



# A Single-Cell Approach to the Elusive Latent Human Cytomegalovirus Transcriptome

Felicia Goodrum,<sup>a</sup> Shannon McWeeney<sup>b</sup>

<sup>a</sup>University of Arizona, Tucson, Arizona, USA

<sup>b</sup>Oregon Health and Sciences University, Portland, Oregon, USA

**ABSTRACT** Herpesvirus latency has been difficult to understand molecularly due to low levels of viral genomes and gene expression. In the case of the betaherpesvirus human cytomegalovirus (HCMV), this is further complicated by the heterogeneity inherent to hematopoietic subpopulations harboring genomes and, as a consequence, the various patterns of infection that simultaneously exist in a host, ranging from latent to lytic. Single-cell RNA sequencing (scRNA-seq) provides tremendous potential in measuring the gene expression profiles of heterogeneous cell populations for a wide range of applications, including in studies of cancer, immunology, and infectious disease. A recent study by Shnayder et al. (mBio 9:e00013-18, 2018, <https://doi.org/10.1128/mBio.00013-18>) utilized scRNA-seq to define transcriptomal characteristics of HCMV latency. They conclude that latency-associated gene expression is similar to the late lytic viral program but at lower levels of expression. The study highlights the numerous challenges, from the definition of latency to the analysis of scRNA-seq, that exist in defining a latent transcriptome.

**KEYWORDS** MARS-Seq, cytomegalovirus, herpesviruses, scRNA-seq, transcriptome

Herpesviruses are incurable due to their ability to establish a latent infection that cannot be targeted by current antivirals. The persistence of herpesviruses through latency results in an enduring health risk, as their reactivation from latency is associated with significant disease, particularly in immunocompromised individuals. Understanding viral latency at the molecular level offers the opportunity to develop strategies to identify and target latently infected cells and to reveal mechanisms of viral persistence.

Viral latency, as classically defined, is the maintenance of viral genomes in the absence of progeny virus production, but with the ability to reactivate for subsequent rounds of replication. During latency, the viral genome is chromatinized, and the latent transcriptional program has been presumed to be restricted to a limited number of viral RNAs, proteins, and microRNAs. Indeed, latency has often been defined experimentally as the absence of a single lytic gene in the presence of a single latency-associated gene. However, highly sensitive, global analysis of gene expression during infection with human cytomegalovirus (HCMV) in cell types supporting a latent infection paints a far more complex picture of viral gene expression associated with latent states (1–5). Similar findings have been made for alphaherpesviruses (6–10). Therefore, Shnayder et al.'s study (11) joins a growing body of herpesvirus literature that challenges the paradigm that latency is a transcriptionally silent state, but with the surprising conclusion that there is no unique latent transcriptome, suggesting a model whereby latency reflects a quantitative, not qualitative, difference in viral gene expression from that of the replicative state.

Two major challenges to defining the latent transcriptome of HCMV exist. First, the viral genome is estimated to be maintained at 2 to 13 copies per cell in  $\leq 1$  in 10,000 mononuclear cells in an infected host (12), and transcription is restricted (13, 14).

**Published** 12 June 2018

**Citation** Goodrum F, McWeeney S. 2018. A single-cell approach to the elusive latent human cytomegalovirus transcriptome. mBio 9:e01001-18. <https://doi.org/10.1128/mBio.01001-18>.

**Copyright** © 2018 Goodrum and McWeeney. 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 Felicia Goodrum, [fgoodrum@email.arizona.edu](mailto:fgoodrum@email.arizona.edu).

For the article discussed, see <https://doi.org/10.1128/mBio.00013-18>.

*The views expressed in this Commentary do not necessarily reflect the views of this journal or of ASM.*

Second, HCMV establishes latency in hematopoietic progenitor cells (HPCs), which are a heterogeneous population of cells with dynamic programs of differentiation that are sensitive to *ex vivo* stimulation. This heterogeneity especially complicates transcriptome analysis, since HCMV reactivation from latency is associated with myeloid lineage differentiation (15, 16). Therefore, it follows that the viral transcriptome may vary along a spectrum from latent to replicative in cells at different stages of differentiation within an HPC population.

To define viral genes associated with natural infection in the host, Shnayder et al. searched the existing bank of transcriptome sequencing (RNA-seq) data for the presence of HCMV transcripts across human tissues (11). The paucity of HCMV gene expression in seropositive (presumably latently infected) individuals is reflected in the fact that an analysis of 1.5 billion aligned sequence reads from 25 samples of hematopoietic origin identified no bona fide CMV reads. Further analysis of the GTEx database (433 billion reads derived from 9,416 RNA-seq samples from 549 individuals spanning 31 tissues) identified only 6 individuals with  $\geq 10$  reads in any tissue. Hierarchical clustering of samples with  $>4$  reads resolved two patterns of gene expression, one dominated by long noncoding RNAs (lncRNAs) (group I) and another dominated by immediate early genes (group II). Group I and II correlated to infection in fibroblasts at late times (24 to 72 h postinfection [hpi]) and early times (3 to 5 hpi), respectively. However, caution must be taken with these correlations given the small number of genes detected and the low abundance of reads. Further, because lncRNAs are the most abundantly expressed transcripts in any context of infection (2, 17–19), the presence of lncRNAs should not be taken as evidence of a lytic infection. One approach to the challenge of low-abundance HCMV reads in seropositive individuals is the use of targeted enrichment approaches to increase the proportion of viral reads in a mixed transcriptome (1).

Single-cell RNA sequencing (scRNA-seq) following *in vitro* infection has the potential to skirt the issues imposed by heterogeneity of the cell population and, indeed, offers a powerful approach to defining the heterogeneity of the viral transcriptome across the hematopoietic hierarchy of differentiation. In the study by Shnayder et al., 3,655 CD14<sup>+</sup> cells were analyzed by massively parallel RNA single-cell sequencing framework (MARS-Seq) analysis and formed 6 clusters based on host and viral gene expression. In clusters 1 to 6, viral gene expression ranged from  $>10\%$  to 0% of the entire transcriptome, respectively (11). Due to its high proportion of viral gene expression, cluster 1 was concluded to represent lytically infected cells. This conclusion is based, in part, on the correlation to the HCMV transcriptome in monocyte-derived macrophages and fibroblasts. The high correlation between the cluster 1 CD14<sup>+</sup> transcriptome and the transcriptome from infection in fibroblasts and macrophages is likely driven by the high expression of lncRNAs that mask the impact of the large number of genes that were expressed in fibroblasts or macrophages but not expressed in CD14<sup>+</sup> monocytes across the different clusters. Additional sensitivity analysis, with the highly abundant lncRNAs removed, to assess the conclusions of both the clustering and the correlation would be important to confidently compare these infection models. Further, it is interesting to note that essentially all cells represented in cluster 1 were from the early time points of infection (3, 4, and 5 days postinfection [dpi]). It is expected that lytically infected cells with this profile of gene expression in the population would be represented in all of the time points and, perhaps to an even greater extent, at later time points postinfection. Because cluster 1 is comprised of cells at early time points postinfection, it is possible that these cells undergo an initial burst of gene expression prior to establishing a more quiescent state and might influence latency, as described for alphaherpesviruses (6, 20–22).

The CD14<sup>+</sup> gene expression clusters continued to segregate by time postinfection, with clusters 5 and 6 being comprised predominantly of cells from 7 and 14 dpi, respectively. The latter clusters are comprised of cells with little to no evident viral gene expression ( $<0.1\%$  of total reads were HCMV reads). In cells where no viral transcripts were detected, the viral genome may be highly repressed or these cells may not have

been initially infected, since the frequency of infections based on green fluorescent protein expression from the viral genome was ~62%. Further, due to the cutoff of  $\geq 1,000$  reads per cell, it is possible that the depth of sequencing may not be able to capture the full transcriptome, particularly for very-low-abundance RNAs. Measuring the maintenance of viral genomes over this time course is essential to understanding whether this apparent loss of HCMV gene expression over time is due to the establishment of a more quiescent but reactivatable state or whether it reflects an abortive infection where viral genomes are lost.

scRNA-seq analysis of infected CD34<sup>+</sup> HPCs presented even greater challenges than CD14<sup>+</sup> cells due to the yet-lower abundance of viral transcripts and to the fact that transcripts were detected in only ~5% of the cells sequenced. Again, it is difficult to ascertain whether the failure to detect transcripts is due to their low abundance or to a truly quiescent infection or because the cells were not infected. Shnayder et al. compared the HCMV transcriptome in CD34<sup>+</sup> HPCs to that in cluster 1 (>10% viral gene expression) from CD14<sup>+</sup> monocytes and concluded that the CD34<sup>+</sup> HPC latency-associated transcriptome largely mirrored that of a lytic infection, being differentiated from a lytic state only by the level of expression (11). The associated correlation (Spearman coefficient of 0.67) is likely heavily driven by high expression of lncRNAs in both contexts, despite the fact that a large number of genes are expressed in a “lytic” infection but not in CD34<sup>+</sup> HPCs. Further, because CD34<sup>+</sup> scRNA-seq was performed at 4 dpi, latency may not yet have been established and the gene expression observed may reflect an early burst of gene expression.

While these studies reveal a transcriptome associated with infection in hematopoietic subpopulations that extends beyond a few genes, consistent with the results of a number of contemporary studies (1, 2), the conclusion that there is no difference in latent and lytic transcriptomes is more difficult to assert. A critical caveat in interpreting scRNA-seq analysis is the sparsity of the data, i.e., the high proportion of zero read counts (23, 24), which often can be due to technical reasons (such as “dropouts,” where a gene is expressed but not detected in a sequencing run) rather than the biological factors of interest (lack of expression in a specific cell type or time point). The limitation in scRNA-seq coverage requires  $\geq 50,000$  reads for unbiased transcriptome definition from low-abundance transcripts (25) and represents a remaining challenge to defining the latent HCMV transcriptome in cells with low levels of viral gene expression. Given this, as well as the inherently stochastic and heterogeneous nature of scRNA-seq, care must be taken with regard to appropriate normalization to address within-sample (e.g., GC content bias, transcript length) and between-sample (e.g., sequencing depth) issues. The unique molecular identifier (UMI)-based protocol utilized by Shnayder et al. can remove sequencing depth bias if the libraries are sequenced to saturation, but UMIs still cannot account for capture efficiency or cellular mRNA content differences (26). While the authors estimated leakage noise in the MARS-seq analysis, no formal normalization was conducted to address these technical issues, and only raw reads were utilized for the 10 $\times$  sequencing data. Further, while scRNA-seq offers immense promise in understanding the variation in profiles of gene expression between cells in a population, challenges in defining complex transcriptomes, such as those comprised of viral and host sequences, remain.

The study by Shnayder et al. brings the power of single-cell sequencing to HCMV infection and demonstrates an array of profiles of viral mRNAs and lncRNAs, from broad to narrow, expressed in CD14<sup>+</sup> and CD34<sup>+</sup> hematopoietic cells infected *in vitro* (11). Going forward, it will be important to understand the profiles of gene expression associated with specific cell types within the larger CD14<sup>+</sup> and CD34<sup>+</sup> populations and how both viral and host gene expression changes over time and cellular differentiation from initial infection to the establishment of latency and, finally, reactivation.

## REFERENCES

1. Cheng S, Caviness K, Buehler J, Smithey M, Nikolich-Zugich J, Goodrum F. 2017. Transcriptome-wide characterization of human cytomegalovirus in natural infection and experimental latency. *Proc Natl Acad Sci U S A* 114:E10586–E10595. <https://doi.org/10.1073/pnas.1710522114>.

2. Rossetto CC, Tarrant-Elorza M, Pari GS. 2013. Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLoS Pathog* 9:e1003366. <https://doi.org/10.1371/journal.ppat.1003366>.
3. Goodrum F, Jordan CT, Terhune SS, High K, Shenk T. 2004. Differential outcomes of human cytomegalovirus infection in primitive hematopoietic cell subpopulations. *Blood* 104:687–695. <https://doi.org/10.1182/blood-2003-12-4344>.
4. Cheung AK, Abendroth A, Cunningham AL, Slobedman B. 2006. Viral gene expression during the establishment of human cytomegalovirus latent infection in myeloid progenitor cells. *Blood* 108:3691–3699. <https://doi.org/10.1182/blood-2005-12-026682>.
5. Goodrum FD, Jordan CT, High K, Shenk T. 2002. Human cytomegalovirus gene expression during infection of primary hematopoietic progenitor cells: a model for latency. *Proc Natl Acad Sci U S A* 99:16255–16260. <https://doi.org/10.1073/pnas.252630899>.
6. Ma JZ, Russell TA, Spelman T, Carbone FR, Tschärke DC. 2014. Lytic gene expression is frequent in HSV-1 latent infection and correlates with the engagement of a cell-intrinsic transcriptional response. *PLoS Pathog* 10:e1004237. <https://doi.org/10.1371/journal.ppat.1004237>.
7. Baird NL, Bowlin JL, Cohrs RJ, Gilden 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>.
8. Kramer MF, Chen SH, Knipe DM, Coen DM. 1998. Accumulation of viral transcripts and DNA during establishment of latency by herpes simplex virus. *J Virol* 72:1177–1185.
9. Nagel MA, Choe A, Traktinskiy I, Cordery-Cotter R, Gilden D, Cohrs RJ. 2011. Varicella-zoster virus transcriptome in latently infected human ganglia. *J Virol* 85:2276–2287. <https://doi.org/10.1128/JVI.01862-10>.
10. Giordani NV, Neumann DM, Kwiatkowski DL, Bhattacharjee PS, McAnany PK, Hill JM, Bloom DC. 2008. During herpes simplex virus type 1 infection of rabbits, the ability to express the latency-associated transcript increases latent-phase transcription of lytic genes. *J Virol* 82:6056–6060. <https://doi.org/10.1128/JVI.02661-07>.
11. Shnyder M, Nachshon A, Krishna B, Poole E, Boshkov A, Binyamin A, Maza I, Sinclair J, Schwartz M, Stern-Ginossar N. 2018. Defining the transcriptional landscape during cytomegalovirus latency with single-cell RNA sequencing. *mBio* 9:e00013-18. <https://doi.org/10.1128/mBio.00013-18>.
12. Slobedman B, Mocarski ES. 1999. Quantitative analysis of latent human cytomegalovirus. *J Virol* 73:4806–4812.
13. Dağ F, Dölken L, Holzik J, Drabig A, Weingärtner A, Schwerk J, Lienenklaus S, Conte I, Geffers R, Davenport C, Rand U, Köster M, Weiß S, Adler B, Wirth D, Messerle M, Hauser H, Čičin-Šain L. 2014. Reversible silencing of cytomegalovirus genomes by type I interferon governs virus latency. *PLoS Pathog* 10:e1003962. <https://doi.org/10.1371/journal.ppat.1003962>.
14. Reeves M, Sinclair J. 2013. Regulation of human cytomegalovirus transcription in latency: beyond the major immediate-early promoter. *Viruses* 5:1395–1413. <https://doi.org/10.3390/v5061395>.
15. Söderberg-Nauclér C, Streblo DN, Fish KN, Allan-Yorke J, Smith PP, Nelson JA. 2001. Reactivation of latent human cytomegalovirus in CD14(+) monocytes is differentiation dependent. *J Virol* 75:7543–7554. <https://doi.org/10.1128/JVI.75.16.7543-7554.2001>.
16. Taylor-Wiedeman J, Sissons P, Sinclair J. 1994. Induction of endogenous human cytomegalovirus gene expression after differentiation of monocytes from healthy carriers. *J Virol* 68:1597–1604.
17. Gatherer D, Seirafian S, Cunningham C, Holton M, Dargan DJ, Baluchova K, Hector RD, Galbraith J, Herzyk P, Wilkinson GWG, Davison AJ. 2011. High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci U S A* 108:19755–19760. <https://doi.org/10.1073/pnas.1115861108>.
18. Van Damme E, Thys K, Tuefferd M, Van Hove C, Aerssens J, Van Loock M. 2016. HCMV displays a unique transcriptome of immunomodulatory genes in primary monocyte-derived cell types. *PLoS One* 11:e0164843. <https://doi.org/10.1371/journal.pone.0164843>.
19. Raftery MJ, Möncke-Buchner E, Matsumura H, Giese T, Winkelmann A, Reuter M, Terauchi R, Schönrich G, Krüger DH. 2009. Unravelling the interaction of human cytomegalovirus with dendritic cells by using SuperSAGE. *J Gen Virol* 90:2221–2233. <https://doi.org/10.1099/vir.0.010538-0>.
20. Proença JT, Coleman HM, Connor V, Winton DJ, Efstathiou S. 2008. A historical analysis of herpes simplex virus promoter activation in vivo reveals distinct populations of latently infected neurones. *J Gen Virol* 89:2965–2974. <https://doi.org/10.1099/vir.0.2008/005066-0>.
21. Raja P, Lee JS, Pan D, Pesola JM, Coen DM, Knipe DM. 2016. A herpesviral lytic protein regulates the structure of latent viral chromatin. *mBio* 7:e00633-16. <https://doi.org/10.1128/mBio.00633-16>.
22. 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>.
23. Kharchenko PV, Silberman L, Scadden DT. 2014. Bayesian approach to single-cell differential expression analysis. *Nat Methods* 11:740–742. <https://doi.org/10.1038/nmeth.2967>.
24. Lun ATL, Bach K, Marioni JC. 2016. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* 17:133. <https://doi.org/10.1186/s13059-016-0998-9>.
25. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, Ramalingam N, Sun G, Thu M, Norris M, Lebofsky R, Toppani D, Kemp DW, Wong M, Clerkson B, Jones BN, Wu S, Knutsson L, Alvarado B, Wang J, Weaver LS, May AP, Jones RC, Unger MA, Kriegstein AR, West JAA. 2014. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol* 32:1053–1058. <https://doi.org/10.1038/nbt.2967>.
26. Vallejos CA, Rizzo D, Scialdone A, Dudoit S, Marioni JC. 2017. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat Methods* 14:565–571. <https://doi.org/10.1038/nmeth.4292>.