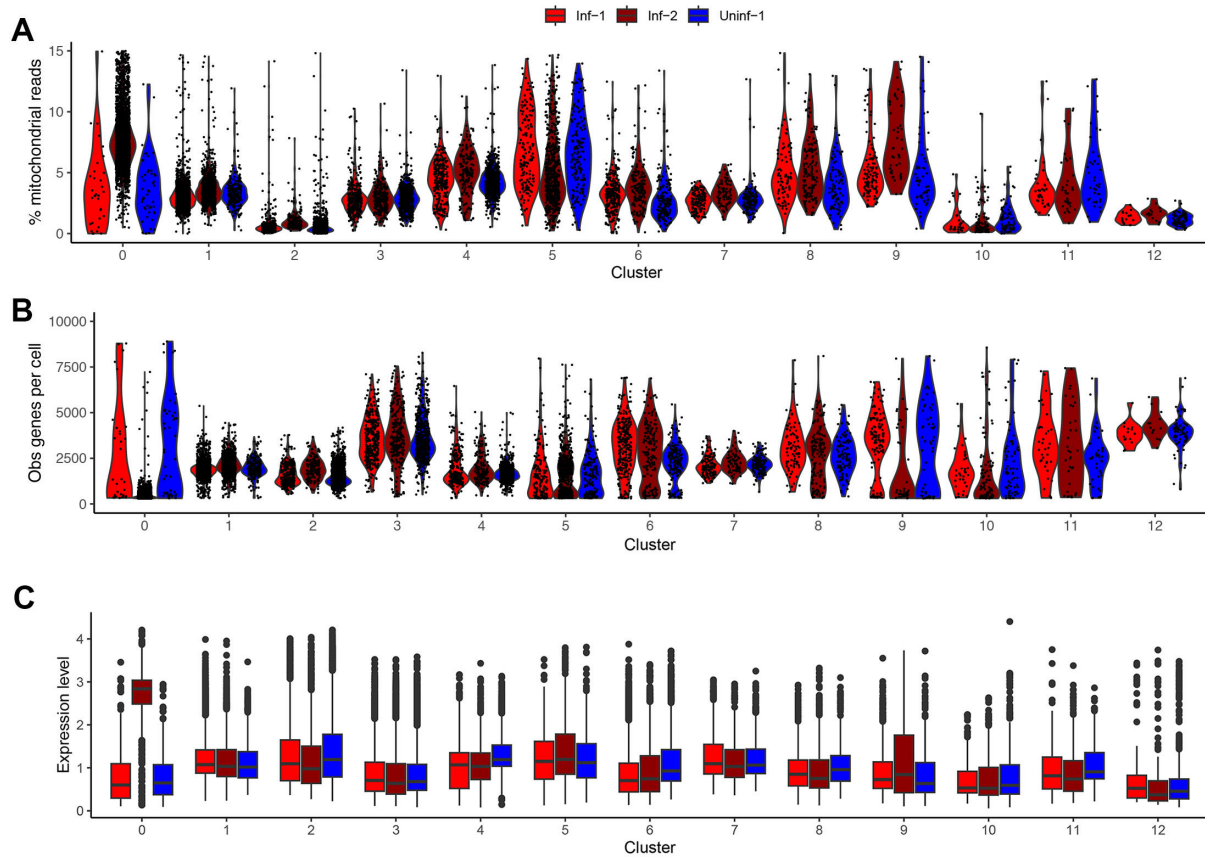
### Off-target capture enables profiling of *LAT* intron lariats by scRNA-Seq

We next switched focus to the reported detection of HSV-1 *LAT* transcripts in multiple clusters. Of particular note is that the stable HSV-1 *LAT* 1.5 and 2 kb introns are not polyadenylated, and that the mature *LAT* RNA is highly unstable (15, 18). Thus, one would not expect to detect intron-derived RNAs by 3' scRNA-Seq, in which the 3' oligo d(T) adapter is designed to prime poly(A) tails. However, reanalysis of the raw fastq files from Wang et al. (8) demonstrated that 74%–92% of viral reads (representing <0.005% of all reads) aligned to the *LAT* intron, while the remaining reads mapped at low density



**FIG 2** Clustering and annotation of cell populations. (A) Aggregated tSNE plot of all three datasets identifies 13 distinct clusters (0–12). (B) The proportion and total number of cells in each cluster shown differs between datasets e.g., cluster 0 is almost entirely composed of cells from the Inf-2 dataset. (C and D) SingleR was used to perform unbiased cell type recognition using both (C) MouseRNAseq and (D) ImmGen databases. For each cluster, the maximum proportion of cells given the same identity (analogous to a confidence score) is shown above the bar plot. (E) Both databases yielded similar results and confidence scores. Notably, clusters 0, 3, 5, and 6 could not be adequately resolved into a single dominant cell type. (F) Bubble plot showing both the proportion of cells in each cluster that express a particular cell type-specific marker (25–27) and the expression level of that marker.



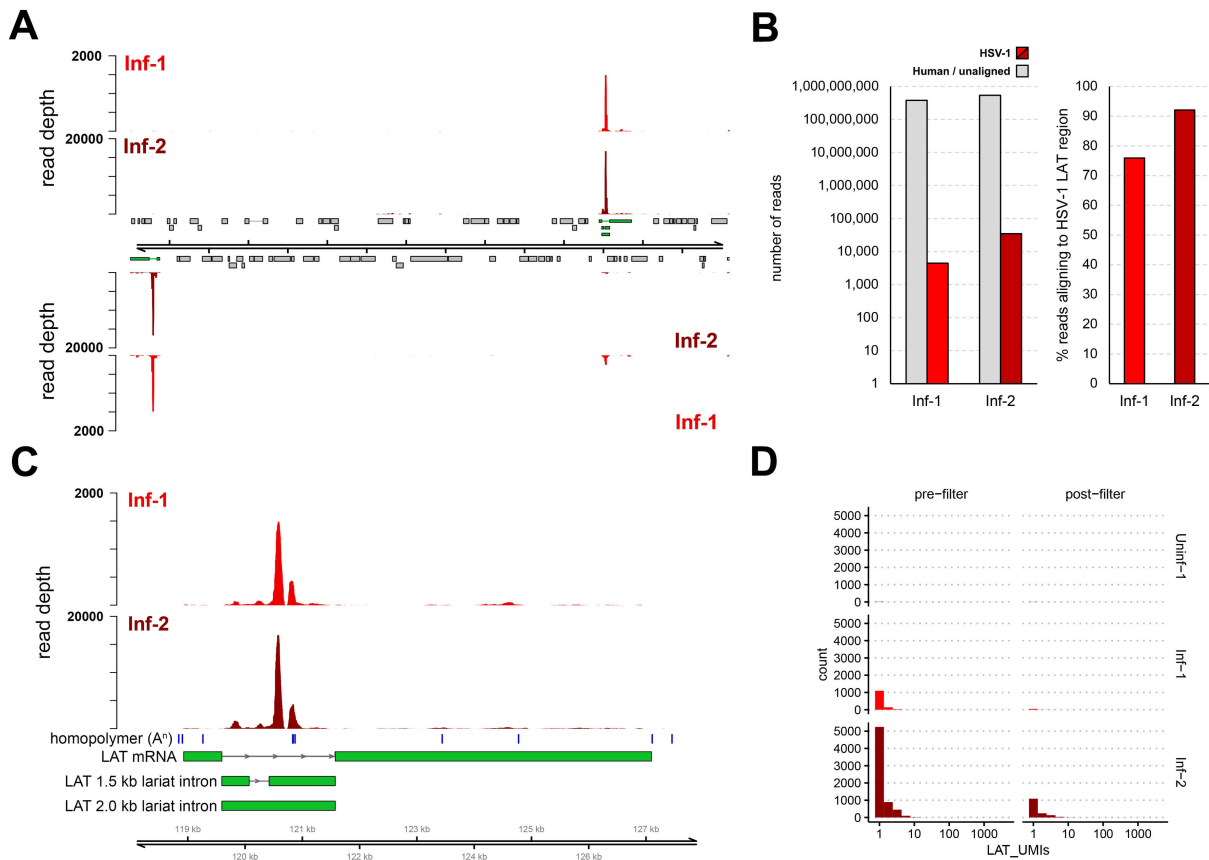


**FIG 3** Further evaluation of cell quality within each cluster. For each cluster, and segregated by dataset, we determined (A) the proportion of mitochondrial reads per cell, (B) the number of distinct genes expressed per cell, and (C) the aggregated expression level of a selection of cell death markers (Table S1).

throughout the rest of the HSV-1 genome (Fig. 4A and B; Table S2). Closer examination of read alignments across the *LAT* locus showed consistent alignments that were associated with short adenosine homopolymers located within the intron and a much smaller peak at the 3' end of the mature *LAT* (Fig. 4C). Taken together, these data show that off-target priming in 3' scRNA-Seq experiments (33) results in the detection of non-polyadenylated HSV-1 *LAT* introns. Similar results have been observed in other scRNA-Seq studies of HSV-1 latently infected ganglia, indicating that the 3' scRNA-Seq approach is compatible with studies of HSV-1 latency models (7, 34), however, the efficiency of this off-target priming remains unknown.

### Loss of HSV-1 *LAT* during filtering suggests cell-free RNA contamination

*LAT* reads were not universally detected in the Inf-1 and Inf-2 datasets, but instead were 10 times more abundant in the Inf-2 dataset (Fig. 4B). In addition, most *LAT* reads in the base dataset (filtered barcode matrices) were excluded during the initial quality control process (Fig. 4D). Notably, most cells with *LAT* reads in the filtered dataset only contained a single *LAT* read as determined by the UMI present in each read (Fig. 4D). Subsequent analysis of the individual clusters demonstrated that the vast majority of cells designated as *LAT*<sup>POS</sup> were associated with Cluster 0 and were almost exclusively from the Inf-2 dataset (Fig. 5A). Similarly, the relative expression of *LAT* was highest in Cluster 0 (Fig. 5B). Because (i) this cluster is composed of dead/dying cells and (ii) HSV-1 *LAT* introns accumulate to high levels in neurons (35), we hypothesized that high background levels of cell-free RNA – originating from HSV-1-infected neurons that were damaged during TG tissue processing – could be the source of *LAT* reads in non-neuronal cells. To test this hypothesis, we compared the number of reads aligning to HSV-1 *LAT* and several cell

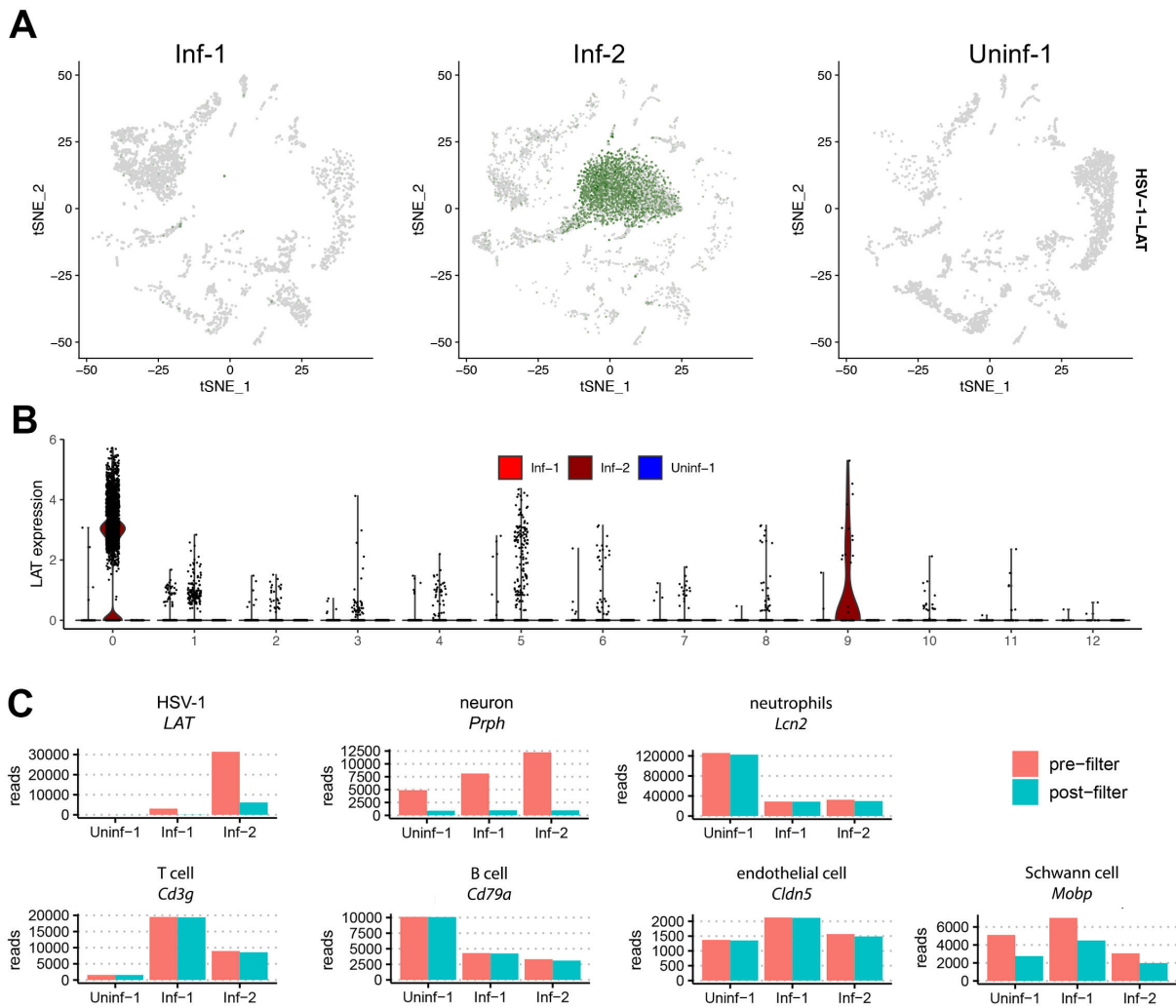


**FIG 4** Abundant detection of non-polyadenylated *LAT* introns. (A) Coverage plot denoting the distribution of HSV-1 reads in the raw (unfiltered) Inf-1 and Inf-2 datasets. Black lines represent the HSV-1 strain McKrae genome, while grey boxes indicate open reading frames and thin lines indicate introns. Top and bottom panels represent HSV-1 reads aligning to the forward and reverse strand of the genome. Both copies of the *LAT* locus are indicated in green. (B) Reads aligning to the HSV-1 genome comprised only a small proportion (< 0.006%) of the Inf-1 and Inf-2 datasets, while most of these reads (74%–92%) aligned to the *LAT* introns located in the *LAT* locus. (C) Coverage plot of the *LAT* locus confirms that the majority of HSV-1 *LAT* reads aligned next to short adenosine homopolymers (blue vertical bars) located within the intron, indicating off-target capture. (D) The majority of cells with *LAT* reads contained only a single copy of *LAT* (i.e., a single UMI) and most of these were removed during the QC filtering step.

type-specific markers in the pre- (Fig. 1A) and post-filtered (Fig. 1B) datasets. Strikingly, this analysis demonstrated a significant loss of both HSV-1 *LAT* and neuronal markers during filtering that was not observed for any of the other major cell types present (Fig. 5C; Fig. 53). Thus, extensive death of (HSV-1 latently-infected) neurons during TG processing is the most likely source of ambient RNA contamination (36, 37).

**DISCUSSION**

The recent study by Wang et al. (8) has challenged the long-held dogma that herpes simplex virus type 1 (HSV-1) exclusively establishes latency in neurons. However, overturning existing dogmas necessitates robust and rigorous evidence that is supported by well-controlled independent experiments using orthologous methodologies. A key premise of Wang et al. (8) is that the presence of HSV-1 *LAT* RNA in a cell is sufficient to conclude that the virus has established latency in the infected cell. HSV-1 latency can be operationally defined as the presence of viral DNA in host cells in the absence of virus particle production, provided that the genome is maintained in a reactivatable state that enables the production of new infectious HSV-1 particles (38, 39). Transcriptional activity of the latent HSV-1 genome is repressed with exclusive expression of *LAT* driven by a neuron-specific promoter (11, 13, 15). Neurons are the only cell type in which HSV-1 latency has been clinically and experimentally demonstrated in both human and



**FIG 5** Abundant loss of reads associated with HSV-1 *LAT* and neuronal markers during filtering. (A) tSNE plot from Fig. 2A split by sample shows that the majority of *LAT* expression (in green) mapped to Cluster 0, which was exclusively present in the Inf-2 dataset. (B) Violin plot showing log normalized expression of *LAT* in each cluster. (C) For HSV-1 *LAT* and representative markers of different cell types, we determined the total number of reads present in the pre-filtered (Fig. 1A) and post-filtered (Fig. 1B) datasets. Data for a wider selection of markers is shown in Fig. S3.

mouse ganglia (2, 9, 40–43). Interestingly, not all HSV-1-infected TG neurons express *LAT* in the HSV-1 mouse model (44), and it is unclear whether all neurons harboring HSV-1 DNA express *LAT* in human TG (45, 46). This raises the question of whether all neurons containing HSV-1 DNA support virus reactivation. Thus, even if low-abundant *LAT* reads were detected in non-neuronal cells, this is not by the currently accepted definitions (38, 39), conclusive evidence of HSV-1 latency.

Principally, our reanalysis of the single-cell RNA sequencing (scRNA-Seq) data from Wang et al. (8) reveals significant problems with the quality of their scRNA-Seq datasets. We conclude from our reanalyses that the reported detection of *LAT* reads from non-neuronal cells is best explained by cell-free RNA originating from latently-infected neurons that were damaged during tissue processing. Moreover, we here specifically establish that the Inf-1 and Inf-2 datasets cannot be considered biological replicates, with the Inf-2 dataset in particular showing evidence of extensive cell destruction during tissue processing. We have also shown that integration of these datasets yields multiple cell clusters that cannot be assigned a specific cellular identity. One of these clusters is dominated by low-quality HSV-1 *LAT*-expressing cells that are almost entirely derived from the Inf-2 dataset. A deep analysis of this cluster identified markers of multiple



cell types including neurons and expression of a large number of programmed cell death markers. The significant loss of reads associated with HSV-1 *LAT* and neuronal markers during QC filtering further supports extensive neuronal cell death during tissue processing and the release of both neuronal RNAs and HSV-1 *LAT* into the homogenized single-cell suspension. Such RNA is easily bound on the surface of other cell types and thus carried into the droplets in which cell lysis and the initial steps of scRNA-Seq library preparation take place (36, 37).

As we had to specifically request that the scRNAseq datasets be made available to us post-publication (and these data are not linked to the publication on the journal website), we feel obligated to reiterate the FAIR (Findability, Accessibility, Interoperability, and Reusability) data principles. These guidelines provide a framework to increase transparency and promote the reuse of data by the scientific community (47, 48), which in turn will accelerate scientific discoveries and provide the opportunity for dataset correction by other analyses. Many funding agencies, universities, and scientific journals aim to promote open science by recommending or requiring researchers to adhere to the FAIR principles and open-access publishing. In the context of scRNA-Seq experiments, this means that all raw sequence files, metadata, raw and filtered matrices, including all code/scripts used for analysis, need to be deposited in publicly available data repositories (e.g., GEO, SRA, and GitHub). References to the location of the data should be provided in the relevant sections of the article. It would have been helpful for Wang et al. (8) to include references to their scRNA datasets in the manuscript to follow journal guidelines, and to provide sufficient details in the Methods section to reproduce all aspects of their data analysis, e.g., the Cell Ranger parameters used for aligning raw sequence data and the construction of the hybrid genome reference were not described. While Methods sections are often written in a concise manner, it is increasingly common, and generally required by journals, that authors make available all scripts used for the analysis of the original data presented. Additionally, we recommend that authors demonstrate the impact of both QC filtering steps, such as those shown in (Fig. 1) and cluster labelling strategies (Fig. 3), on each individual biological replicate. By showing this as Supporting Data in manuscripts describing scRNA-Seq data, it becomes easier for experts in the field and other interested parties to evaluate the results.

In summary, our reanalysis of recently published scRNA-Seq data of HSV-1-infected mouse TG does not support the reported detection of HSV-1 *LAT* RNA in non-neuronal cells. While studies investigating the virus and host factors contributing to viral latency and reactivation at the single-cell resolution will undoubtedly advance our understanding of these processes, it is critical to adhere to the accepted best practices for the design and analysis of scRNA-Seq data and to share both the datasets and code used for analysis.

## MATERIALS AND METHODS

### Data sourcing

The raw data files (FASTQ) associated with the original study of Wang et al. (8) (SRA PRJNA937697) were downloaded from the sequence read archive using `fastq_dump` from the SRA tool kit v2.10.9 (<https://github.com/ncbi/sra-tools>). Filtered count matrices generated by Wang et al. (8) were downloaded from the Gene Expression Omnibus archive (GEO GSE225839) in order to reproduce the analyses. It was not possible to reproduce the Cell Ranger (10X Genomics) analysis as the original authors did not provide sufficient detail on how reference genomes and annotations were generated, nor what parameters were used when running Cell Ranger.

### Analysis of read alignments against HSV-1

The 10 x Genomics 3' v3.1 datasets comprise three sets of reads, the I1 reads which contain the sample index, the R1 reads which contain the cellular barcodes and UMIs,

and the R2 reads which contain the transcriptome sequences. To examine the nature and numbers of reads derived from the HSV-1 McKrae transcriptome, we first performed quality and adapter trimming of the R2 reads using TrimGalore (clip\_R1 3 -q 30-length 50) (<https://github.com/FelixKrueger/TrimGalore>) before performing a competitive alignment against the *Mus musculus* (mm10) and HSV-1 McKrae (MN136524.1) genomes using STAR (49). Resulting SAM files were parsed using SAMTools (50) and BEDTools (51) to generate bedGraph files that could be visualized using R v4.2.3 in RStudio v 2023.06.2 using the packages Gviz (52) and GenomicFeatures (53). The scripts detailing this analysis (hsv1\_lat\_region\_analysis.sh) and visualization (CoveragePlotting-HSV1-McKrae.R) are available via Zenodo (see data availability section).

### Preprocessing of Wang et al. filtered count matrices

Filtered count matrices from all three datasets (Uninf-1, Inf-1, and Inf-2) were imported and analyzed using the Seurat package (v5) in R (v4.3.2) (54). Since the Wang et al. (8) original analysis scripts were not provided, we attempted to use parameters from the manuscript where possible to recreate the analysis. However, our analysis differed from Wang et al. in two ways. Firstly, although Wang et al. mentions integration of the three sample datasets, no information was provided about the method or parameters used. While the integration of datasets is not strictly required and is generally more useful for multimodal datasets, we tested whether integration affected clustering and cell-type annotation results. Integration was performed using the workflow described in [https://satijalab.org/seurat/archive/v4.3/integration\\_introduction](https://satijalab.org/seurat/archive/v4.3/integration_introduction) with default settings. Results were compared against a simple merge() of the datasets. Secondly, Wang et al. (8) provided RNA- and UMI-level filter metrics for LAT- cells only, and it was unclear what filters were used on the full dataset. We chose not to apply different filters to different subsets of data as this could not reasonably be rationalized and remains a significant error of the original analysis. Mitochondrial filtering was performed at the same level as Wang et al. (8) (15%), and only cells with 300–9,000 features (i.e., distinct transcript identities) were retained.

### Merging of datasets and clustering

Datasets for the three samples (Uninf-1, Inf-1 and Inf-2) were combined using the merge() function in Seurat. For tSNE and clustering, we used parameters from Wang et al. (8) where available: e.g., top 20 principal components and resolution 0.1.

### Cell type identification and cell death marker analysis

Cell type annotation was performed using the same package (singleR (28)) and reference databases as Wang et al. (8) (MouseRNAseqData (29) and ImmGenData (30) from the celldex package (<https://github.com/LTLA/celldex>). Low-confidence assignments were marked as NA using pruneScores() with default settings. Cell death markers were obtained from DeathBase (31) a curated database of core genes involved in cell death. All core genes involved in apoptosis and necroptosis (combined 40 unique genes) were included. We extracted the log-normalized counts for each of these genes from the imported Wang et al. barcode matrices and plotted all non-zero expression values.

### Doublet identification

Doublet identification was performed using DoubletFinder (<https://github.com/chris-mcginnis-ucsf/DoubletFinder>) either on the merged dataset, or, in the case where integration was performed, on individual sample datasets prior to integration. In both cases, cells identified as doublets were removed by subsetting cell barcodes. Clustering and cell type identification were performed as described above.

## ACKNOWLEDGMENTS

W.J.D.O. and G.G.M.G.M.V. are supported by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) under award number R01-AI151290. Computational analyses were supported by Fred Hutch Scientific Computing (NIH ORIP grant S10OD028685). P.R. and K.R.J. are supported by the NIAID grant R01-AI132599. A.L.C. is supported by Australian NHMRC Level 3 Investigator grant no. APP1177942. D.M.K. and G.G.M.G.M.V. are supported by NIH grants 75N93019C00063 and R01-AG064800. P.R.K. is supported by the NIAID grant R01-AI151290 and a National Eye Institute CORE Award (P30 EY08098). I.M. is supported by the NIH grants R01-AI073898, R01-AI152543, R01-AI176335, and R01-GM056927. A.C.W. is supported by the NIAID grants R01-AI176335 and R01-AI170583. D.P.D. is supported by a German Centre for Infection Research (DZIF) Professorship, the NIAID grants R01-AI170583 and R01-AI152543, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2155 - project number 390874280.

## AUTHOR AFFILIATIONS

<sup>1</sup>HerpesLabNL, Department of Viroscience, Erasmus Medical Center, Rotterdam, the Netherlands

<sup>2</sup>Department of Laboratory Medicine and Pathology, University of Washington, Seattle, Washington, USA

<sup>3</sup>Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Center, Seattle, Washington, USA

<sup>4</sup>Centre for Virus Research, The Westmead Institute for Medical Research, Sydney, New South Wales, Australia

<sup>5</sup>Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia

<sup>6</sup>Department of Medicine, University of Washington, Seattle, Washington, USA

<sup>7</sup>Department of Global Health, University of Washington, Seattle, Washington, USA

<sup>8</sup>Department of Translational Research, Benaroya Research Institute, Seattle, Washington, USA

<sup>9</sup>Department of Ophthalmology and of Molecular Microbiology and Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

<sup>10</sup>Department of Microbiology, New York University School of Medicine, New York, New York, USA

<sup>11</sup>Institute of Virology, Hannover Medical School, Hannover, Germany

<sup>12</sup>German Center for Infection Research (DZIF) partner site Hannover-Braunschweig, Hannover, Germany

<sup>13</sup>Cluster of Excellence RESIST (EXC 2155), Hannover Medical School, Hannover, Germany

## AUTHOR ORCID*s*

Werner J. D. Ouwendijk  <http://orcid.org/0000-0001-8393-296X>

Paul R. Kinchington  <http://orcid.org/0000-0002-1901-9970>

Georges G. M. G. M. Verjans  <http://orcid.org/0000-0002-2465-2674>

Daniel P. Depledge  <http://orcid.org/0000-0002-4292-0599>

## FUNDING

Funder	Grant(s)	Author(s)
HHS   National Institutes of Health (NIH)	R01-AI151290	Werner J. D. Ouwendijk Paul R. Kinchington Georges G. M. G. M. Verjans
HHS   National Institutes of Health (NIH)	R01-AI132599	Pavitra Roychoudhury

Funder	Grant(s)	Author(s)
		Keith R. Jerome
DHAC   National Health and Medical Research Council (NHMRC)	APP1177942	Anthony L. Cunningham
HHS   National Institutes of Health (NIH)	75N93019C00063, R01-AG064800	David M. Koelle Georges G. M. G. M. Verjans
HHS   NIH   National Eye Institute (NEI)	P30 EY08098	Paul R. Kinchington
HHS   National Institutes of Health (NIH)	R01-AI073898, R01-GM056927	Ian Mohr
HHS   National Institutes of Health (NIH)	R01-AI152543	Ian Mohr Daniel P. Depledge
HHS   National Institutes of Health (NIH)	R01-AI170583	Angus C. Wilson Daniel P. Depledge
HHS   National Institutes of Health (NIH)	R01-AI176335	Ian Mohr Angus C. Wilson

## AUTHOR CONTRIBUTIONS

Werner J. D. Ouwendijk, Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review and editing | Pavitra Roychoudhury, Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing | Anthony L. Cunningham, Conceptualization, Supervision, Writing – review and editing | Keith R. Jerome, Conceptualization, Funding acquisition, Supervision, Writing – review and editing | David M. Koelle, Conceptualization, Supervision, Writing – review and editing | Paul R. Kinchington, Conceptualization, Supervision, Writing – review and editing | Ian Mohr, Conceptualization, Supervision, Writing – review and editing | Angus C. Wilson, Conceptualization, Supervision, Writing – review and editing | Georges G. M. G. M. Verjans, Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review and editing, Project administration, Resources | Daniel P. Depledge, Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing

## DATA AVAILABILITY

All scripts used in the analyses presented herein are available from [10.5281/zenodo.10551298](https://doi.org/10.5281/zenodo.10551298).

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Fig. S1 (JVI01858-23-s0001.tif).** Doublet identification in merged datasets.

**Fig. S2 (JVI01858-23-s0002.tif).** Doublet identification in integrated datasets.

**Fig. S3 (JVI01858-23-s0003.tif).** Further evidence that abundant loss of reads during filtering is specifically associated with HSV-1 LAT and neuronal markers.

**Table S1 (JVI01858-23-s0004.xlsx).** Cell death markers.

**Table S2 (JVI01858-23-s0005.xlsx).** Alignment metrics.

## REFERENCES

- Knipe DM, Heldwein EE, Mohr I, Sodroski CN. 2021. Herpes Simplex viruses: Mechanisms of Lytic and latent infection, p 235–296. In *Field's Virology: DNA viruses*, 7th ed. Lippincott Williams & Wilkins, Philadelphia.
- Bastian FO, Rabson AS, Yee CL, Tralka TS. 1972. Herpesvirus hominis: isolation from human trigeminal ganglion. *Science* 178:306–307. <https://doi.org/10.1126/science.178.4058.306>
- Vrabec JT, Payne DA. 2001. Prevalence of herpesviruses in cranial nerve ganglia. *Acta Otolaryngol* 121:831–835. <https://doi.org/10.1080/00016480152602285>
- Nagel MA, Rempel A, Huntington J, Kim F, Choe A, Gilden D. 2014. Frequency and abundance of alphaherpesvirus DNA in human thoracic sympathetic ganglia. *J Virol* 88:8189–8192. <https://doi.org/10.1128/JVI.01070-14>
- Kennedy DP, Clement C, Arceneaux RL, Bhattacharjee PS, Huq TS, Hill JM. 2011. Ocular herpes simplex virus type 1: is the cornea a reservoir for viral latency or a fast pit stop? *Cornea* 30:251–259. <https://doi.org/10.1097/ICO.0b013e3181ef241d>
- Remeijer L, Duan R, van Dun JM, Wefers Bettink MA, Osterhaus ADME, Verjans GMGM. 2009. Prevalence and clinical consequences of herpes simplex virus type 1 DNA in human cornea tissues. *J Infect Dis* 200:11–19. <https://doi.org/10.1086/599329>
- Hu H-L, Srinivas KP, Wang S, Chao MV, Lionnet T, Mohr I, Wilson AC, Depledge DP, Huang TT. 2022. Single-cell transcriptomics identifies Gadd45b as a regulator of herpesvirus-reactivating neurons. *EMBO Rep* 23:e53543. <https://doi.org/10.15252/embr.202153543>
- Wang S, Song X, Rajewski A, Santiskulvong C, Ghiasi H. 2023. Stacking the odds: multiple sites for HSV-1 latency. *Sci Adv* 9:eadf4904. <https://doi.org/10.1126/sciadv.adf4904>
- Stevens JG, Wagner EK, Devi-Rao GB, Cook ML, Feldman LT. 1987. RNA complementary to a herpesvirus alpha gene mRNA is prominent in latently infected neurons. *Science* 235:1056–1059. <https://doi.org/10.1126/science.2434993>
- Rock DL, Nesburn AB, Ghiasi H, Ong J, Lewis TL, Lokensgard JR, Wechsler SL. 1987. Detection of latency-related viral RNAs in trigeminal ganglia of rabbits latently infected with herpes simplex virus type 1. *J Virol* 61:3820–3826. <https://doi.org/10.1128/JVI.61.12.3820-3826.1987>
- Deatly AM, Spivack JG, Lavi E, Fraser NW. 1987. RNA from an immediate early region of the type 1 herpes simplex virus genome is present in the trigeminal ganglia of latently infected mice. *Proc Natl Acad Sci U S A* 84:3204–3208. <https://doi.org/10.1073/pnas.84.10.3204>
- Umbach JL, Nagel MA, Cohrs RJ, Gilden DH, Cullen BR. 2009. Analysis of human alphaherpesvirus microRNA expression in latently infected human trigeminal ganglia. *J Virol* 83:10677–10683. <https://doi.org/10.1128/JVI.01185-09>
- 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 61. *Nat Commun* 9:1167. <https://doi.org/10.1038/s41467-018-03569-2>
- Umbach JL, Kramer MF, Jurak I, Karnowski HW, Coen DM, Cullen BR. 2008. MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature* 454:780–783. <https://doi.org/10.1038/nature07103>
- Zwaagstra JC, Ghiasi H, Slanina SM, Nesburn AB, Wheatley SC, Lillycrop K, Wood J, Latchman DS, Patel K, Wechsler SL. 1990. Activity of herpes simplex virus type 1 latency-associated transcript (LAT) promoter in neuron-derived cells: evidence for neuron specificity and for a large LAT transcript. *J Virol* 64:5019–5028. <https://doi.org/10.1128/JVI.64.10.5019-5028.1990>
- Puga A, Notkins AL. 1987. Continued expression of a poly(A)+ transcript of herpes simplex virus type 1 in trigeminal ganglia of latently infected mice. *J Virol* 61:1700–1703. <https://doi.org/10.1128/JVI.61.5.1700-1703.1987>
- Dobson AT, Sederati F, Devi-Rao G, Flanagan WM, Farrell MJ, Stevens JG, Wagner EK, Feldman LT. 1989. Identification of the latency-associated transcript promoter by expression of rabbit beta-globin mRNA in Mouse sensory nerve ganglia latently infected with a recombinant herpes simplex virus. *J Virol* 63:3844–3851. <https://doi.org/10.1128/JVI.63.9.3844-3851.1989>
- Devi-Rao GB, Goodart SA, Hecht LM, Rochford R, Rice MK, Wagner EK. 1991. Relationship between polyadenylated and nonpolyadenylated herpes simplex virus type 1 latency-associated transcripts. *J Virol* 65:2179–2190. <https://doi.org/10.1128/JVI.65.5.2179-2190.1991>
- Wu TT, Su YH, Block TM, Taylor JM. 1996. Evidence that two latency-associated transcripts of herpes simplex virus type 1 are nonlinear. *J Virol* 70:5962–5967. <https://doi.org/10.1128/JVI.70.9.5962-5967.1996>
- Wu TT, Su YH, Block TM, Taylor JM. 1998. Atypical splicing of the latency-associated transcripts of herpes simplex type 1. *Virology* 243:140–149. <https://doi.org/10.1006/viro.1998.9036>
- Phelan D, Barrozo ER, Bloom DC. 2017. HSV1 latent transcription and non-coding RNA: a critical retrospective. *J Neuroimmunol* 308:65–101. <https://doi.org/10.1016/j.jneuroim.2017.03.002>
- Heumos L, Schaar AC, Lance C, Litnetskaya A, Drost F, Zappia L, Lücken MD, Strobl DC, Henao J, Curion F, Single-cell Best Practices Consortium, Schiller HB, Theis FJ. 2023. Best practices for single-cell analysis across modalities. *Nat Rev Genet* 24:550–572. <https://doi.org/10.1038/s41576-023-00586-w>
- Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, Teichmann SA. 2016. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* 17:29. <https://doi.org/10.1186/s13059-016-0888-1>
- Osorio D, Cai JJ. 2021. Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control. *Bioinformatics* 37:963–967. <https://doi.org/10.1093/bioinformatics/btaa751>
- Franzén O, Gan L-M, Björkegren JLM. 2019. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* 2019:baz046. <https://doi.org/10.1093/database/baz046>
- Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, Bai J, Zhao T, Li X, Xiao Y. 2019. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res* 47:D721–D728. <https://doi.org/10.1093/nar/gky900>
- Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, et al. 2018. Mapping the mouse cell atlas by microwell-seq. *Cell* 172:1091–1107. <https://doi.org/10.1016/j.cell.2018.02.001>
- Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, Butte AJ, Bhattacharya M. 2019. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 20:163–172. <https://doi.org/10.1038/s41590-018-0276-y>
- Benayoun BA, Pollina EA, Singh PP, Mahmoudi S, Harel I, Casey KM, Dulken BW, Kundaje A, Brunet A. 2019. Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res* 29:697–709. <https://doi.org/10.1101/gr.240093.118>
- Heng TSP, Painter MW, Elpek K, Lukacs-Kornek V, Mauermann N, Turley SJ. 2008. The immunological genome project: networks of gene expression in immune cells. *Nat Immunol* 9:1091–1094. <https://doi.org/10.1038/ni1008-1091>
- Díez J, Walter D, Muñoz-Pinedo C, Gabaldón T. 2010. DeathBase: a database on structure, evolution and function of proteins involved in apoptosis and other forms of cell death. *Cell Death Differ* 17:735–736. <https://doi.org/10.1038/cdd.2009.215>
- McGinnis CS, Murrow LM, Gartner ZJ. 2019. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* 8:329–337. <https://doi.org/10.1016/j.cels.2019.03.003>
- Svoboda M, Frost HR, Bosco G. 2022. Internal oligo(dT) priming introduces systematic bias in bulk and single-cell RNA sequencing count data. *NAR Genom Bioinform* 4:lqac035. <https://doi.org/10.1093/nargab/lqac035>
- Aubert M, Strongin DE, Roychoudhury P, Loprieno MA, Haick AK, Klouner LM, Stensland L, Huang M-L, Makhosous N, Tait A, De Silva Felixge HS, Galetto R, Duchateau P, Greninger AL, Stone D, Jerome KR. 2020. Gene editing and elimination of latent herpes simplex virus *in vivo*. *Nat Commun* 11:4148. <https://doi.org/10.1038/s41467-020-17936-5>



35. Kramer MF, Coen DM. 1995. Quantification of transcripts from the ICP4 and thymidine kinase genes in mouse ganglia latently infected with herpes simplex virus. *J Virol* 69:1389–1399. <https://doi.org/10.1128/JVI.69.3.1389-1399.1995>
36. Young MD, Behjati S. 2020. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* 9:giaa151. <https://doi.org/10.1093/gigascience/giaa151>
37. Janssen P, Kliemete Z, Vieth B, Adiconis X, Simmons S, Marshall J, McCabe C, Heyn H, Levin JZ, Enard W, Hellmann I. 2023. The effect of background noise and its removal on the analysis of single-cell expression data. *Genome Biol* 24:140. <https://doi.org/10.1186/s13059-023-02978-x>
38. Sawtell NM, Thompson RL. 2016. Herpes simplex virus and the lexicon of latency and reactivation: a call for defining terms and building an integrated collective framework. *F1000Res* 5:F1000 Faculty Rev-2038. <https://doi.org/10.12688/f1000research.8886.1>
39. Cliffe AR, Wilson AC. 2017. Restarting lytic gene transcription at the onset of herpes simplex virus reactivation. *J Virol* 91:e01419-16. <https://doi.org/10.1128/JVI.01419-16>
40. Cook ML, Bastone VB, Stevens JG. 1974. Evidence that neurons harbor latent herpes simplex virus. *Infect Immun* 9:946–951. <https://doi.org/10.1128/iai.9.5.946-951.1974>
41. Stevens JG, Cook ML. 1971. Latent herpes simplex virus in spinal ganglia of mice. *Science* 173:843–845. <https://doi.org/10.1126/science.173.3999.843>
42. Paine TF. 1964. Latent herpes simplex infection in man. *Bacteriol Rev* 28:472–479. <https://doi.org/10.1128/br.28.4.472-479.1964>
43. Goodpasture EW. 1993. Herpetic infection, with especial reference to involvement of the nervous system. 1929. *Medicine (Baltimore)* 72:125–132. <https://doi.org/10.1097/00005792-199303000-00006>
44. Yang L, Voytek CC, Margolis TP. 2000. Immunohistochemical analysis of primary sensory neurons latently infected with herpes simplex virus type 1. *J Virol* 74:209–217. <https://doi.org/10.1128/jvi.74.1.209-217.2000>
45. Wang K, Lau TY, Morales M, Mont EK, Straus SE. 2005. Laser-capture microdissection: refining estimates of the quantity and distribution of latent herpes simplex virus 1 and varicella-zoster virus DNA in human trigeminal ganglia at the single-cell level. *J Virol* 79:14079–14087. <https://doi.org/10.1128/JVI.79.22.14079-14087.2005>
46. Maroui MA, Callé A, Cohen C, Streichenberger N, Texier P, Takissian J, Rousseau A, Pocard N, Welsch J, Corpet A, Schaeffer L, Labetoulle M, Lomonte P. 2016. Latency entry of herpes simplex virus 1 is determined by the interaction of its genome with the nuclear environment. *PLoS Pathog* 12:e1005834. <https://doi.org/10.1371/journal.ppat.1005834>
47. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3:160018. <https://doi.org/10.1038/sdata.2016.18>
48. Rocca-Serra P, Gu W, Ioannidis V, Abbassi-Daloui T, Capella-Gutierrez S, Chandramouliswaran I, Splendiani A, Burdett T, Giessmann RT, Henderson D, et al. 2023. The FAIR cookbook - the essential resource for and by FAIR doers. *Sci Data* 10:292. <https://doi.org/10.1038/s41597-023-02166-3>
49. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
50. 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>
51. 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>
52. 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)
53. 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>
54. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of single-cell data. *Cell* 177:1888–1902. <https://doi.org/10.1016/j.cell.2019.05.031>