Single cell analysis reveals human cytomegalovirus drives latently infected cells towards an anergic-like monocyte state

Document Version:
Accepted author manuscript (peer-reviewed)

Citation for published version:
Shnayder, M, Nachshon, A, Rozman, B, Bernshtein, B, Lavi, M, Fein, N, Poole, E, Avdic, S, Blyth, E, Gottlieb, D, Abendroth, A, Slobedman, B, Sinclair, J, Stern-Ginossar, N & Schwartz, M 2020, 'Single cell analysis reveals human cytomegalovirus drives latently infected cells towards an anergic-like monocyte state', *eLife*, vol. 9, e52168. https://doi.org/10.7554/eLife.52168

Total number of authors:
15

Digital Object Identifier (DOI):
10.7554/eLife.52168

Published In:
eLife

(article begins on next page)

Single cell analysis reveals human cytomegalovirus drives latently infected cells towards an anergic-like monocyte state

Miri Shnayder[1], Aharon Nachshon[1] , Batsheva Rozman[1], Biana Bernstein[1], Michael Lavi[1], Noam Fein[1], Emma Poole[2], Selmir Avdic[3], Emily Blyth[3,4], David Gottlieb[3,4], Allison Abendroth[5], Barry Slobedman[5] , John Sinclair[2], Noam Stern-Ginossar[1*] and Michal Schwartz[1*]

[1] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

[2] Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge CB20QQ, UK

[3] Sydney Cellular Therapies Laboratory, Westmead, Sydney, NSW, Australia

[4] Blood and Bone Marrow Transplant Unit, Westmead Hospital, Sydney, NSW, Australia

[5] Discipline of Infectious Diseases and Immunology, Faculty of Medicine and Health, Charles Perkins Centre, University of Sydney, Sydney, NSW, Australia.

* To whom correspondence should be addressed; noam.stern-ginossar@weizmann.ac.il or michalsc@weizmann.ac.il

Abstract

Human cytomegalovirus (HCMV) causes a lifelong infection through establishment of latency. Although reactivation from latency can cause life-threatening disease, our molecular understanding of HCMV latency is incomplete. Here we use single cell RNA-seq analysis to characterize latency in monocytes and hematopoietic stem and progenitor cells (HSPCs). In monocytes, we identify host cell surface markers that enable enrichment of latent cells harboring higher viral transcript levels, which can reactivate more efficiently, and are characterized by reduced intrinsic immune response that is important for viral gene expression. Significantly, in latent HSPCs, viral transcripts could be detected only in monocyte progenitors and were also associated with reduced immune-response. Overall, our work indicates that regardless of the developmental stage in which HCMV infects, HCMV drives hematopoietic cells towards a weaker immune-responsive monocyte state and that this anergic-like state is crucial for the virus ability to express its transcripts and to eventually reactivate.

Introduction

Human cytomegalovirus (HCMV) is a prevalent pathogen of the beta-herpesvirus family, infecting the majority of the human population worldwide (Staras et al., 2006). Following primary infection, HCMV persists through the lifetime of the host by establishing latency. In the latent state, no viral progeny is produced but the virus maintains the capacity to reactivate. Reactivation in immunocompromised individuals, such as transplant recipients and HIV patients, leads to severe illness and mortality (Crough and Khanna, 2009). Despite the significant health burden that accompanies HCMV reactivation from latency, to date there is no treatment that targets the latent stage and the processes governing latency and reactivation are far from fully understood.

HCMV has a wide cell tropism within its human host (Mocarski, E.S., Shenk, T., 2013), with most cell types supporting lytic replication (C. Sinzger et al., 2008). In contrast, latent infection has so far been characterized in cells of the early myeloid lineage, including CD34+ hematopoietic stem and progenitor cells (HSPCs) and CD14+ monocytes *in vivo* (Mendelson et al., 1996; Taylor-Wiedeman et al., 1991; Von Laer et al., 1995). Since CD14+ monocytes are short-lived cells it has been proposed that the latent reservoir resides in hematopoietic stem cells (HSCs) (Slobedman et al., 2010) and that latent monocytes support viral spread and persistence within the infected host (Stevenson et al., 2014). Latent cells in HCMV seropositive individuals are scarce and were estimated by PCR-driven *in situ* hybridization, at 1:10,000 to 25,000 with a copy number of 2 to 13 genomes per infected cell (Slobedman and Mocarski, 1999). Using highly sensitive methodologies, such as digital PCR, viral genomes were detected in less than half of seropositive individuals and viral load was estimated at less than 10 genomes in 10,000 cells in most individuals (Jackson et al., 2017; Parry et al., 2016). Due to the scarcity of HCMV-infected cells in the natural context, *in vitro* HCMV infection of primary cells, mainly HSPCs and monocytes were developed as models. The caveat of these systems is their heterogeneity and the possibility that they may represent dynamic differentiation states. Additionally, although both models are being widely used, the differences between them are not well understood.

It is becoming increasingly evident that the repertoire of viral genes expressed during latent infections is broader than initially appreciated (Cheng et al., 2017; Schwartz and Stern-Ginossar, 2019; Shnayder et al., 2018). Despite low expression of viral transcripts, a number of studies have described infection driven changes in host cells during HCMV latency (Chan et al., 2010; Kew et al., 2017; Reeves et al., 2012; Slobedman et al., 2002; Smith et al., 2007, 2004). Monocyte infection was proposed to promote differentiation to macrophages with specific polarization towards genes that mark M1 phenotype with some atypical attributes (Chan et al., 2008; Smith et al., 2004). On the other hand it was shown that the viral homolog of human interleukin-10, encoded by UL111A, polarizes monocytes into an anti-inflammatory M2 subset (Avdic et al., 2013). The UL7 viral protein was found to bind Fms-like tyrosine kinase 3 receptor (Flt-3R), inducing differentiation of HSPCs to monocytes and of monocytes to macrophages (Crawford et al., 2018). Finally, analysis of HCMV infected HSPCs revealed reprogramming of HSPCs into immune-suppressive monocytes (Zhu et al., 2018). Thus, although it is clear that latent HCMV infection affects the differentiation state of infected HSPCs and monocytes, the nature of these effects is still enigmatic and controversial.

The studies to date examining host responses to latent HCMV infection focused on differences between infected and uninfected cells or on the effect of specific viral transcripts. Since the experimental systems for HCMV latency rely on primary immune cell populations, which are heterogeneous, and since infection is variable within the culture, it is likely that analyses of bulk populations could miss important signatures. Single cell-RNA-seq (scRNA-seq) provides a unique opportunity to depict viral and host heterogeneity simultaneously and thus to uncover functional connections between the cellular environment and viral gene expression. Indeed, several recent works that applied single cell transcriptomics, revealed novel insights into the complexity of the host response and cellular permissiveness for a number of well-studied viruses (Douam et al., 2017; Drayman et al., 2019; Galinato et al., 2018; Rato et al., 2017; Russell et al., 2018; Steuerman et al., 2018; Wyler et al., n.d.; Zanini et al., 2018).

Using single cell RNA data, we analyzed host determinants that are associated with HCMV latency. In CD14+ monocytes, we identified two cellular cell surface markers,

MHCII and its chaperon CD74, whose expression is inversely-correlated with viral transcript levels. We demonstrate these markers allow separating between cells harboring higher and lower viral transcript levels, that these differences are induced by HCMV infection and that the cells exhibiting higher viral transcript levels support more efficient reactivation of HCMV from latency. Using these markers, we show that latently infected cells display an intrinsic weaker immune response state that is important for the viral ability to express its genes and reactivate. Furthermore, analysis of 7500 infected HSPCs revealed very heterogeneous populations, but viral transcripts were only detected in cells expressing monocyte lineage markers, such as CD14. Remarkably, also in these HSPC-derived monocytes, higher viral transcript levels were associated with lower expression of CD74 and reduced immune response gene signature. Taken together, our findings highlight cell surface proteins associated with viral transcript levels and establish that both HSPC and monocyte infection models lead to establishment of HCMV latency in a similar anergic-like state of monocytic cells.

Results

Expression of host genes correlates with viral transcript levels

We have previously performed single cell RNA sequencing (scRNA-seq) on HCMV infected CD14+ monocytes as a model for studying HCMV latency (Shnayder et al., 2018). We used the Massively Parallel RNA Sequencing (MARS-seq) platform (Jaitin et al., 2014; Paul et al., 2015) to analyze experimentally infected primary CD14+ monocytes at different days post infection (dpi). This approach provided high coverage of cellular and viral transcriptomes, spanning expression of more than 11,000 genes in 3,655 cells. We previously used this dataset to examine the viral transcriptome during HCMV latency, showing it largely mirrors a late lytic viral program, albeit at much lower levels of expression (Shnayder et al., 2018).

An inherent advantage of scRNA-seq is the ability to track viral and host expression within the same cell, thus allowing analysis of viral-host interactions while keeping information on cell-to-cell and infection heterogeneity. Viral transcript levels in the majority of the infected monocyte population was low to undetectable and projection of the cells using t-distributed stochastic neighbor embedding (t-SNE) demonstrated that the cell distribution is determined mainly according to host gene expression variations (Shnayder et al., 2018). Nevertheless, the cells are organized according to viral transcript levels (Shnayder et al., 2018 and Figure 1A), suggesting that much of the differences in host gene expression are associated with variation in viral transcript levels. We therefore calculated Spearman correlation coefficient between the expression of each host gene and the total number of viral transcripts across all cells (Figure 1B). 319 cellular genes showed significant positive or negative correlations with viral gene expression (Z score>2, Table S1). Since a key challenge in transcriptome analysis is to connect between snapshots of gene expression profiles and a functional outcome, we first focused on identifying cell surface markers that exhibited strong association with viral gene expression in the scRNA-seq data as these may allow us to enrich for cells with higher viral transcript levels within the population. Among the strongest co-varying genes were genes encoding for MHC class II (variants HLA-DRB1/ HLA-DPB1/ HLA-DRA/ HLA-DQB1/ HLA-DPA1/ HLA-DQA1/ HLA-DMA/ HLA-DRB5/

HLA-DQA2 and HLA-DMB) and the transcript of CD74, an MHCII chaperon as well as a cell surface receptor on its own (Bergmann, 2012). Indeed, across the infected monocyte population, the expression level of MHC class II (MHCII, Figure 1C) and CD74 (Figure 1D) showed clear inverse-correlation to viral transcript levels (Figure 1A).

Cell-surface levels of CD74 and MHCII inversely-correlate with viral transcript levels

Viral gene expression levels in latent infection decreases with time in culture, likely due to continuous repression of the viral genome. This time related reduction in viral transcript levels is also apparent in our CD14+ monocytes scRNA-seq data (Figure 1E), thus the increase in CD74 and MHCII expression could be related to time in culture and to only indirectly inversely-correlate with viral transcript levels (Figure 1- figure supplement 1). Indeed, we observed that the expression of both CD74 and MHCII increase over time in culture, both in infected and in uninfected cells (Figure 1- figure supplement 2). We therefore tested whether this inverse-correlation between CD74 and MHCII expression and viral transcript levels is upheld within a single infected population at single time points. To this end, we infected primary CD14+ monocytes with HCMV strain TB40/E-GFP (O'Connor and Murphy, 2012; Christian Sinzger et al., 2008). At 3 days post infection (dpi) the cells were FACS-sorted according to the cell-surface levels of CD74 and MHCII (Figure 2A and Figure 2- figure supplement 1). RT-qPCR analysis of viral gene expression in these populations confirmed that both $CD74^{low}$ and $MHCII^{low}$ monocyte populations express higher levels of viral transcripts compared to their high expressing counterparts (Figure 2B and Figure 2- figure supplement 1). Importantly, the association of higher viral transcript levels with lower cell-surface levels of CD74 was also maintained at 6dpi (Figure 2- Figure supplement 2). We next examined if MHCII and CD74 are independent markers or are co-expressed and therefore mark the same population. mRNA expression analysis of CD74 and MHCII cells sorted according to the cell-surface levels of CD74 and MHCII as well as analysis of their co-expression in the scRNA-seq data, confirmed that the two markers are co-expressed and probably can be used to sort similar subpopulations within the infected cell population (Figure 2C and Figure 2- Figure supplement 3). We therefore

conducted all further experiments using CD74 based sorts. qPCR Analysis of viral DNA levels demonstrated that higher viral transcript levels in CD74^{low} monocytes are concurrent with higher abundance of viral genomes (Figure 2D) suggesting that differential loads of viral genome templates probably contribute to differential viral transcript levels and therefore to the effect on the host. Finally, we examined CD74 cell surface expression in infected and uninfected cells at different time points post infection but no major changes in CD74 distribution following infection were observed (Figure 2-figure supplement 4).

Changes in CD74 and MHCII expression are induced by infection

There are two alternative explanations for the inverse-correlation between viral transcript levels and CD74 cell-surface levels, several days post infection with HCMV. The first possibility is that viral entry is more efficient in CD74^{low} monocytes compared to CD74^{high} monocytes, leading to more incoming viral genomes and higher viral transcript levels. In this case, differences in viral levels between CD74^{high} and CD74^{low} monocytes should be evident immediately following viral entry to the cells. An alternative option is that the differential expression of CD74 is driven by HCMV infection. In this case, the viral DNA and RNA levels in early stages of infection should be independent of CD74 cell-surface levels, and at later time points, higher load of virus leads to the observed differences in CD74 expression. To test these possibilities, uninfected freshly isolated CD14+ monocytes were FACS sorted based on CD74 cell-surface levels and then infected separately with TB40E-GFP. At 8 and 72 hours post infection (hpi) viral DNA and RNA were analyzed by qPCR. We confirmed that indeed the CD74^{high} and CD74^{low} sorted cells exhibited differences in CD74 transcript levels negating the possibility that the separation is only due to variations associated with the cell surface staining (Figure 3A). No significant differences between viral DNA load (Figure 3B) or viral transcript levels (Figure 3C) in CD74^{high} and CD74^{low} monocytes were observed at either 8 or 72hpi, indicating there are no major differences in the efficiency of viral entry between the two populations. Taken together, these results

indicate that the observed variation in CD74 cell-surface levels is induced following HCMV infection.

CD74^{low} monocytes reactivate more efficiently

An important characteristic that defines latent infection is the ability of the virus to reactivate. Therefore, a key challenge in gene expression analysis is to connect between snapshots of gene expression profiles and the infection status of the cells, which is defined by this functional outcome. The identification of cellular cell surface markers that inversely-correlate with viral transcript levels provides a handle to connect between gene expression and the ability of the virus to reactivate. To test the association between viral transcript levels and reactivation efficiency, HCMV infected primary CD14+ monocytes were sorted by CD74 cell-surface levels, and viral reactivation was induced by two complementary methods; cytokine driven differentiation to dendritic cells (DCs) (Reeves et al., 2005), followed by incubation with an indicator fibroblasts monolayer, or long term co-culturing with fibroblasts. Quantification of GFP positive plaques showed that reactivation was significantly more frequent in CD74^{low} cells compared to CD74^{high} cells, in both protocols (Figure 4A and 4B). Lysed samples of monocytes plated onto indicator fibroblasts produced no plaques, confirming that these infected cells did not produce any detectable infectious virus, consistent with latency prior to reactivation induction. These results demonstrate a functional difference in reactivation efficiency associated with cell-surface CD74 levels as well as with viral transcript levels and genome load.

CD74 can be used to enrich for HCMV harboring CD14+ monocytes from viremic patients

Based on our experimental infection results, we next tested whether CD74 expression on CD14+ monocytes could be used as a cell surface marker that will allow enrichment of cells that contain HCMV genomes in healthy seropositive individuals. We used digital droplet PCR (ddPCR) to detect viral genome loads in CD14+ monocytes from

seven seropositive donors. Despite using a highly sensitive platform (Figure 5- figure supplement 1) and detection of several positive events, testing 100,000-400,000 cells per donor did not provide us with levels of detection and reproducibility that allowed relative quantification (Figure 5- figure supplement 2A, 2B and Table S2). Recently, it was suggested that latent HCMV resides in a unique B7H4-positive monocyte subset (Zhu et al., 2018). To examine whether we can enrich for a cell population in which we can detect robust levels of HCMV genomes, we collected CD14+ monocytes from 5 healthy seropositive individuals. In all five donors we could not detect a distinct population of B7H4-positive cells (Figure 5- figure supplement 2C). We nevertheless sorted the top 2% and the bottom 70% B7H4 stained cells and measured HCMV genome loads by ddPCR but did not detect higher levels of HCMV genomes in the top 2% B7H4 cells (Figure 5- figure supplement 2D). We therefore next analyzed CD14+ monocytes from seven hematopoietic stem cell transplant (HSCT) recipient samples in which HCMV viremia was detected, for the presence of viral genomes. In three of the tested samples no virus could be detected in CD14+ monocytes (Figure 5- figure supplement 2E). Monocytes from the four additional samples, in which viral genomes were detected (Figure 5- figure supplement 2E), were sorted according to CD74 cell-surface levels (Figure 5A). In agreement with our results in the experimental infection model, CD74 ^{low} monocytes were significantly enriched for the virus compared to CD74^{high} monocytes, illustrating that also in this natural infection setting, CD74 expression can be used to enrich for monocytes carrying viral genomes (Figure 5B and C).

Cells with higher viral load express lower immune-responsive gene signatures

To further characterize the differences in host pathways associated with the variations in viral transcript levels we conducted RNA-seq on latently infected monocytes, sorted according to their CD74 cell-surface levels. As was seen for single viral genes (Figure 2B), we found that on a genome-wide level, viral gene expression is higher in CD74^{low} cells compared to CD74^{high} cells (Figure 6A and Figure 6- figure supplement 1A). Furthermore, in accordance with our previous findings (Shnayder et al., 2018) the viral

gene expression profile in these cells was correlated with late lytic profile (Figure 6-figure supplement 1B), but viral transcript levels were very low (~ 0.1% of mRNA reads originated from the virus). We performed differential gene expression analysis comparing between the $CD74^{low}$ and $CD74^{high}$ populations. This analysis revealed 113 differentially expressed cellular genes (FDR<0.05, Table S3). Gene set enrichment analysis (GSEA) show that compared to $CD74^{low}$,$CD74^{high}$ monocytes, which exhibit lower viral transcript levels, are enriched for many immune response related pathways (Figure 6B and Table S4) including adaptive immune response and response to interferon gamma (Figure 6C). Since these cells were extracted from the same culture, these results suggest that monocytes carrying higher viral load are driven towards an intrinsic anergic-like phenotype. Reassuringly, many of the cellular genes and pathways exhibiting higher expression in $CD74^{high}$ cells were also inversely-correlated with viral transcript levels in the scRNA-seq data (Figure 6D, Figure 6- figure supplement 2 and Table S5). The $CD74^{low}$ monocyte population, which exhibits higher viral transcript levels, was enriched in genes that are related to transcription and virus life cycle (Figure 6B and Table S4). These pathways were not significantly correlated with viral expression levels in the single cell dataset, perhaps due to the sparser nature of our single cell measurements, which due to the sampling of limited number of cells may not detect some weaker effects.

To decipher the differentiation status of latent monocytes we compared our data to a recently published single cell analysis of lineage commitment during hematopoiesis (Velten et al., 2017). The genes that showed elevated expression in the $CD74^{high}$ cell population compared to the $CD74^{low}$ population were enriched for genes associated with monocyte lineage priming (Figure 6E), implying that the cells exhibiting higher viral transcript levels express less commitment markers and therefore may represent a less differentiated state. Significantly, this gene signature of monocyte lineage priming was also inversely-correlated with viral transcript levels in the scRNA-seq data (Figure 6F). To further investigate the cell differentiation state, we analyzed gene sets that were previously associated with M1 or M2 polarization (Gerrick et al., 2018) and found that genes that were more highly expressed in $CD74^{high}$ monocytes were significantly enriched for genes previously associated with M2 phenotype (Pval=2.25×10^{-8}). This

suggests that the infected CD14+ monocytes in our culture conditions may be polarized towards a cell-state associated with M2 phenotype, however higher viral transcript levels attenuates this differentiation trajectory.

Inhibition of interferon signaling increases viral gene expression and reactivation

Since a main feature of CD74^{low} monocytes is intrinsic reduced responsiveness to immune signals, we reasoned that this feature could contribute to the higher ability of the virus to express viral transcripts and eventually reactivate. To test this we examined whether inhibition of interferon signaling affects the expression level of viral transcripts and its ability to reactivate in infected CD14+ monocytes. We used ruxolitinib, a potent and selective Janus kinase (JAK) 1 and 2 inhibitor (Lin et al., 2009) that blocks the signaling downstream of interferon receptors. Treatment with ruxolitinib immediately after infection (3hpi) resulted in a considerable reduction in the levels of interferon-induced genes at 3 and 6dpi (Figure 7A) and in increased expression of essentially all viral transcripts (Figure 7B), in an overall uniform manner (Figure 7- figure supplement 1). In order to test how long after infection blocking interferon signaling still affects viral transcript levels, we applied ruxolitinib at different time points along infection and measured viral transcripts by RNA-seq. Interestingly, although the effect was smaller when the inhibitor was added later along infection, blocking interferon signaling still increased viral transcript levels even when added 3dpi (Figure 7C). These results therefore indicate there is continuous expression of viral genes and at 3dpi viral genes are still transcribed. However, the reduced effect of ruxolitinib with time also points that in addition there is gradual repression of viral gene expression as was also captured in our scRNA-seq analysis (Shnayder et al., 2018 and Figure 1E). Importantly, blocking interferon signaling led to more efficient reactivation of the virus (Figure 7D). This demonstrates that interferon signaling has a major role in repression of viral gene expression in infected monocytes and that the anergic-like state of monocytes exhibiting higher viral transcript levels is likely important for latency maintenance and reactivation.

Viral transcript levels in HCMV-infected CD34+ HSPCs are associated with priming towards the monocyte lineage and reduced immune-response

We have previously also performed scRNA-seq analysis of HCMV infected CD34+ HSPCs at 4dpi (Shnayder et al., 2018). To gain insight into the effects of HCMV on HSPC differentiation, we inferred differentiation trajectories using Monocle, a strategy that allows placing single cells along a pseudotime continuum based on their gene expression (Trapnell et al., 2014) (Figure 8A) . The cells expressing HCMV transcripts were clustered in a region that exhibited the latest pseudotime (i.e the most differentiated state, Figure 8B). Remarkably, although we infected bone marrow derived CD34+ HSPCs, viral transcripts were expressed only in cells expressing monocyte lineage markers (Velten et al., 2017), such as IRF7, IRF8 (Figure 8C) and CD14 (Figure 8D). Taking into account that it was previously shown that HCMV infects multipotent hematopoietic stem cells (Goodrum et al., 2004), these results indicate that HCMV induces differentiation of infected HSPCs towards the monocyte lineage. Interestingly, more detailed analysis of the distinct group of cells that expressed monocyte markers, revealed that this group could be split into two close yet distinct clusters, with viral reads being detected in much more cells in one of the clusters. Interestingly, the cluster showing higher viral transcript levels exhibited lower expression of CD74 (Pval=1.73×10^{-95}, Figure 8E and 8F) as well as gene signatures associated with lower immune response to interferon gamma and lower adaptive immune response (Pval<0.001 and Pval=0.003, respectively, Figure 8G and 8H). Thus, cells within the HSPC population that exhibit the highest viral gene expression belong to the CD14+ monocyte cell lineage and are associated with similar anergic-like signatures as infected monocytes.

Discussion

HCMV establishes latency in its host in progenitor cells of the myeloid system (Mendelson et al., 1996; Taylor-Wiedeman et al., 1991; Von Laer et al., 1995). Nevertheless, in cell culture experimental systems it is apparent that not all cells have the ability to reactivate, indicating that there are variations in the levels or dynamics of latent infection. This heterogeneity means that bulk assays, comparing infected and uninfected cell populations can capture host responses to HCMV infection but likely miss specific responses in the group of cells in which latency is established, and will enable reactivation down the line.

Recent single cell RNA-seq data portray low-level expression of a broad spectrum of canonical viral lytic genes during HCMV latent infection of various cells of the hematopoietic system (Galinato et al., 2018; Shnayder et al., 2018). Our analysis of latent HCMV infected CD14+ monocytes revealed a continuous population. Essentially, all of the cells were infected; however, they varied in the levels of viral transcripts. We exploited the host and viral heterogeneity, revealed simultaneously in scRNA-seq data, to look for associations between viral transcript levels and the human transcriptome. We found that among the genes that were highly inversely-correlated with viral transcript levels, is a group of genes encoding MHCII as well as the cell surface marker CD74, which serves as an MHCII chaperone. The finding that there are host genes that show specific correlation with viral transcript levels indicates that there is direct interaction between the host and the virus during latent infection- either there is a preference of the virus to infect specific cell types or that the virus actively affects the host transcriptome. Indeed, these markers could be used to enrich for cells with increased viral genome load and viral transcript levels. Importantly, this was supported by analysis of CD14+ monocytes from viremic patients, which also shows a clear inverse-correlation between CD74 cell-surface levels and viral genome load. These results are in line with previous findings showing that human, murine and rat cytomegalovirus down regulate the surface expression of MHCII molecules in cells of the myeloid lineage (Baca Jones et al., 2009; Elder et al., 2019; Lee et al., 2011; Slobedman et al., 2002; Yunis et al., 2018). In addition, CIITA, a transcription factor

that regulates CD74 and MHCII, was shown to be down regulated by HCMV (Lee et al., 2011).

The inverse-association between the expression of CD74 and MHCII and viral transcript levels could be related to differences in permissivity for the virus. By cell sorting according to CD74 levels prior to infection, we show that this is likely not the case; instead, these changes seem to be induced by viral infection. Moreover, we see that higher viral transcript levels in CD74^{low} cells is also accompanied by higher abundance of viral genomes, indicating that the level of viral transcripts is determined, at least initially, by the amount of incoming genomes, and this contributes to the extent of the effect on the host. Importantly, reactivation from HCMV latency happens in a very small population of cells even in experimental systems where the majority of cells are infected. We show here that CD74^{low} monocytes, which carry higher viral transcript and viral genome levels, reactivate more efficiently. This indicates that the ability of monocytes to reactivate is associated with viral transcript expression and that the cells carrying higher viral loads in these models are functionally the latent cell population, as they are more likely to reactivate. Recently, B7H4 was suggested as a marker for monocytes with higher levels of HCMV genomes from HCMV seropositive individuals (Zhu et al., 2018), however, we could not detect expression of B7H4 mRNA in any of our RNA-seq samples. Moreover, by staining for B7H4, we could not identify a distinct B7H4 positive monocyte population and did not detect higher levels of HCMV genomes in the top 2% B7H4 sorted cells. According to available human datasets, B7H4 is indeed not expressed in healthy monocytes ("Blood atlas - VTCN1 - The Human Protein Atlas"), however, these differences may stem from sampling of different donors or due to other variables related to the isolation of the cells.

Our analysis indicates that the cells that have higher expression of viral genes are less immune-responsive including reduced response to interferons. Since the CD74^{high} and CD74^{low} monocytes we analyzed grew in the same culture, they were exposed to the same immune-extrinsic signals. Hence, this difference is intrinsic to specific cells, and may be actively induced by the virus. This is supported by a recent report showing that interferon induced genes are downregulated during latent infection of CD14+ monocytes (Elder et al., 2019). This anergic-like state is functionally related to the

ability of the virus to express viral transcripts and to reactivate, as inhibiting the response to interferons resulted in increased expression of viral transcripts as well as increased reactivation efficiency. The importance of the role interferon plays in latent infection is further supported by the fact that CMV replication can be inhibited in otherwise permissive cells by treatment with interferon beta (Dağ et al., 2014). Similarly, extrinsic interferon stimuli inhibited HSV-1 reactivation from latency in neuronal cells (Linderman et al., 2017). Importantly, blocking interferon signaling increases viral gene expression even when done up to 3dpi, suggesting that there is continuous viral gene expression. However, the effect was smaller when the inhibitor was added later along infection, indicating a gradual repression of viral gene expression with time. Overall, these findings indicate that a major aspect of the maintenance of latency and of the ability to reactivate at the cellular level is a balance between opposing forces which also affect each other- the intrinsic immune response, specifically the interferon pathway, and viral transcript levels. Future work will have to delineate the mechanisms by which these immune response pathways control general viral transcription and how viral transcripts and viral proteins modulate the immune response during latent infection.

Previous works have described changes in the differentiation state of monocytes in response to infection or to specific viral genes (Avdic et al., 2013; Chan et al., 2008; Smith et al., 2004). The use of the CD74 marker, allowed us to focus on differentiation processes unique to the CD14+ monocytes with higher viral transcript levels and higher reactivation efficiencies. We show that higher viral transcript levels are associated with less expression of monocyte priming signature and lower expression of a gene signature associated with M2 phenotype. It was previously shown that the HCMV encoded IL-10 homolog polarizes monocytes towards an M2c macrophage phenotype (Avdic et al., 2013). It is possible that the total population is indeed polarized towards this direction; however, this process may be inhibited in cells with higher viral transcript levels. These findings may suggest that the virus promotes attenuation of differentiation processes that monocytes are undergoing in culture and in the blood (Patel et al., 2017).

Examining the differentiation state in HSPCs following infection is far more complicated than in monocytes, as CD34+ HSPCs are a mix of pluripotent cells as well

as progenitor cells in different stages of lineage commitment (Velten et al., 2017). By applying single cell trajectory analysis, which allows recovery of gene expression kinetics of differentiating cells (Trapnell et al., 2014), and aligning our data with data of lineage commitment during hematopoiesis (Velten et al., 2017), we show that cells containing detectable viral transcripts belong largely to one specific population- cells primed towards the monocyte lineage. From our data, we cannot determine whether this is due to preferential infection by HCMV of monocyte lineage committed cells within the HSPC compartment or whether HCMV infects multipotent cells, which are then either preferentially skewed towards the monocytic lineage or viral gene expression is initiated only when cells start to differentiate in the monocytes lineage. However since it was previously demonstrated that multipotent cells are infected with HCMV (Goodrum et al., 2004), the latter options are more likely. Previous studies demonstrated that HCMV infection leads to an increase in monocyte markers (Zhu et al., 2018) and similar results were shown specifically for the viral gene UL7 (Crawford et al., 2018). Our results are in line with these studies and show directly and in an unbiased manner that several days post infection, viral gene expression can be found only in cells in the monocyte lineage. Moreover, within the monocyte lineage primed cells, the cells exhibiting higher viral transcript levels show the same markers and characteristics as we found when infecting CD14+ monocytes, primarily lower expression of CD74 and weaker immune responsiveness. This is especially interesting as it suggests that regardless of the developmental stage in which HCMV infects, the end point, a few days after infection, are cells with very similar characteristics.

Overall, we use here single cell data to pinpoint the characteristics of the latently infected cells in an unbiased manner. Our analyses indicate that HCMV drives human HSPCs and monocytes into a monocyte state characterized by anergic-like gene signature. These findings shed light on the characteristics of the latent reservoir, which may help in the effort of developing strategies to eradicate the latently infected cells.

Acknowledgments

We thank the members of the Stern-Ginossar lab for critical reading of the manuscript. We thank Eain A. Murphy for the TB40E-GFP virus strain. This research was supported by Infect-ERA (TANKACY) and the European Research Council starting grant (StG-2014-638142) to NS-G, and by the Cambridge NIHR BRC Cell Phenotyping Hub and a grant from the British Medical Research Council (Grant G0701279) to JS. NS-G is incumbent of the skirball career development chair in new scientist. B.S. and E.B. were supported by a Biomed-Connect Grant from the University of Sydney.

Materials and Methods

Cells and virus stocks

Primary CD14+ monocytes were isolated from fresh venous blood, obtained from healthy donors, using Lymphoprep (Stemcell Technologies) density gradient followed by magnetic cell sorting with CD14+ magnetic beads (Miltenyi Biotec). The cells were cultured in X-Vivo15 media (Lonza) supplemented with 2.25mM L-glutamine at 37^0C in 5% CO2 (Fortunato, 2014). Primary human foreskin fibroblasts (HFF) (ATCC CRL-1634) were maintained in DMEM with 10% fetal bovine serum (FBS), 2mM L-glutamine, and 100 units/ml penicillin and streptomycin (Beit-Haemek, Israel).

The TB40E virus containing an SV40-GFP tag (TB40E-GFP) was described previously (O'Connor and Murphy, 2012; Christian Sinzger et al., 2008). Virus was propagated by electroporation of infectious bacterial artificial chromosome (BAC) DNA into fibroblasts using the Amaxa P2 4D-Nucleofector kit (Lonza) according to the manufacturer's instructions. Viral stocks were concentrated by centrifugation at 26000xg, 4^0C for 120 minutes. Infectious virus yields were assayed on THP-1 cells (ATCC TIB-202).

Infection and reactivation procedures

For experimental latent infection, CD14+ monocytes were incubated with the virus for 3 hours, washed twice and supplemented with fresh media. To assess infection efficiency, a sample of the infected cell population was FACS analyzed for GFP expression at 3dpi. All experiments were conducted when there was a shift in GFP intensity of the entire population following infection, indicating all cells were infected. HCMV latency was validated by absence of GFP positive plaques on fibroblasts incubated with infected monocytes cell lysate.

For reactivation assays, infected monocytes were counted, plated and co-cultured with primary fibroblasts immediately or following differentiation into dendritic cells (DCs). DC differentiation was done by incubation of cells with granulocyte-macrophage CSF and interleukin-4 (Peprotech) at 1,000 U/ml for 5 days, followed by stimulation with 500 ng/ml of LPS (Sigma) for 48 hours (as previously described in (Cobbs et al., 2014)). Release of infectious virions was assayed by quantification of GFP positive plaques on the fibroblasts monolayer.

Ruxolitinib treatment

Ruxolitinib (Ruxo) was added at a concentration of 4uM, either immediately after infection (3hpi) or at later time points (1, 2, or 3dpi). Monocytes were washed to remove residual Ruxo before co-culturing with fibroblasts.

Quantitative real-time PCR analysis

For analysis of RNA expression, total RNA was extracted using Tri-Reagent (Sigma) according to manufacturer's protocol. cDNA was prepared using qScript cDNA Synthesis Kit (Quanta Biosciences) according to manufacturer's protocol. For analysis of DNA levels, cells were lysed in a 1:1 mixture of PCR solutions A (100 mM KCl, 10 mM Tris–HCl pH 8.3, and 2.5 mM MgCl2) and B (10 mM Tris–HCl pH 8.3, 2.5 mM MgCl2, 1% Tween 20, 1% Non-idet P-40, and 0.4 mg/ml Proteinase K), for 60 min at 60°C followed by a 10 min 95°C incubation, as described in (Roback et al., 2001). Real time PCR was performed using the SYBR Green PCR master-mix (ABI) on the QuantStudio 12K Flex (ABI) with the following primers (forward, reverse):

UL 138 (GTGTCTTCCCAGTGCAGCTA, GCACGCTGTTTCTCTGGTTA)

UL22 (TTACTAGCCGTGACCTTGACG, CAGAAATCGAAGCGCAGCG)

RNA 2.7 (TCCTACCTACCACGAATCGC, GTTGGGAATCGTCGACTTTG)

CD74 (TGGAAGGTCTTTGAGAGCTGGATG, TTCCTGGCACTTGGTCAGTA)

MHCII-HLA-DQA1 (CTTCATCATCCAAGGCCTGC,
CGGGCCAGAGAATAGTGCTA)

ANXA5 (AGTCTGGTCCTGCTTCACCT, CAAGCCTTTCATAGCCTTCC)

Viral DNA was quantified with RNA2.7 primers, host DNA was measured with the following primers (forward, reverse):

B2M (TGCTGTCTCCATGTTTGATGTATCT, TCTCTGCTCCCCACCTCTAAGT)

Cell staining for flow cytometry and sorting

Cells were counted, and stained in cold MACS buffer (PBS, 5% BSA, 2 mM EDTA). Cell staining was done using the following antibodies: anti-human APC-CD74 (Clone: 5-329, Miltenyi Biotec), anti-human PE-HLA-DR, DP, DQ (clone: REA332, Miltenyi Biotec),

anti-human FITC-CD14 (Clone: TÜK4, Miltenyi Biotec), anti-human APC-B7H4 (clone: MIH43, Biolegend), APC-Mouse IgG1 isotype control (clone: MOPC-21, Biolegend) according to manufacturer's instructions. Cells were analyzed and sorted on a BD FACSAriaIII.

Detection of viral genomes by digital PCR

Detection of viral DNA in monocytes from natural latent samples was done using the QX200 droplet digital PCR system (Bio-Rad), using FAM labeled HCMV primer and probe (Human CMV HHV5 kit for qPCR using a glycoprotein B target (PrimerDesign) and HEX labeled RPP30 copy number assay for ddPCR (Bio-Rad), as previously described (Jackson et al., 2017). Calibration curve was ran in duplicate, using CMV positive control template (PrimerDesign). The limit of detection was at 3 events per sample, with accuracy improved at 10 copies and higher (Figure 5- figure supplement 1). For sample preparation cells were counted, dry pelleted, and stored at −80°C prior to DNA extraction. DNA was extracted from the cell pellet in a 1:1 mixture of PCR solutions A (100 mM KCl, 10 mM Tris–HCl pH 8.3, and 2.5 mM MgCl2) and B (10 mM Tris–HCl pH 8.3, 2.5 mM MgCl2, 0.25% Tween 20, 0.25% Non-idet P-40, and 0.4 mg/ml Proteinase K), for 60 min at 60°C followed by a 10 min 95°C incubation, according to the description in (Roback et al., 2001).

RNA library construction

RNA libraries were generated from samples of ~10,000 cells according to the MARS-seq protocol (Jaitin et al., 2014; Keren-Shaul et al., 2019).

Sequencing and data analysis

RNA-Seq libraries (pooled at equimolar concentration) were performed in duplicates and sequenced using NextSeq 500 (Illumina), with read parameters: Read1: 72 cycles and Read2: 15 cycles.

Analysis of bulk MARS-seq of CD14+ monocytes, sorted according to the CD74 cell-surface levels, was done as described previously (Shnayder et al., 2018). The number of Unique Molecular Identifiers (UMIs) were: 976,294, and 947,474 for the $CD74^{high}$ samples, 902,150, and 844,872 for the $CD74^{low}$ samples (Figure 6 A-E); 2,422,356 and

2,395,329 for the –Ruxo 3dpi samples, 2,187,290 and 925,362 for the +Ruxo 3dpi samples, 4,228,168 and 3,854,339 for the –Ruxo 6dpi samples and 3,449,188 and 1,900,301 for the +Ruxo 6dpi samples (Figure 7A and 7B); 2,076,741 and 1,858,871 for the +Ruxo at 1dpi samples, 1,914,289 and 1,163,469 for the +Ruxo at 2dpi samples, 1,554,298 and 1,685,834 for the +Ruxo at 3dpi samples, 2,004,859 and 2,450,039 for the –Ruxo control (Figure 7C). Reads for gene expression and correlation analyses were normalized using DEseq2.

Based on the t-SNE plot of the CD14+ monocyte cells in (Shnayder et al., 2018), the latent cells were defined as all cells besides the small distinct group of cells that show high levels of viral reads (mean 9.5%) . All the analyses in this paper include only the remaining 3,416 cells.

Correlation and t-SNE coloring

For calculating the Spearman correlation between host genes and either viral expression levels or dpi, in each cell, the sum of viral reads was normalized to the total number of reads in the cell, and the number of reads of each host gene was normalized to the total number of host reads in the cell. The correlation was calculated across 1,448 cells for 6,997 genes. Cells with no viral reads, or with less than 700 different host genes expressed, were omitted. Genes with total number of reads less than 20 were ignored. The Z-score was calculated based on the mean and the standard deviation calculated over all 6,997 genes.

Color coding of t-SNE (Figure 1C, Figure 6D and Figure 7B) or Monocle (e.g. Figure 7C, G, H) plots for expression levels of groups of genes was calculated according to the average relative expression as follows: First, genes with low level of expression, i.e. expressed in less than 50 cells, were omitted. Next, the number of host reads per cell was normalized so that all cells will include the same number of reads. Next, for each cell and each gene, the relative expression is calculated as $log_2(\frac{r_{ij}+1}{\overline{r_i}+1})$ where r_{ij} is the number of reads of gene i in cell j and $\overline{r_i}$ is the average number of reads of gene i over all cells. Finally, for each cell the average log_2 relative expression over all genes in the group was calculated.

Differential Expression and Enrichment analysis

The differential expression analysis was done with DESeq2 (version 1.22.2) (Love et al., 2014) using default parameters, with the number of reads in each of the samples as an input.

The normalized number of reads according to DESeq2 were used for enrichment analysis using GSEA (version 3.0) (Subramanian et al., 2005). The genes sets that were used were GO biological process (c5.bp) and REACTOME (c2.cp.reactome) from MSigDB (version 6.2) (Liberzon et al., 2011) and the monocyte progenitor gene list from (Velten et al., 2017). The GSEA plots were created based on the GSEA output with the R package enrichplot (Yu, n.d.). To calculate differential expression and pathway enrichment within the distinct group of cells that expressed monocyte priming markers, we used the graph-based clustering of the Cell Ranger software ("10x CellRanger 2.0.0," n.d.), stratifying this group of cells into two distinct clusters and GSEA (version 3.0) (Subramanian et al., 2005) for pathway enrichment. Differential expression and enrichment analysis were done on these two clusters. Enrichment of M1 and M2 associated genes was done by hypergeometric test taking all genes with mean expression >=5 as background.

Monocle analysis

CD34+ cells were ordered according to the predicted pseudo-time using Monocle (version 2.10.1) (Trapnell et al., 2014). The cells used in this analysis were filtered to have at least 1,000 expressed genes and not more than 10,000 UMIs, while the genes were filtered to have mean expression greater than 0.1, and empirical dispersion greater than the global dispersion fit.

Ethics statement

All fresh peripheral blood samples were obtained after approval of protocols by the Weizmann Institutional Review Board (IRB application 92-1). The study using HSCT recipient samples was approved by the Human Research Ethics Committee of the University of Sydney and the Western Sydney Local Health District. Informed consent was obtained from all study participants prior to enrolment in accordance with the Declaration of Helsinki.

Data availability

All next-generation sequencing data files were deposited in Gene Expression Omnibus under accession number GSE138838.

References

10x CellRanger 2.0.0. n.d. https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome

Avdic S, Cao JZ, McSharry BP, Clancy LE, Brown R, Steain M, Gottlieb DJ, Abendroth A, Slobedman B. 2013. Human cytomegalovirus interleukin-10 polarizes monocytes toward a deactivated M2c phenotype to repress host immune responses. *J Virol* **87**:10273–82. doi:10.1128/JVI.00912-13

Baca Jones CC, Kreklywich CN, Messaoudi I, Vomaske J, McCartney E, Orloff SL, Nelson JA, Streblow DN. 2009. Rat cytomegalovirus infection depletes MHC II in bone marrow derived dendritic cells. *Virology* **388**:78–90. doi:10.1016/J.VIROL.2009.02.050

Bergmann H. 2012. Business as usual: The p35 isoform of human CD74 retains function in antigen presentation. *Immunol Cell Biol* **90**:839–840. doi:10.1038/icb.2012.51

Blood atlas - VTCN1 - The Human Protein Atlas. n.d. https://www.proteinatlas.org/ENSG00000134258-VTCN1/blood

Chan G, Bivins-Smith ER, Smith MS, Smith PM, Yurochko AD. 2008. Transcriptome analysis reveals human cytomegalovirus reprograms monocyte differentiation toward an M1 macrophage. *J Immunol* **181**:698–711. doi:10.4049/jimmunol.181.1.698

Chan G, Nogalski MT, Bentz GL, Smith MS, Parmater A, Yurochko AD. 2010. PI3K-Dependent Upregulation of Mcl-1 by Human Cytomegalovirus Is Mediated by Epidermal Growth Factor Receptor and Inhibits Apoptosis in Short-Lived Monocytes. *J Immunol* **184**:3213–3222. doi:10.4049/jimmunol.0903025

Cheng S, Caviness K, Buehler J, Smithey M, Nikolich-Žugich J, Goodrum F. 2017. Transcriptome-wide characterization of human cytomegalovirus in natural infection and experimental latency. *Proc Natl Acad Sci*. doi:10.1073/pnas.1710522114

Cobbs CS, Matlaf L, Harkins LE. 2014. Human Cytomegaloviruses **1119**:165–196. doi:10.1007/978-1-62703-788-4

Crawford LB, Kim JH, Collins-Mcmillen D, Lee B-J, Landais I, Held C, Nelson JA, Yurochko AD, Caposio P. 2018. Human Cytomegalovirus Encodes a Novel FLT3 Receptor Ligand Necessary for Hematopoietic Cell Differentiation and Viral Reactivation. doi:10.1128/mBio

Crough T, Khanna R. 2009. Immunobiology of Human Cytomegalovirus: from Bench to Bedside. *Clin Microbiol Rev* **22**:76–98. doi:10.1128/CMR.00034-08

Dağ F, Dölken L, Holzki 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. doi:10.1371/journal.ppat.1003962

Douam F, Hrebikova G, Albrecht YES, Sellau J, Sharon Y, Ding Q, Ploss A. 2017. Single-cell tracking of flavivirus RNA uncovers species-specific interactions with the immune system dictating disease outcome. *Nat Commun* **8**:14781. doi:10.1038/ncomms14781

Drayman N, Patel P, Vistain L, Tay S. 2019. HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. *Elife* **8**:566489. doi:10.7554/eLife.46339

Elder EG, Krishna BA, Williamson J, Lim EY, Poole E, Sedikides GX, Wills M, O'connor CM, Lehner PJ, Sinclair J. 2019. Interferon-Responsive Genes Are Targeted during the Establishment of Human Cytomegalovirus Latency.

Fortunato E. 2014. Human Cytomegaloviruses. *Hum Cytomegaloviruses Methods Protoc* **1119**:47–58. doi:10.1007/978-1-62703-788-4

Galinato M, Shimoda K, Aguiar A, Hennig F, Boffelli D, McVoy MA, Hertel L. 2018. Single-cell transcriptome analysis of CD34+ stem cell-derived myeloid cells identifies a CFU-GEMM-like population permissive to human cytomegalovirus infection. *bioRxiv* 438457. doi:10.1101/438457

Gerrick KY, Gerrick ER, Gupta A, Wheelan SJ, Yegnasubramanian S, Jaffee EM. 2018. Transcriptional profiling identifies novel regulators of macrophage polarization.

PLoS One **13**. doi:10.1371/journal.pone.0208602

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. doi:10.1182/blood-2003-12-4344

Jackson SE, Sedikides GX, Okecha G, Poole EL, Sinclair JH, Wills MR. 2017. Latent cytomegalovirus (CMV) infection does not detrimentally alter T cell responses in the healthy old, but increased latent CMV carriage is related to expanded CMV-specific T cells. *Front Immunol* **8**:733. doi:10.3389/fimmu.2017.00733

Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I. 2014. Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science (80-)* **343**:776–779. doi:10.1126/science.1247651

Keren-Shaul H, Kenigsberg E, Jaitin DA, David E, Paul F, Tanay A, Amit I. 2019. MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. *Nat Protoc* **14**:1841–1862. doi:10.1038/s41596-019-0164-4

Kew V, Wills M, Reeves M. 2017. HCMV activation of ERK-MAPK drives a multi-factorial response promoting the survival of infected myeloid progenitors. *J Mol Biochem* **6**:13–25.

Lee AW, Wang N, Hornell TMC, Harding JJ, Deshpande C, Hertel L, Lacaille V, Pashine A, Macaubas C, Mocarski ES, Mellins ED. 2011. Human cytomegalovirus decreases constitutive transcription of MHC class II genes in mature Langerhans cells by reducing CIITA transcript levels. *Mol Immunol* **48**:1160–1167. doi:10.1016/J.MOLIMM.2011.02.010

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**:1739–1740. doi:10.1093/bioinformatics/btr260

Lin Q, Meloni D, Pan Y, Xia M, Rodgers J, Shepard S, Li M, Galya L, Metcalf B, Yue TN, Liu P, Zhou J. 2009. Enantioselective synthesis of janus kinase inhibitor

INCB018424 via an organocatalytic aza-michael reaction. *Org Lett* **11**:1999–2002. doi:10.1021/ol900350k

Linderman JA, Kobayashi M, Rayannavar V, Fak JJ, Darnell RB, Chao M V, Wilson AC, Mohr I. 2017. Immune Escape via a Transient Gene Expression Program Enables Productive Replication of a Latent Pathogen. *Cell Rep* **18**:1312–1323. doi:10.1016/j.celrep.2017.01.017

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

Mendelson M, Monard S, Sissons P, Sinclair J. 1996. Detection of endogenous human cytomegalovirus in CD34+ bone marrow progenitors. *J Gen Virol* **77**:3099–3102. doi:10.1099/0022-1317-77-12-3099

Mocarski, E.S., Shenk, T. et al. 2013. CytomegalovirusesFields' Virology 6th Edition. Lippincott Williams & Wilkins. pp. 1960–2014.

O'Connor CM, Murphy EA. 2012. A Myeloid Progenitor Cell Line Capable of Supporting Human Cytomegalovirus Latency and Reactivation, Resulting in Infectious Progeny. *J Virol* **86**:9854–9865. doi:10.1128/JVI.01278-12

Parry HM, Zuo J, Frumento G, Mirajkar N, Inman C, Edwards E, Griffiths M, Pratt G, Moss P. 2016. Cytomegalovirus viral load within blood increases markedly in healthy people over the age of 70 years. *Immun Ageing* **13**:1. doi:10.1186/s12979-015-0056-6

Patel AA, Zhang Y, Fullerton JN, Boelen L, Rongvaux A, Maini AA, Bigley V, Flavell RA, Gilroy DW, Asquith B, Macallan D, Yona S. 2017. The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J Exp Med* **214**:1913–1923. doi:10.1084/JEM.20170355

Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, Winter D, Lara-Astiaso D, Gury M, Weiner A, David E, Cohen N, Lauridsen FKB, Haas S, Schlitzer A, Mildner A, Ginhoux F, Jung S, Trumpp A, Porse BT, Tanay A, Amit I. 2015. Transcriptional Heterogeneity and Lineage Commitment in Myeloid

Progenitors. *Cell* **163**:1663–1677. doi:10.1016/j.cell.2015.11.013

Rato S, Rausell A, Muñoz M, Telenti A, Ciuffi A. 2017. Single-cell analysis identifies cellular markers of the HIV permissive cell. *PLOS Pathog* **13**:e1006678. doi:10.1371/journal.ppat.1006678

Reeves MB, Breidenstein A, Compton T. 2012. Human cytomegalovirus activation of ERK and myeloid cell leukemia-1 protein correlates with survival of latently infected cells. *Proc Natl Acad Sci U S A* **109**:588–593. doi:10.1073/pnas.1114966108

Reeves MB, Lehner PJ, Sissons JGP, Sinclair JH. 2005. An in vitro model for the regulation of human cytomegalovirus latency and reactivation in dendritic cells by chromatin remodelling. *J Gen Virol* **86**:2949–2954. doi:10.1099/vir.0.81161-0

Roback JD, Hillyer CD, Lawrence Drew W, Laycock ME, Luka J, Mocarski ES, Slobedman B, Smith JW, Soderberg-Naucler C, Todd DS, Woxenius S, Busch MP. 2001. Multicenter evaluation of PCR methods for detecting CMV DNA in blood donors. *Transfusion* **41**:1249–1257. doi:10.1046/j.1537-2995.2001.41101249.x

Russell AB, Trapnell C, Bloom JD. 2018. Extreme heterogeneity of influenza virus infection in single cells. *Elife* **7**. doi:10.7554/eLife.32303

Schwartz M, Stern-Ginossar N. 2019. The Transcriptome of Latent Human Cytomegalovirus. *J Virol* **93**:e00047-19. doi:10.1128/JVI.00047-19

Shnayder 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. doi:10.1128/mBio.00013-18

Sinzger C., Digel M, Jahn G. 2008. Cytomegalovirus cell tropism. *Curr Top Microbiol Immunol* **325**:63–83. doi:10.1007/978-3-540-77349-8_4

Sinzger Christian, Hahn G, Digel M, Katona R, Sampaio KL, Messerle M, Hengel H, Koszinowski U, Brune W, Adler B. 2008. Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *J Gen Virol* **89**:359–368. doi:10.1099/vir.0.83286-0

Slobedman B, Cao JZ, Avdic S, Webster B, McAllery S, Cheung AK, Tan JC, Abendroth A. 2010. Human cytomegalovirus latent infection and associated viral gene expression. *Future Microbiol* **5**:883–900. doi:10.2217/fmb.10.58

Slobedman B, Mocarski ES. 1999. Quantitative Analysis of Latent Human Cytomegalovirus. *J Virol* **73**:4806–4812.

Slobedman B, Mocarski ES, Arvin AM, Mellins ED, Abendroth A. 2002. Latent cytomegalovirus down-regulates major histocompatibility complex class II expression on myeloid progenitors, Blood. doi:10.1182/blood.V100.8.2867

Smith MS, Bentz GL, Alexander JS, Yurochko AD. 2004. Human Cytomegalovirus Induces Monocyte Differentiation and Migration as a Strategy for Dissemination and Persistence. *J Virol* **78**:4444–4453. doi:10.1128/jvi.78.9.4444-4453.2004

Smith MS, Bivins-Smith ER, Tilley AM, Bentz GL, Chan G, Minard J, Yurochko AD. 2007. Roles of Phosphatidylinositol 3-Kinase and NF- B in Human Cytomegalovirus-Mediated Monocyte Diapedesis and Adhesion: Strategy for Viral Persistence. *J Virol* **81**:7683–7694. doi:10.1128/jvi.02839-06

Staras SAS, Dollard SC, Radford KW, Flanders WD, Pass RF, Cannon MJ. 2006. Seroprevalence of Cytomegalovirus Infection in the United States, 1988-1994. *Clin Infect Dis* **43**:1143–1151. doi:10.1086/508173

Steuerman Y, Cohen M, Peshes-Yaloz N, Valadarsky L, Cohn O, David E, Frishberg A, Mayo L, Bacharach E, Amit I, Gat-Viks I. 2018. Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing. *Cell Syst* **6**:679-691.e4. doi:10.1016/j.cels.2018.05.008

Stevenson E V., Collins-McMillen D, Kim JH, Cieply SJ, Bentz GL, Yurochko AD. 2014. HCMV reprogramming of infected monocyte survival and differentiation: A goldilocks phenomenon. *Viruses* **6**:782–807. doi:10.3390/v6020782

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102**:15545–15550. doi:10.1073/PNAS.0506580102

Taylor-Wiedeman J, Sissons JGP, Borysiewicz LK, Sinclair JH. 1991. Monocytes are a major site of persistence of human cytomegalovirus in peripheral blood mononuclear cells. *J Gen Virol* **72**:2059–2064. doi:10.1099/0022-1317-72-9-2059

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**:381–386. doi:10.1038/nbt.2859

Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D, Boch T, Hofmann W-K, Ho AD, Huber W, Trumpp A, Essers MAG, Steinmetz LM. 2017. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* **19**:271–281. doi:10.1038/ncb3493

Von Laer D, Meyer-Koenig U, Serr A, Finke J, Kanz L, Fauser AA, Neumann- Haefelin D, Brugger W, Hufert FT. 1995. Detection of cytomegalovirus DNA in CD34+cells from blood and bone marrow. *Blood* **86**:4086–4090.

Wyler E, Franke V, Menegatti J, Kocks C, Boltengagen A, Praktiknjo S, Walch-Rückheim B, Bosse J, Rajewsky N, Grässer F, Akalin A, Landthaler M. n.d. Single-cell RNA-sequencing of herpes simplex virus 1-infected cells connects NRF2 activation to an antiviral program. doi:10.1038/s41467-019-12894-z

Yu G. n.d. enrichplot: Visualization of Functional Enrichment Result. https://github.com/GuangchuangYu/enrichplot

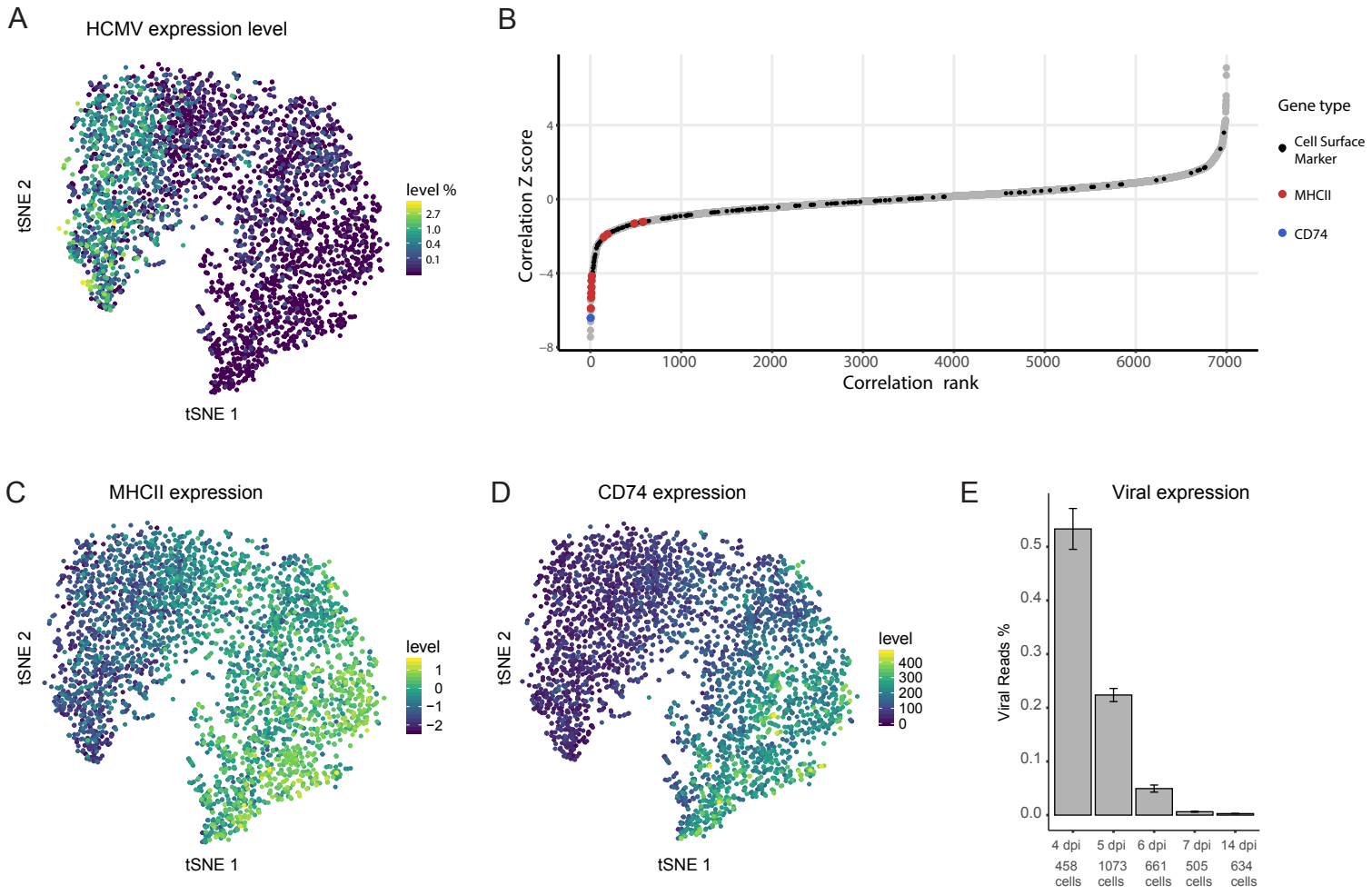
Yunis J, Farrell HE, Bruce K, Lawler C, Sidenius S, Wyer O, Davis-Poynter N, Stevenson PG. 2018. Murine cytomegalovirus degrades MHC class II to colonize the salivary glands. *PLoS Pathog* **14**:e1006905. doi:10.1371/journal.ppat.1006905

Zanini F, Pu SY, Bekerman E, Einav S, Quake SR. 2018. Single-cell transcriptional dynamics of flavivirus infection. *Elife* **7**:e32942. doi:10.7554/eLife.32942

Zhu D, Pan C, Sheng J, Liang H, Bian Z, Liu Y, Trang P, Wu J, Liu F, Zhang C-Y, Zen K. 2018. Human cytomegalovirus reprogrammes haematopoietic progenitor cells into immunosuppressive monocytes to achieve latency. *Nat Microbiol* **3**:503–513. doi:10.1038/s41564-018-0131-9

Figures

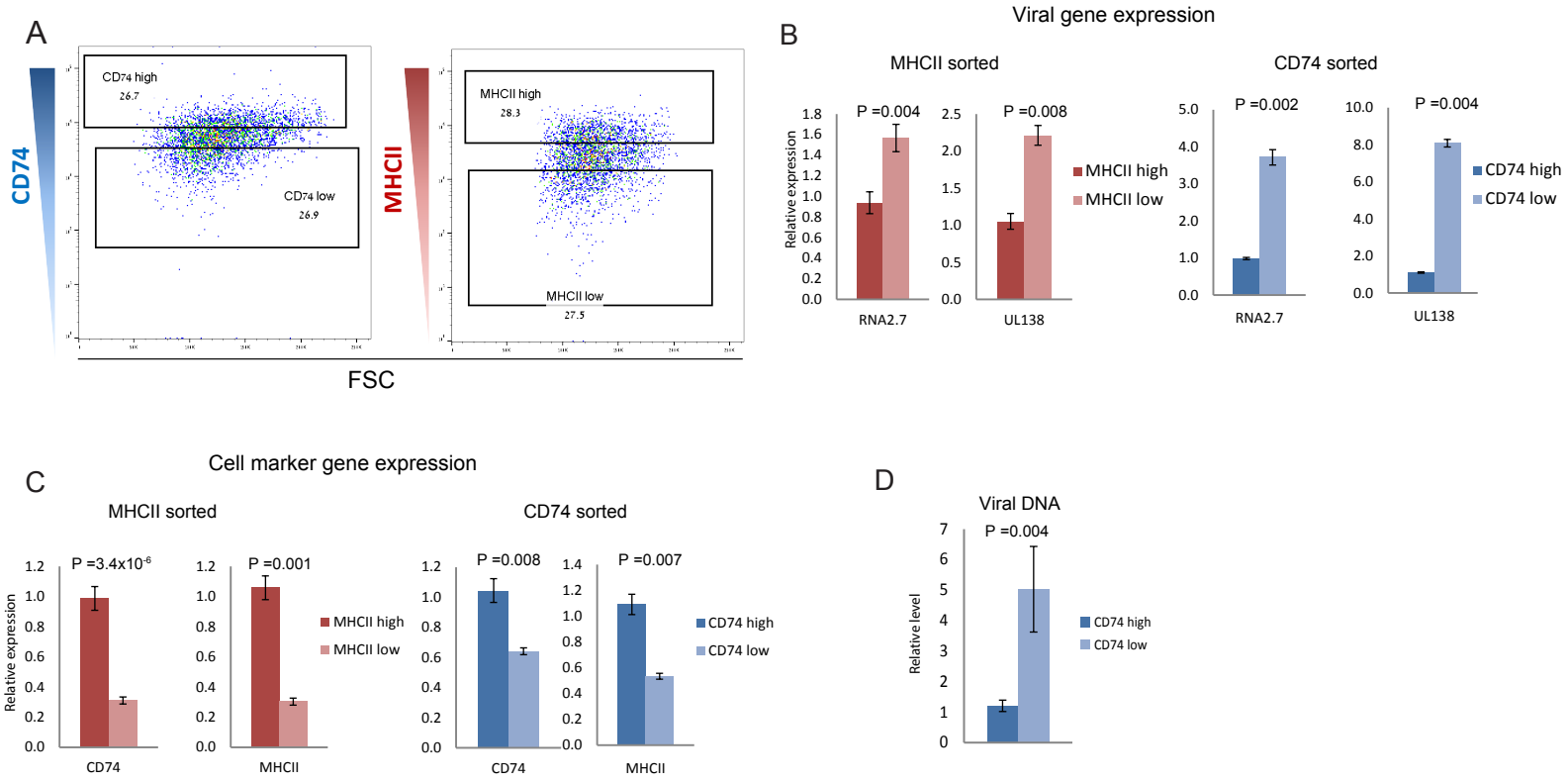
Figure 1



Cellular gene expression co-vary with viral transcript levels

(A) t-SNE plot of 3,416 latently infected CD14+ monocytes based on host and viral gene expression (Shnayder et al., 2018), colored by percentage of HCMV reads per cell. (B) Distribution of Z scores of the Spearman correlation coefficients between host transcript levels and total HCMV transcript levels across a single population of infected CD14+ monocytes. Black dots mark cells surface markers coding transcripts, blue marks CD74 and red marks transcripts of MHCII (isoforms HLA-RB1/ HLA-DPB1/ HLA-DRA/ HLA-DQB1/ HLA-DPA1/ HLA-DQA1/ HLA-DMA/ HLA-DRB5/ HLA-DQA2 and HLA-DMB, left to right). (C and D) t-SNE plots of monocytes as presented in A colored coded by their expression levels of MHCII (HLA-RB1/ HLA-DPB1/ HLA-DRA/ HLA-DQB1/ HLA-DPA1/ HLA-DQA1/ HLA-DMA/ HLA-DRB5/ HLA-DQA2 and HLA-DMB) (C) or CD74 (D) transcripts. (E) Percentage of viral reads measured in all single cells by days post infection (dpi). Error bars represent standard deviation across the single cells.

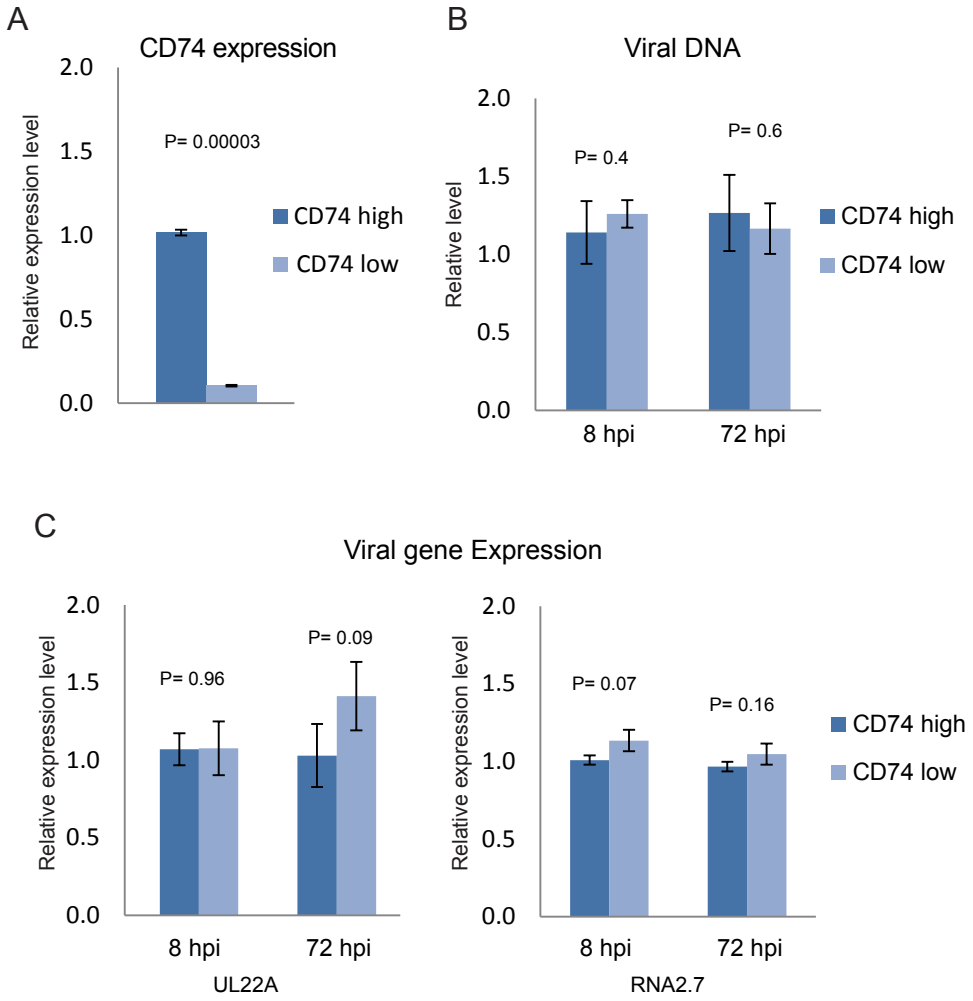
Figure 2



CD74 and MHCII cell-surface levels in HCMV infected monocytes inversely-correlate with viral transcript levels

(A) HCMV infected monocytes were FACS sorted according to cell-surface levels of CD74 or MHCII, at 3dpi. High and low gates were determined as highest and lowest 30% of the population, respectively. (B and C) Relative expression level of the viral transcripts RNA2.7 and UL138 (B) or MHCII and CD74 transcripts (C), as measured by RT-qPCR in HCMV infected cells, sorted by either MHCII (left) or CD74 (right) cell-surface levels at 3dpi. (D) Relative abundance of viral DNA in HCMV infected CD74high and CD74low monocytes, a t 3dpi as measured by qPCR. Graphs show a representative experiment of 3 biological repeats, error bars reflect standard deviation of 3 measurements. P values as calculated by t-test are indicated.

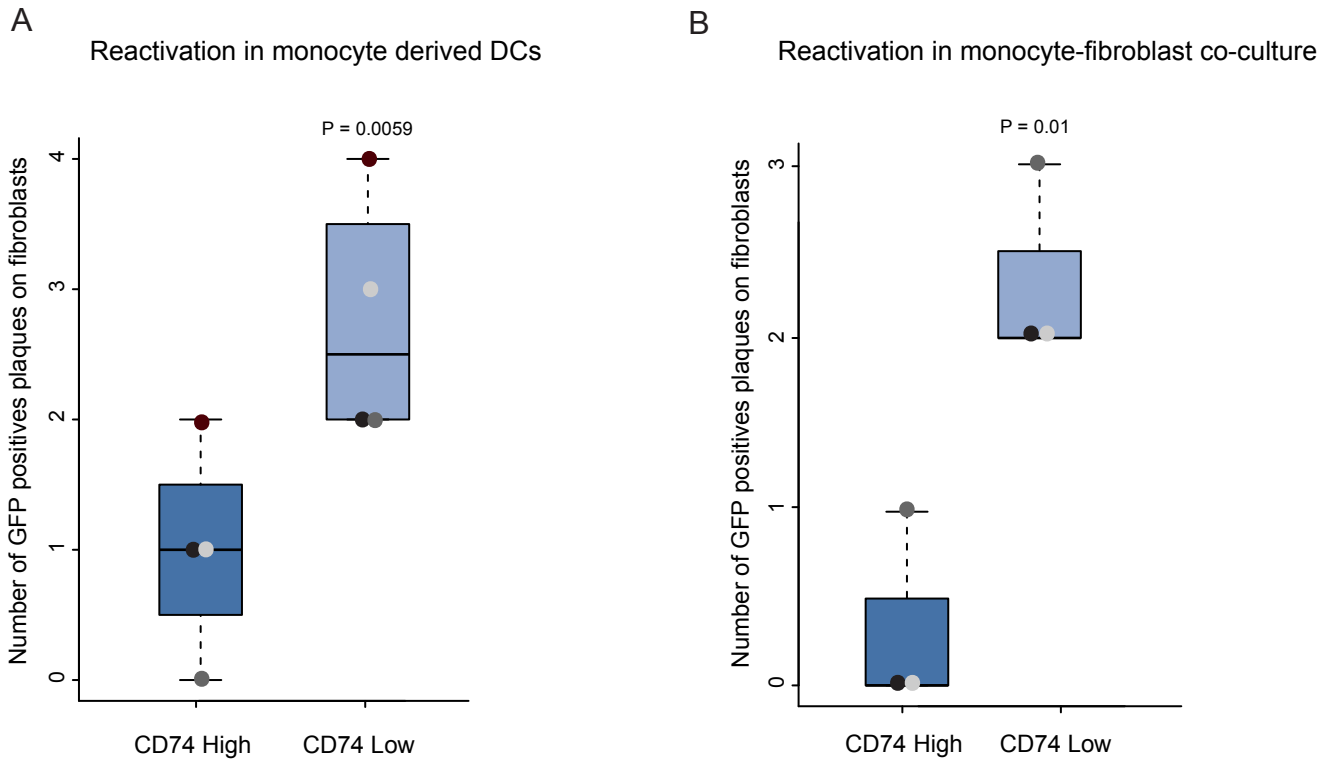
Figure 3



Changes in CD74 expression are induced by infection

Uninfected primary monocytes were FACS sorted according to cell-surface levels of CD74. Equivalent numbers of $CD74^{high}$ and $CD74^{low}$ cells were infected with HCMV and differences in CD74 RNA levels and in viral DNA and RNA levels between these two cell populations were assessed by qPCR. (A) Relative CD74 transcript levels in $CD74^{high}$ and $CD74^{low}$ cells at 8hpi. (B) Relative abundance of viral DNA in $CD74^{high}$ and $CD74^{low}$ cells at 8hpi and 72hpi. (C) Relative expression level of the viral transcripts UL22A and RNA2.7 in $CD74^{high}$ and $CD74^{low}$ cells as measured at 8hpi and 72 hpi. Graphs show a representative experiment of 3 biological repeats, error bars reflect standard deviation of 3 measurements. P values were calculated by t-test and reflect non-significant differences.

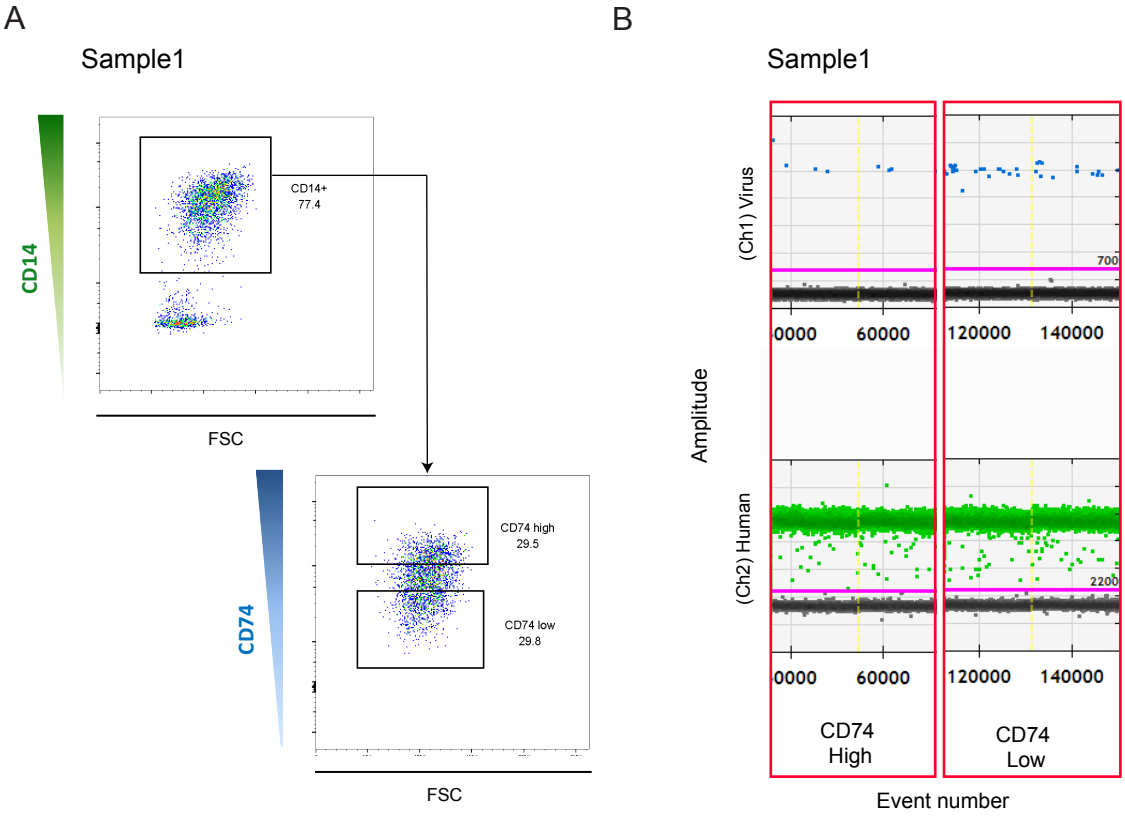
Figure 4

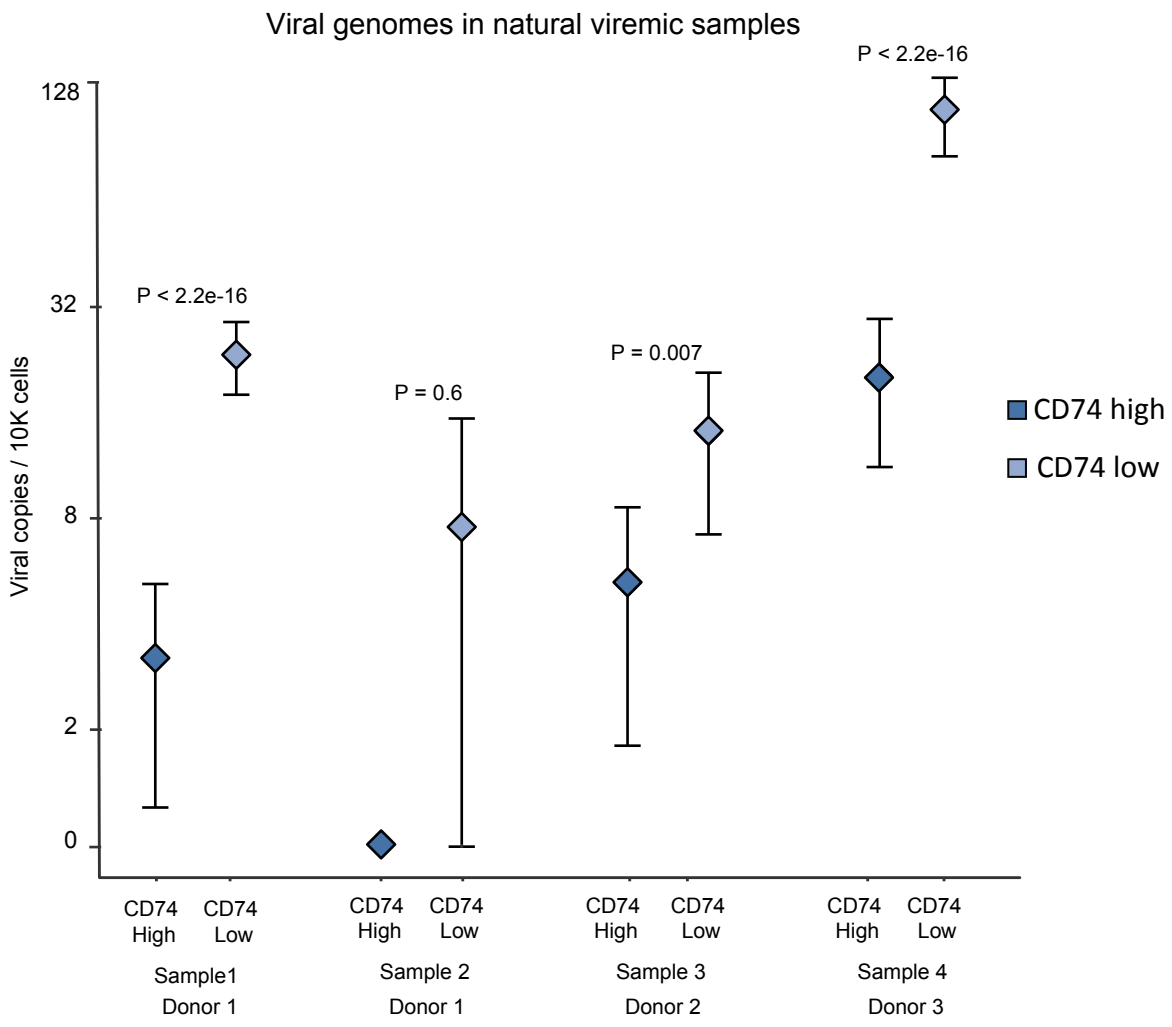


CD74^{low} infected monocytes reactivate more efficiently

HCMV infected CD14+ Monocytes were sorted according to cell-surface levels of CD74 at 3dpi. Equivalent numbers of cells were plated and (A) differentiated into DCs followed by co-culture with fibroblasts to quantify reactivation events or (B) directly co-cultured with fibroblasts to induce reactivation. Data is shown as number of infectious centers formed by CD74^{high} vs. CD74^{low} cells. Means and error bars (reflecting standard deviation) were generated from 4 independent experiments (A) and 3 independent experiments (B). Dot colors indicate sets from the same experiment. P values as calculated by t-test are indicated.

Figure 5

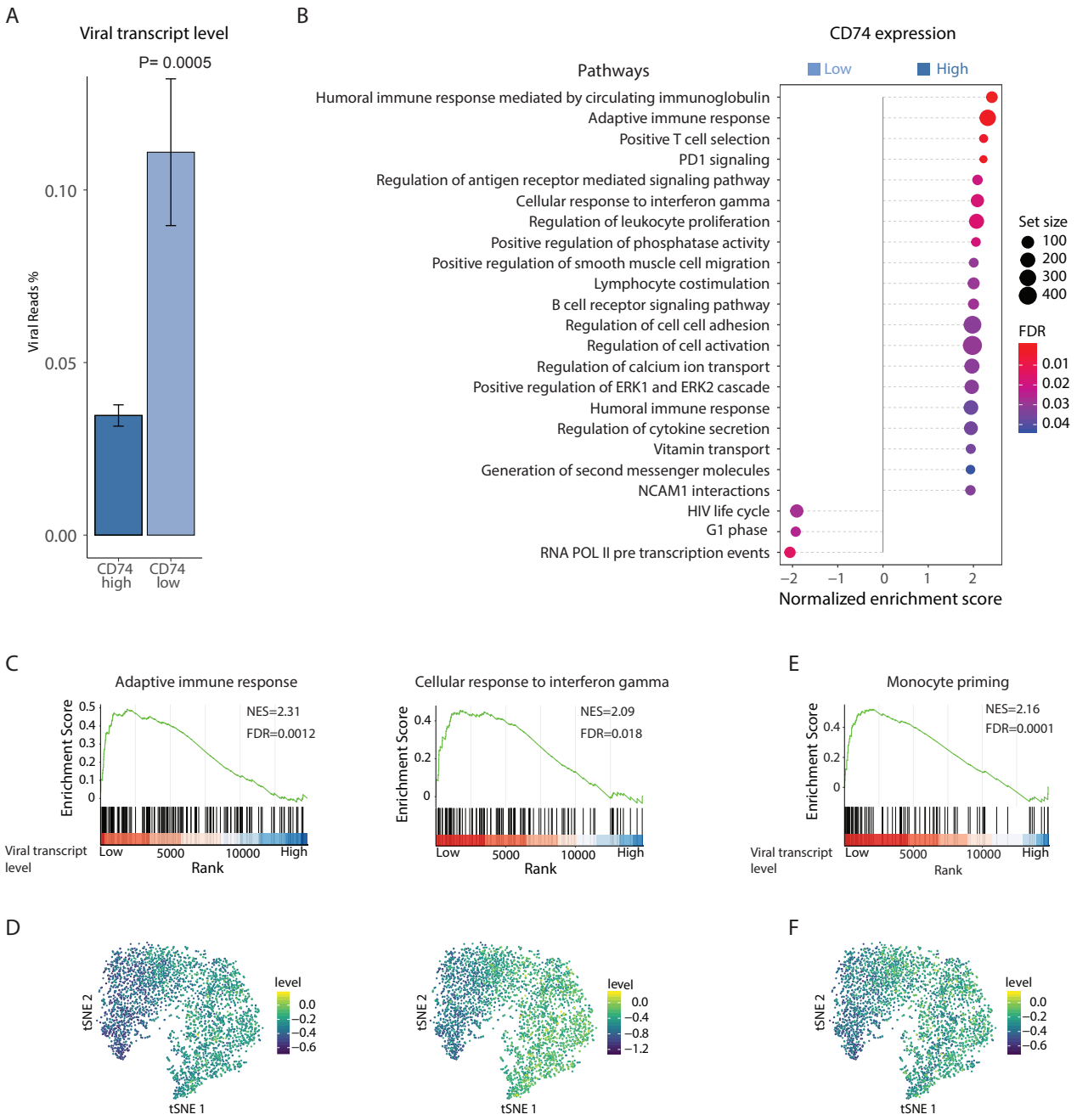

A
Sample1
CD14
CD14+
77.4
FSC
CD74
CD74 high
29.5
CD74 low
29.8
FSC
B
Sample1
(Ch1) Virus
700
0000
60000
120000
140000
Amplitude
(Ch2) Human
2200
0000
60000
120000
140000
CD74
High
CD74
Low
Event number


C
Viral genomes in natural viremic samples
Viral copies / 10K cells
128
32
8
2
0
P < 2.2e-16
P = 0.6
P = 0.007
P < 2.2e-16
CD74 high
CD74 low
CD74 High
CD74 Low
Sample1
Donor 1
Sample 2
Donor 1
Sample 3
Donor 2
Sample 4
Donor 3

CD74 cell-surface levels allow enrichment of CD14+ monocytes carrying HCMV genomes from viremic patients

CD14+ monocytes from HSCT recipients with HCMV reactivation were sorted according to CD74 cell-surface levels. In total, 4 samples were tested, originating from 3 different donors (samples 1 and 2 were collected at different time points from the same donor), and viral genome abundance was measured by ddPCR. (A) A representative FACS sort of the cells. The highest and lowest 30% of the population were collected as $CD74^{high}$ and $CD74^{low}$ samples. (B) ddPCR results of two representative replicates from a single sample, separated by yellow vertical line. Upper panel shows detection of viral DNA, lower panel reflects detection of host genomes. The magenta line marks the threshold. (B) Quantification of viral genomes in $CD74^{high}$ and $CD74^{low}$ cells from four different samples, presented as copies per 10,000 cells. Graph reflects mean and 95% CV of poisson distribution, calculated from 5 technical replicates for each donor. P values as calculated by Fisher test are indicated.

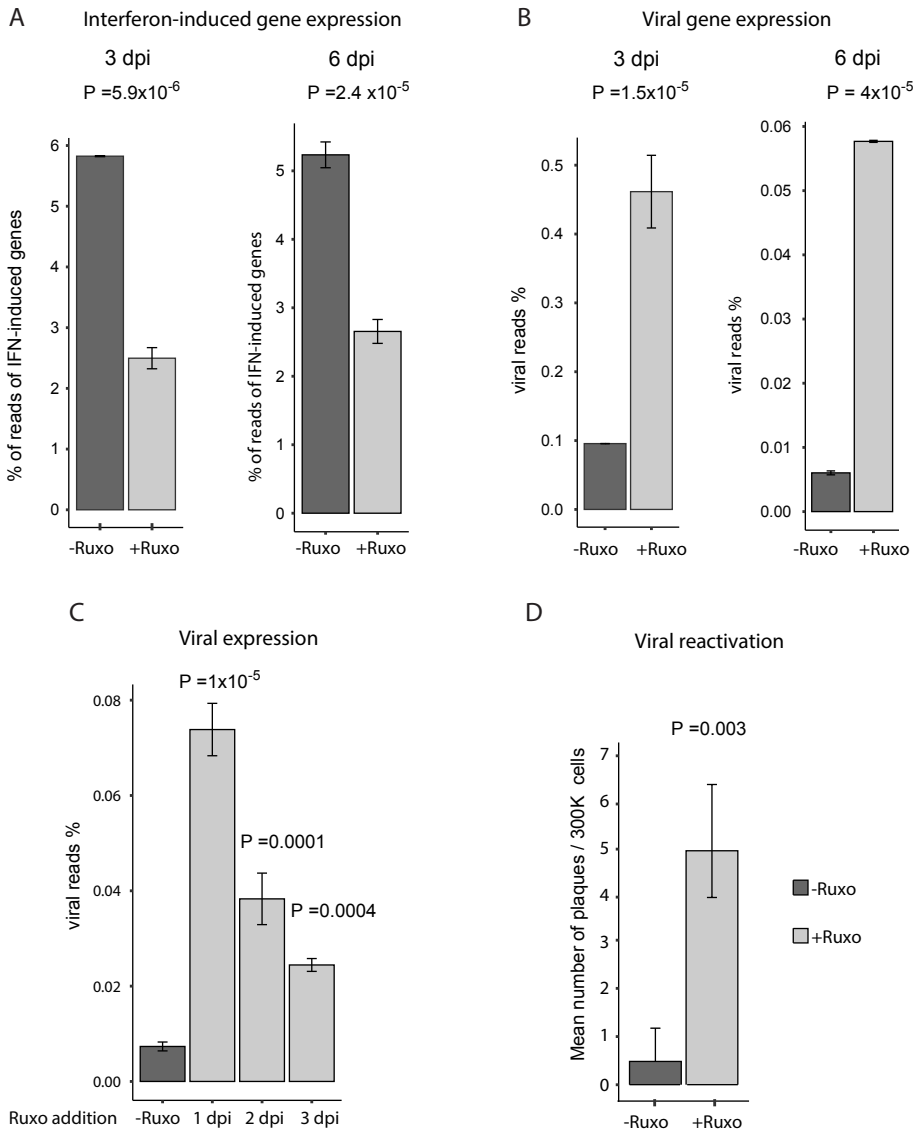
Figure 6



HCMV latency in monocytes is associated with reduced immune-response gene signatures

RNA-seq was performed on HCMV infected CD14+ Monocytes that were sorted according to cell-surface levels of CD74 at 3dpi. (A) Normalized viral gene expression in CD74^{high} and CD74^{low} cells. P value calculated using likelihood ratio test on logistic regression of viral reads, is indicated. (B) Summary of gene set enrichment analysis (GSEA) of differential expressed genes identified in RNA-seq analysis of CD74^{high} and CD74^{low} cells using annotated GO biological processes and Reactome pathways. (C) Representative pathways from GSEA of genes ranked by their differential expression between CD74^{high} and CD74^{low} cells. (D) tSNE plot of scRNA-seq of latent monocytes colored by expression level of the pathways shown in C. (E) Monocyte priming gene set from (Velten et al., 2017) analyzed on GSEA. Genes are ranked by their differential expression between CD74^{high} and CD74^{low} monocytes. (F) tSNE plot of scRNA-seq of latent monocytes colored by the expression level of the monocyte priming gene set from (Velten et al., 2017).

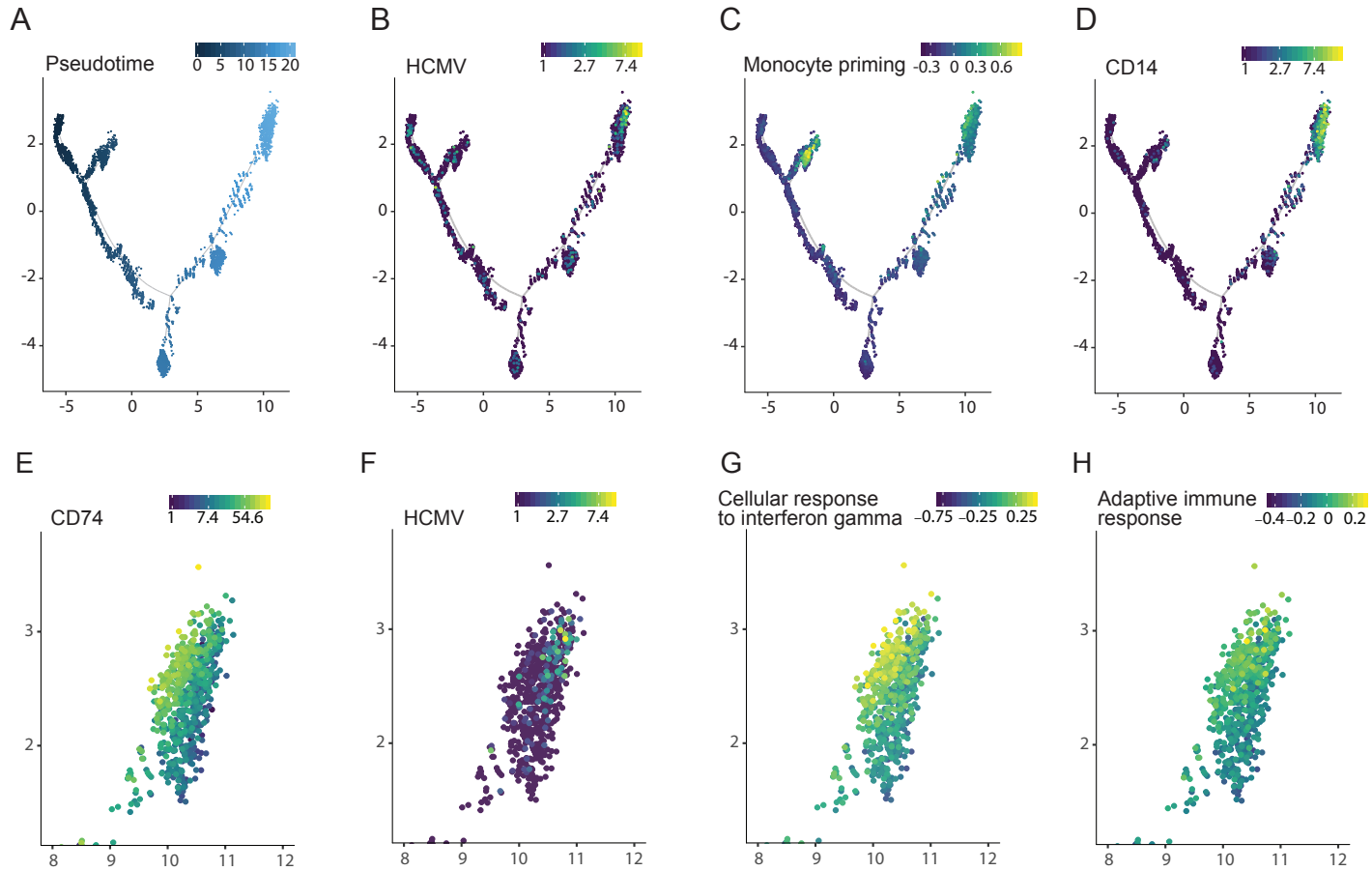
Figure 7



Inhibition of interferon signaling increases viral gene expression and promotes reactivation

(A and B) HCMV infected monocytes were treated at 3 hpi with interferon signaling inhibitor, ruxolitinib (ruxo), or left untreated, and analyzed for gene expression level, by RNA-seq. (A) Expression of interferon-induced genes at 3dpi (left) and 6dpi (right). (B) Expression of viral genes at 3 dpi (left) and 6 dpi (right). (C) Viral gene expression in HCMV infected monocytes that were treated with ruxo at 1, 2, 3 dpi or left untreated was measured by RNA-seq at 5 dpi. P values for (A-C) calculated using likelihood ratio test on logistic regression of viral reads, are indicated. (D) HCMV infected monocytes were treated at 3hpi with interferon signaling inhibitor, ruxolitinib (ruxo), or left untreated, and at 6 dpi, equivalent numbers of monocytes were co-cultured with fibroblasts to induce reactivation. Viral reactivation in ruxo treated vs. untreated cells was assessed by count of GFP positive plaques formed on the fibroblasts. Means and error bars (reflecting standard deviation) were calculated from 2 independent biological repeats. P value as calculated using likelihood ratio test on Poisson regression of positive plaque events is indicated.

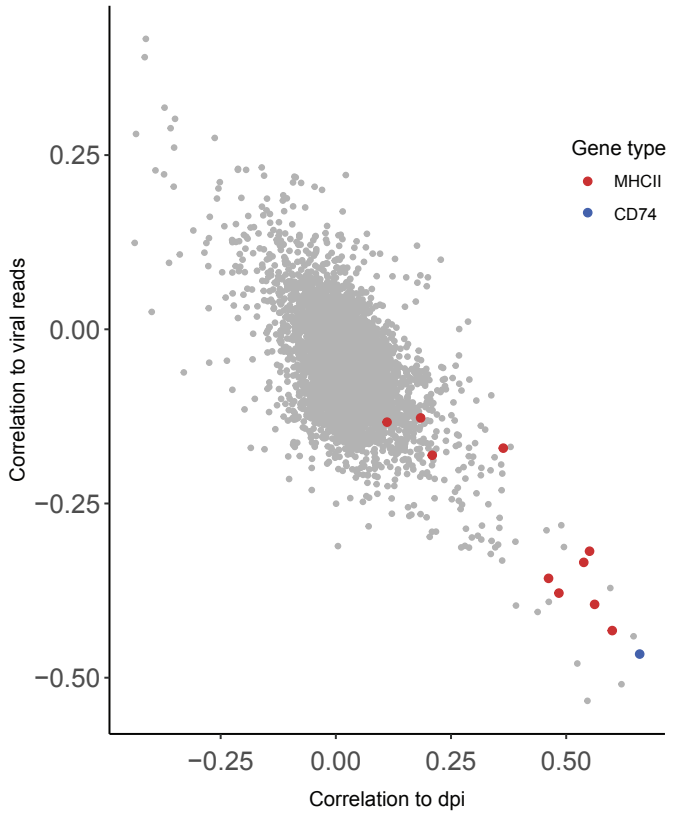
Figure 8



Viral transcript levels in HCMV-infected CD34+ HSPCs is associated with priming towards the monocyte lineage and with reduced immune-response

Pseudotime trajectory of single cell transcriptomes from HCMV infected CD34+ HSPCs (n= 7,634, (Shnayder et al., 2018)) colored by pseudotime (A), viral transcript levels (B), expression of monocyte priming gene signature (Velten et al., 2017)(C) and CD14 expression (D). Zoom-in representation of the pseudotime trajectory in A-D, on the region exhibiting higher viral transcript level (top right group of cells) , colored by expression of CD74 (E), viral transcript levels (F), expression level of the cellular response to interferon gamma pathways (G) and of the adaptive immune response pathway (H).

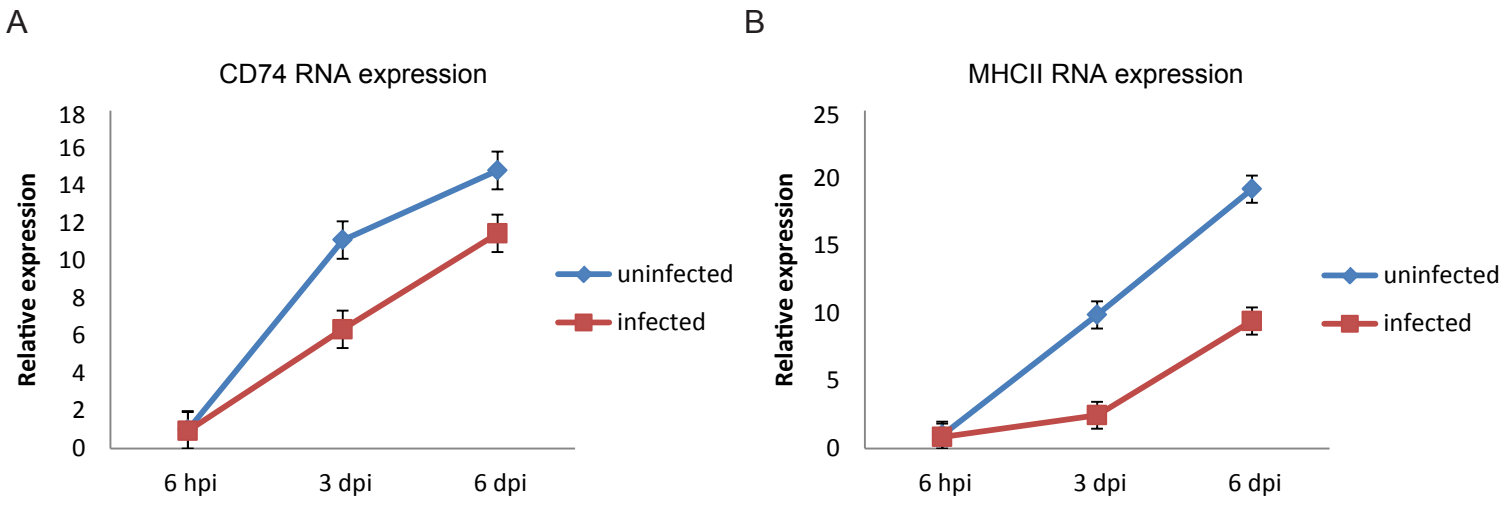
Figure 1- figure supplement 1



High association between the correlation of host gene expression vs. viral gene expression and host gene expression vs. time post infection

Scatterplot showing correlation of host transcript levels and total HCMV transcript levels compared to correlation of host transcript levels and time along infection.

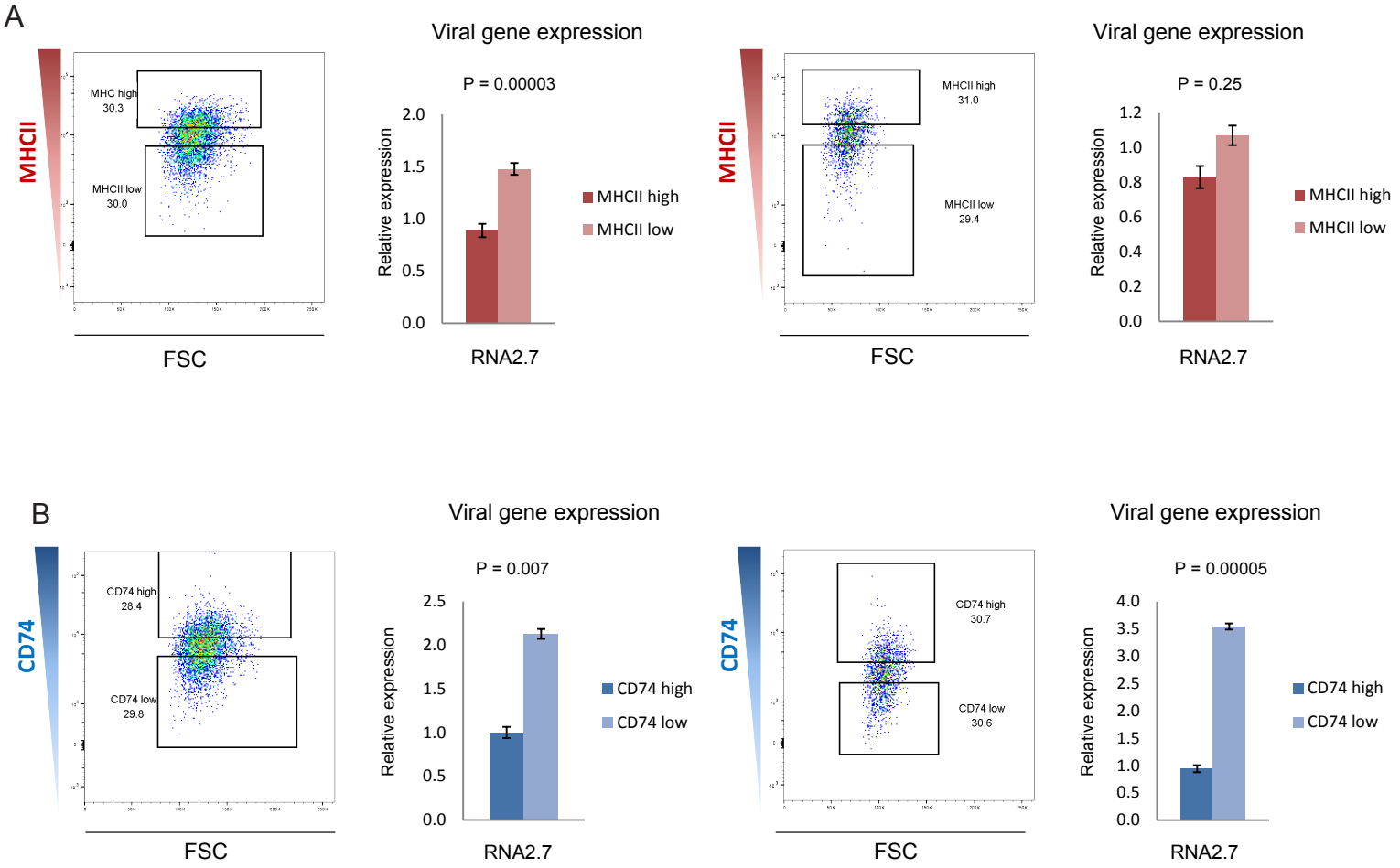
Figure 1- figure supplement 2



CD74 and MHCII expression increases with culturing time of CD14+ monocytes

Relative expression levels of CD74 (A) and MHCII (B) in HCMV infected cells and uninfected cells, measured by RT-qPCR at 6 hpi, 3 dpi and 6 dpi. Error bars reflect standard deviation of 3 measurements. A representative analysis of two independent experiments is shown.

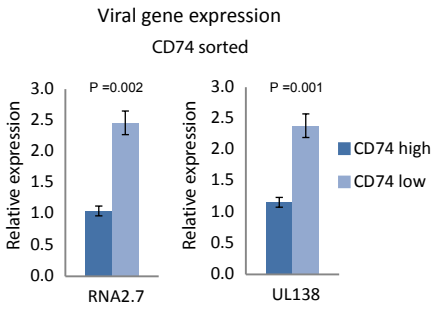
Figure 2- figure supplement 1



CD74 and MHCII cell-surface levels in HCMV infected monocytes inversely-correlate with viral transcript levels

HCMV infected monocytes were FACS sorted according to cell-surface levels of (A) MHCII or (B) CD74, at 3 dpi. High and low gates were determined as highest and lowest 30% of the population, respectively. Two biological repeats are shown. Viral expression levels in cells from each gate were assessed by RT-qPCR measurement of the viral transcript RNA2.7. Standard deviation and P values, by t-test, were calculated from 3 technical replicates.

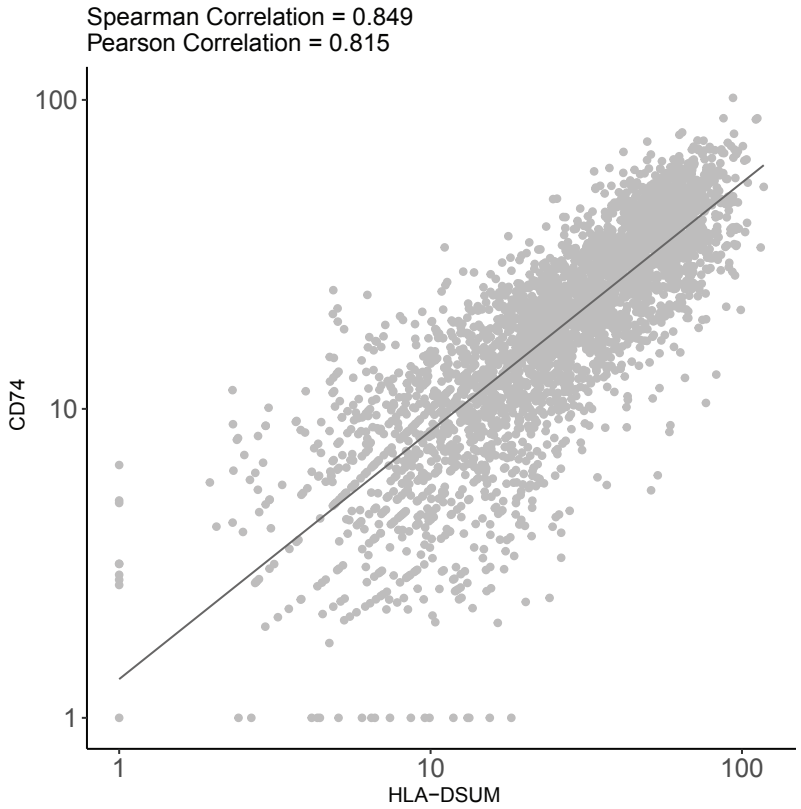
Figure 2- figure supplement 2



Cell-surface CD74 levels in HCMV infected monocytes at 6dpi inversely-correlate with viral gene expression

Expression level of viral transcripts RNA2.7 and UL138, as RT-qPCR measured in HCMV infected cells, sorted by CD74 cell-surface levels at 6 dpi. Error bars reflect standard deviation of 3 technical replicates. P value, as calculated by t-test, is indicated.

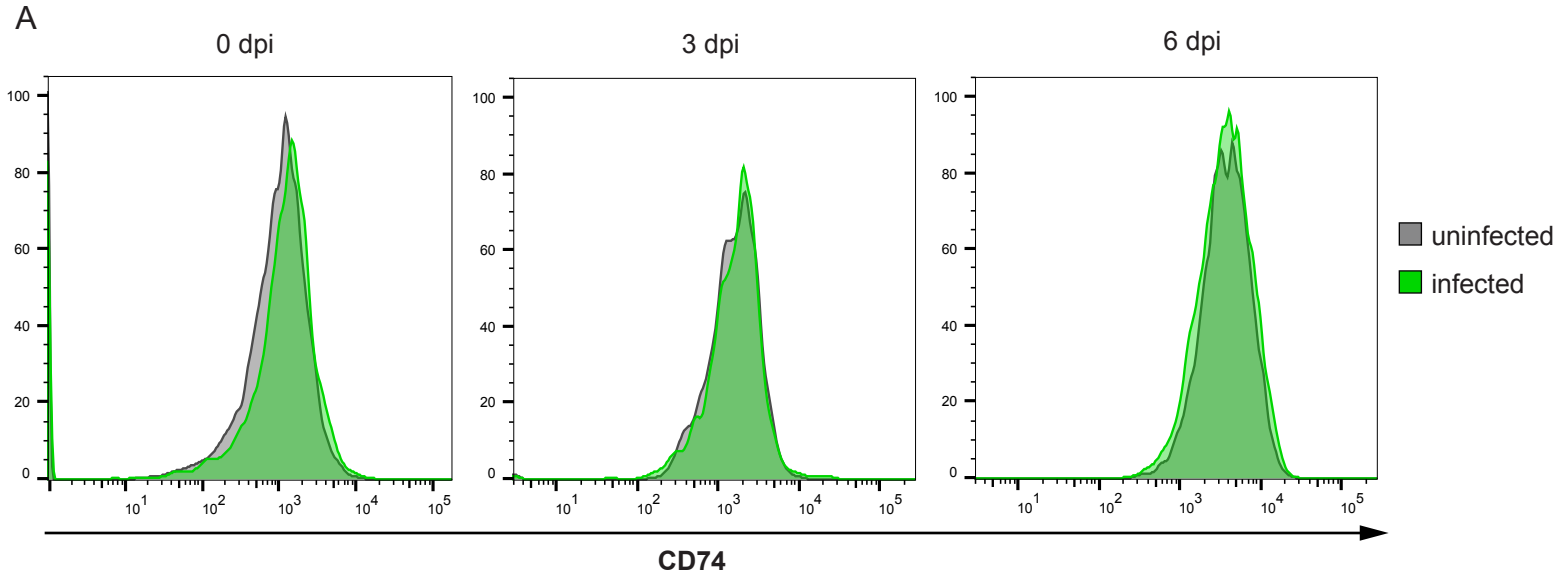
Figure 2- figure supplement 3



CD74 and MHCII genes are co-expressed in HCMV infected monocytes

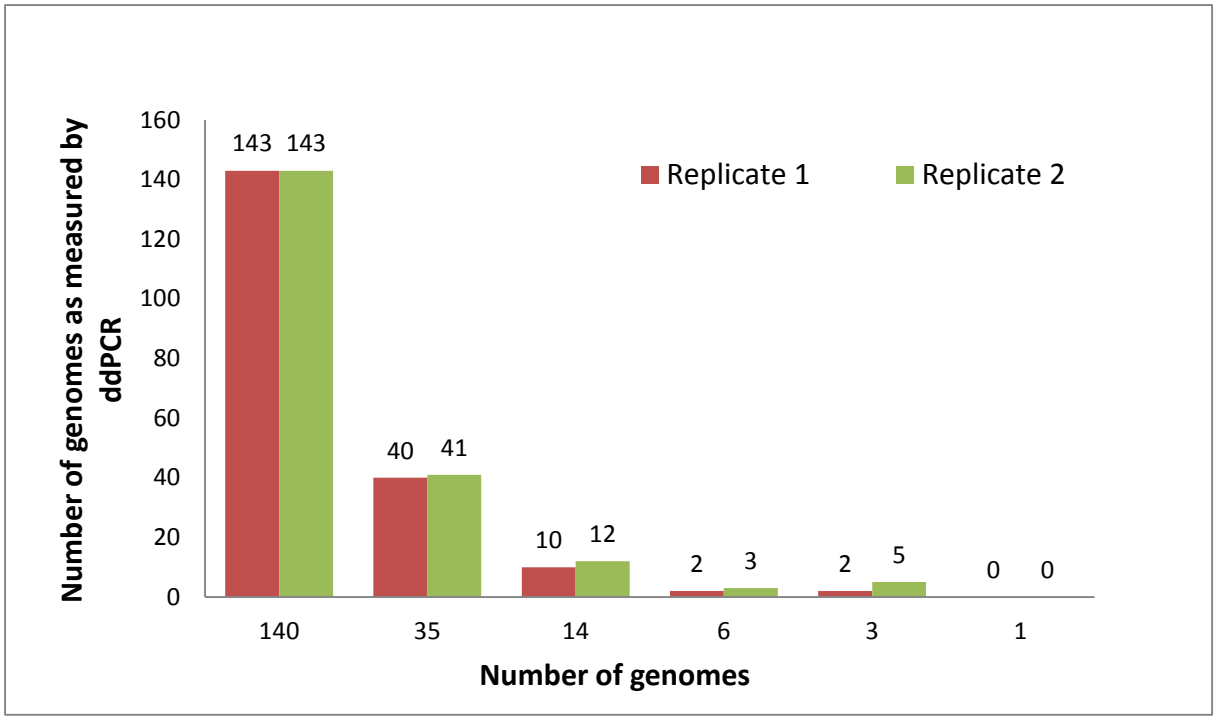
Scatterplot showing read number of CD74 vs. read number of MHCII genes in single HCMV infected monocytes according to scRNA-seq data (Shnayder et al., 2018).

Figure 2- figure supplement 4



Surface expression distribution of CD74 does not change in uninfected and infected cell populations
Infected (green) and uninfected (grey) cells were stained for surface expression of CD74 and analyzed by flow cytometry at 0, 3 and 6 dpi.

Figure 5- figure supplement 1

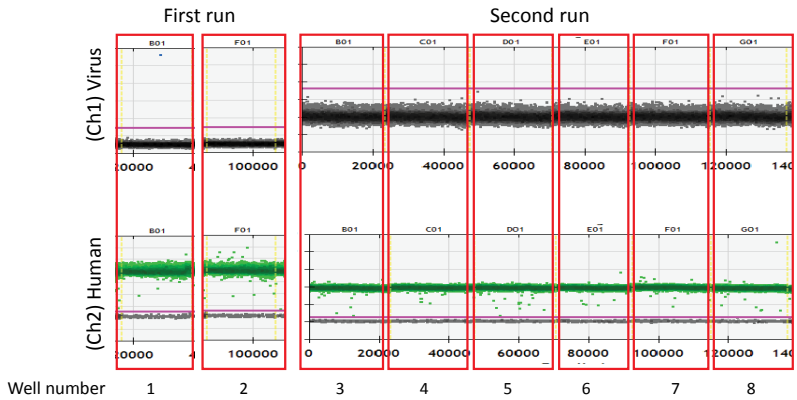


Sensitivity of viral genomes detection by ddPCR

Standard curve measuring the sensitivity and accuracy of viral DNA detection by ddPCR, using a CMV positive control template in copy number of: 1, 3, 6, 14, 35 and 140. Detection was possible down to 3 copies per sample.

Figure 5- figure supplement 2

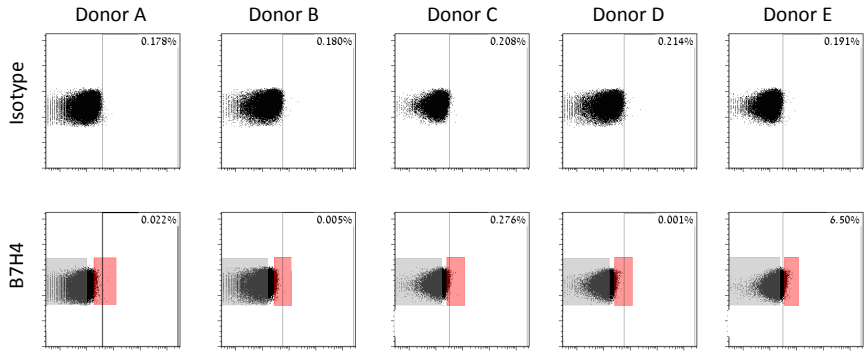
A



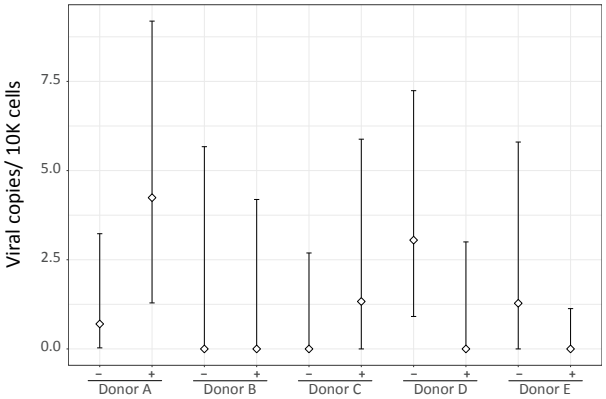
B

Well number	Viral genomes/ Sample	Cells/ Sample
1	1	53100
2	0	52300
3	0	49200
4	0	51900
5	0	49700
6	0	50600
7	0	50100
8	0	50500

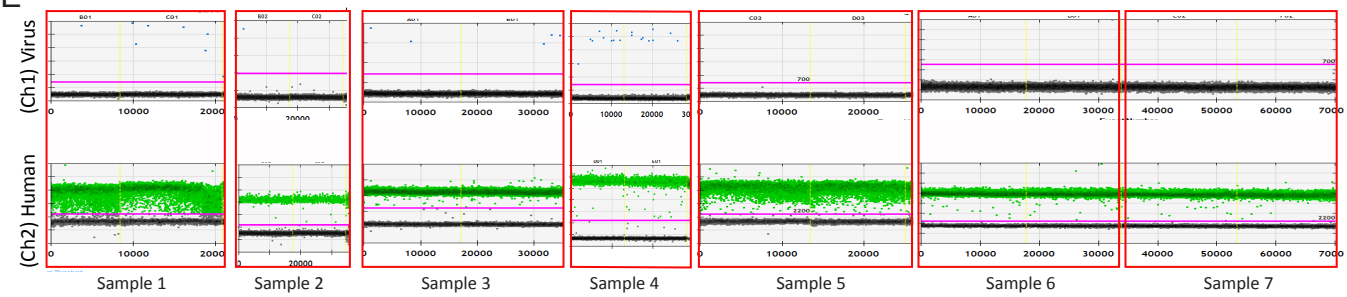
C



D



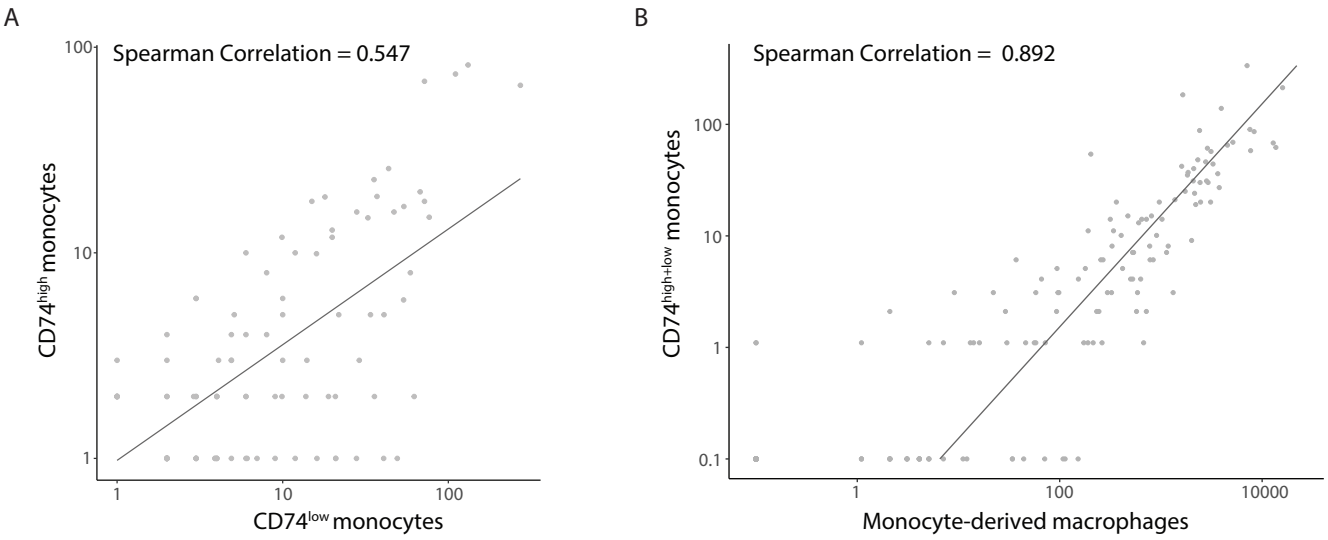
E



ddPCR analysis of natural samples

(A) Analysis of samples from a healthy seropositive donor by ddPCR, done in eight replicates. Upper panel shows detection of viral DNA (1 positive event), lower panel reflects detection of host genomes. The magenta line marks the threshold. (B) Summary of all positive events detected in A. (C) Monocytes from 5 healthy seropositive individuals were FACS sorted according to B7H4 levels. A gate defining positive cells, as determined by minimal staining with isotype control (0.2%-0.8%) for each donor, and the sorting gates indicating the top 2% and bottom 70% are shown. (D) Quantification of viral genomes in the bottom 70% and top 2% B7H4 cells, designated as – and +, respectively, from 5 different donors, presented as copies per 10,000 cells. Graph reflects mean and 95% CV of poisson distribution, calculated from 2-4 technical replicates for each donor. (E) ddPCR analysis of monocytes from 7 samples originating from 6 HSCT patients with HCMV viremia. Each sample was tested in a duplicate for presence of viral genomes, shown separated by yellow vertical lines. Monocytes from donors that were positive for viral genomes by ddPCR (samples 1-4) were used for CD74 analysis, as shown in Figure 5.

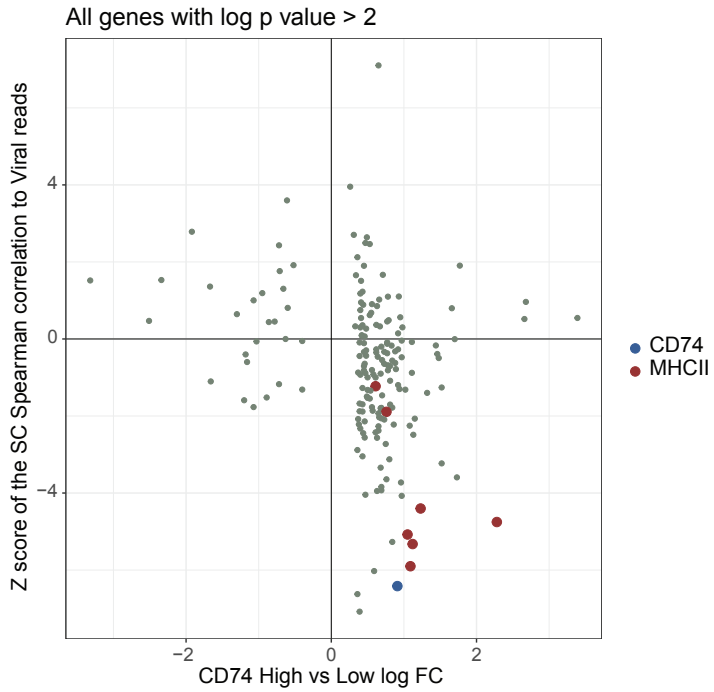
Figure 6- figure supplement 1



Viral gene expression profile of infected CD14+ monocytes correlates between CD74^{high} and CD74^{low} cells and to late lytic profile

Scatterplot showing read number of all viral genes in CD74^{high} compared with CD74^{low} cells at 3dpi (A). Scatterplot showing read number of all viral genes in CD74^{high} +CD74^{low} cells at 3 dpi compared to lytically infected monocyte-derived macrophages at 4dpi (B).

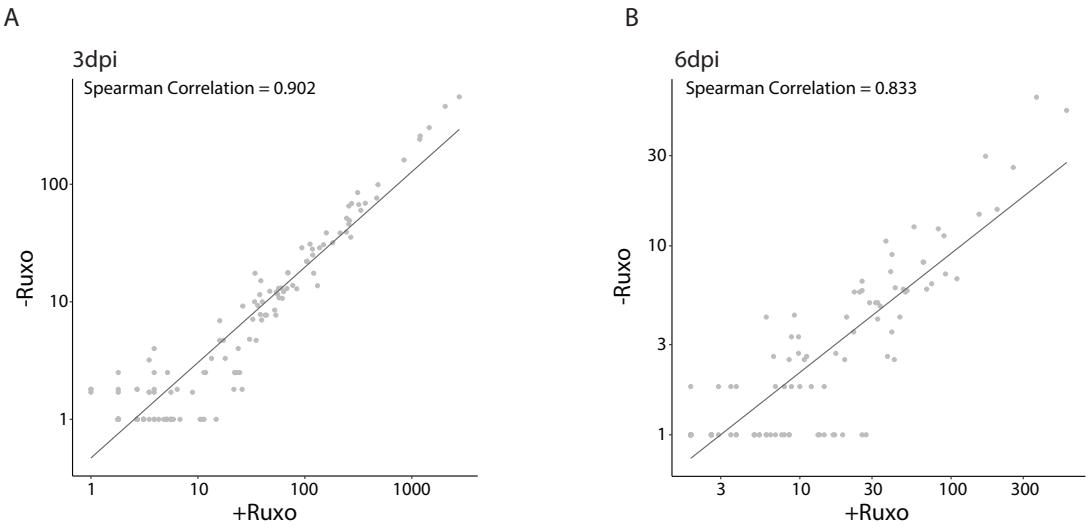
Figure 6- figure supplement 2



Comparison of changes detected in bulk RNA-seq and scRNA-seq data

Relationship between the fold changes in RNA-seq of CD74^{high} vs. CD74^{low} infected monocytes and correlation with viral transcript levels in infected monocyte scRNA-seq. Only genes that were differentially expressed in the bulk RNA-seq with P values< 0.01 are shown.

Figure 7- figure supplement 1



Uniform induction of expression across viral genes by inhibition of interferon signaling

Scatterplot showing read number of all viral genes in ruxolitinib treated cells compared to untreated cells at (A) 3 dpi and (B) 6 dpi.