

# Impact of dual baculovirus infection on the Sf9 insect cell transcriptome during AAV production using single-cell RNA-seq

Nikolaus Virgolini<sup>1,2</sup>, Marco Silvano<sup>1,2</sup>, ryan hagan<sup>3,4</sup>, Ricardo Correia<sup>1,2</sup>, Paula Alves<sup>1,2</sup>, Colin Clarke<sup>3,4</sup>, Antonio Roldao<sup>1,2</sup>, and Inês Isidro<sup>1,2</sup>

<sup>1</sup>Instituto de Biologia Experimental e Tecnológica

<sup>2</sup>ITQB NOVA, Instituto de Tecnologia Química e Biológica António Xavier

<sup>3</sup>National Institute for Bioprocessing Research and Training

<sup>4</sup>School of Chemical and Bioprocess Engineering, University College Dublin

February 22, 2024

## Abstract

The insect cell-baculovirus expression vector system (IC-BEVS) has shown to be a powerful platform to produce complex biopharmaceutical products, such as recombinant proteins and VLPs. More recently IC-BEVS has been also used as an alternative to produce adeno-associated virus (AAV). However, little is known about the variability of insect cell populations and the potential effect of heterogeneity on product titer and/or quality. In this study, transcriptomics analysis of Sf9 insect cells during the production of recombinant AAV using a low multiplicity of infection, dual-baculovirus system was performed *via* single-cell RNA-seq (scRNA-seq). Before infection, the principal source of variability in Sf9 insect cells was associated to cell cycle. Over the course of infection, an increase in transcriptional heterogeneity was detected, this being linked to the expression of baculovirus genes as well as to differences in AAV transgenes (*rep*, *cap* and *gfp*) expression. Noteworthy, at 24 hours post-infection (hpi) only 29 % of cells showed to enclose all three necessary AAV transgenes to produce packed AAV particles, indicating limitations of the dual baculovirus system. In addition, the trajectory analysis herein performed highlighted biological processes such as protein folding, metabolic processes, translation and stress response has been significantly altered upon infection. Overall, this work reports the first application of scRNA-seq to the IC-BEVS and highlights significant variations in individual cells within the population, providing insight for rational cell and process engineering towards improved AAV production in IC-BEVS.

## Introduction

Since the development of the baculovirus expression vector system (BEVS) (Smith et al., 1983), insect cells (IC) have been established as a powerful platform for recombinant protein production. The IC-BEVS is scalable, capable of producing proteins with complex post-translational modifications (Ailor and Betenbaugh, 1999; Fernandes et al., 2022a), and induces high expression of transgenes leading to reduced production times (Correia et al., 2020; Correia et al., 2022; Fernandes et al., 2022b). These advantages position IC-BEVS as a potential alternative to traditional systems in the production of complex products like virus-like particles (VLPs) and viral vectors such as recombinant adeno-associated virus (AAV). Nevertheless, the lack of a comprehensive understanding of the severe impact of the viral infection on the host cell machinery still poses major challenges to further improve this expression system.

Efforts have been made to characterize the insect cell host response to virus infection and/or heterologous gene expression. Technologies such as metabolomics (Carinhas et al., 2010; Monteiro et al., 2012; Monteiro

et al., 2016), transcriptomics (Chen et al., 2014; Koczka et al., 2018; Silvano et al., 2022; Wei et al., 2017 and Virgolini et al., 2022) and proteomics (Carinhas et al., 2011; Nayyar et al., 2017; Yu et al., 2016) have been employed by our group and others to shed light on the underlying biological mechanisms of cultured insect cells and BEVS. Exploring the response to virus infection on a gene expression level might be especially interesting, as the baculovirus has shown to take-over the cells gene expression machinery, activating stress response mechanisms such as apoptosis, and impacting protein folding and translation mechanisms, among others (Koczka et al., 2018; Wei et al., 2017; Virgolini et al., 2022).

Recent studies have provided genome and transcriptome references for *Spodoptera frugiperda* and *Trichoplusia ni*, (Chen et al., 2019; Xiao et al., 2020) and such advancements could thereby be a milestone in the study of gene expression alterations in Sf9 and High Five insect cells during baculovirus infection and/or heterologous protein expression. Indeed, an emerging number of transcriptomic studies in both cell lines can be observed (Koczka et al., 2018; Silvano et al., 2022; Virgolini et al., 2022), highlighting the interest in understanding host cell response to guide rational design of targeted cell line development and process engineering approaches. Nevertheless, these studies are limited by their approach of assessing cell populations instead of single cells, thus ignoring potential heterogeneity within the production platform. Moreover, single-cell RNA sequencing (scRNA-seq) would have the added benefit of identifying sub-populations of cells with distinct gene expression profiles, which might result in phenotypically beneficial traits (Ke et al., 2022).

Single-cell transcriptomics technologies have matured rapidly, allowing the study of gene expression patterns of tens of thousands of single cells in an accurate and cost-effective manner. ScRNA-seq has become standard in evaluating the transcriptome profiles of different cell types within tissues. Nevertheless, first applications in clonally derived cell lines as well as virus infection processes have been shown, each indicating significant heterogeneity within the respective cell population (Russell et al., 2018; Sun et al., 2020; Tzani et al., 2021). While this could indicate the potential complexity and/or heterogeneity in production processes using insect cells and BEVS, the use of single-cell transcriptome analysis to characterize this expression system is so far inexistent.

In this study, we implemented for the first time scRNA-seq in Sf9 insect cells during the production of AAV using a low multiplicity of infection (MOI), dual-baculovirus infection process, assessing population heterogeneity and gene expression profiles prior to and along infection.

## Material and Methods

### Cell culture

*Spodoptera frugiperda* Sf9 cells (Invitrogen, Cat#: 11496-015) were routinely sub-cultured every 2-3 days at  $0.6-1 \times 10^6$  cell.mL<sup>-1</sup> in serum-free Sf900-II<sup>TM</sup> SFM medium (Thermo Fisher Scientific) when the cell concentration reached  $2-4 \times 10^6$  cell.mL<sup>-1</sup>. Cells were cultured in shake flasks (Corning) using 10% (w/v) working volume and maintained at 27 °C in an Inova 44R shaking incubator (orbital motion diameter of 2.54 cm; Eppendorf) at 100 rpm.

### Baculovirus amplification and storage

Two recombinant baculoviruses (rBAC) were used for AAV production, one incorporating a GFP transgene flanked by inverted terminal repeats of AAV serotype 2 (AAV2) and under control of the cytomegalovirus promoter (hereby named rBAC-GFP, kindly provided by Généthon) and a second rBAC carrying AAV2 *rep* and *cap* genes (hereby named rBAC-REP/CAP), produced in-house using Addgene plasmid #65214 (Pais et al., 2019; Smith et al., 2009). Amplification of baculovirus stocks and storage was performed as described previously (Virgolini et al., 2022). Baculovirus titers were determined using the MTT assay as described elsewhere (Mena et al., 2003; Roldão et al., 2009).

## AAV production

AAV production was carried out in a 0.5 L stirred tank bioreactor (Sartorius BIOSTAT Qplus) by infecting Sf9 cells at  $2 \times 10^6$  cell.mL<sup>-1</sup> with rBAC-REP/CAP and rBAC-GFP, both at a multiplicity of infection (MOI) of 0.05 pfu/cell. Each vessel was equipped with one Rushton impeller; gas flow of 0.01 VVM was supplied through a ring sparger. pO<sub>2</sub>(partial pressure of oxygen) setpoint of 30% of air saturation was set, which was achieved through varying the agitation rate (70 - 300 rpm) and percentage of O<sub>2</sub> in the gas flow (0 - 100 % of O<sub>2</sub>).

Cultures were maintained for up to 96 hours post-infection (hpi). Samples for assessment of cell concentration, viability, intracellular AAV titer and metabolite concentration were taken every 24 hours. For scRNA-seq analysis, samples were taken at 0, 10 and 24 hpi.

## Analytics

### *Cell concentration and viability*

Cell concentration and viability were determined with the trypan blue dye exclusion method (Tennant, 1964) using the Cedex HiRes Analyzer (Roche).

### *Intracellular AAV viral genomes quantification*

Cell culture samples were collected and intracellular AAV titer quantified as described elsewhere (Pais et al., 2019). Intracellular AAV viral genomes (VG) were quantified by real-time quantitative PCR (RT-qPCR), as established previously (Virgolini et al., 2022).

## Single-cell RNA sequencing

Single-cell isolation and sample processing for scRNA-seq was performed, using the BD Rhapsody<sup>TM</sup> Express Single-Cell Analysis System (BD Biosciences) according to manufacturer's instructions. In short, cultured cells were centrifuged (300 × g, 4 °C, 5 min), washed and strained (30 μm mesh size – CellTrics<sup>®</sup>) prior dilution to the recommended cell concentration to target 6,000 single cells. Cells were then captured in nanowell-containing cartridges, lysed, and the released mRNA isolated using poly(dT)-coated magnetic beads.

Sequencing libraries were prepared using the mRNA Whole Transcriptome Analysis (WTA) Library Preparation Protocol (BD Biosciences), as recommended by the manufacturer, and using unique index primers for each library. The quality of each library was assessed using a high-sensitivity D5000 kit (Agilent) and quantification was done by Qubit analysis (Thermo Fisher Scientific). Finally, libraries were pooled to achieve 40,000 reads per cell and sent for sequencing (Illumina NovaSeq6000) elsewhere, spiked with 20% PhiX.

## Single-cell RNA sequencing data analysis

### *UMI count matrix generation*

Raw FASTQ files were processed using the BD Rhapsody WTA Analysis Pipeline (Seven Bridges Genomics), using default parameters. A customised reference genome, which included reference sequences of *S. frugiperda* (RefSeq assembly accession: GCF\_011064685.1)( Xiao et al., 2020), *Autographa californica* multiple nucleopolyhedrovirus (ViralProj14023)(Ayres et al., 1994) and AAV transgene sequences (*rep*, *cap* and *gfp*), was supplied for mapping using STAR version 2.5.2b (Dobin et al., 2013). Finally, recursive substitution error correction (RSEC)-adjusted molecule count matrices were generated and used for downstream analysis.

### *Sample pre-processing*

Downstream analysis was performed in R (v4.2.1) using the Seurat package (v4.1.1) (Hao et al., 2021); default parameters were used to create the Seurat objects. Low-quality cells ([?] 5% mitochondrial UMI counts per cell) were removed.

### *Cell cycle scoring*

To determine the cell cycle phase of each cell, a score indicating the likelihood of cells being in either S or G2/M phase was assigned, based on the supplied reference gene list for respective phases from Seurat (according to mouse reference genes published from Tirosh et al., 2016). The list of mouse cell cycle genes was associated to the *S. frugiperda* genome using a protein BLAST search (e-value cut-off 0.01). Then, identified sequences were blasted back to the *S. frugiperda* genome. Genes validated after both steps were considered for cell cycle association.

### *Seurat analysis*

To evaluate cell heterogeneity along infection, data sets (from 0, 10 and 24 hpi samples) were merged prior to log-normalization. Next, 2,000 variable genes of each sample were identified using the “vst” method. Data was scaled, regressing effects caused by the total number of UMIs and dimensionality reduced using principal component analysis (PCA) using all variable genes. A specified number of principal components was chosen for subsequent UMAP analysis. Relative gene expression, differential expression analysis and gene correlation analysis were performed using the data slot of the Seurat object. Gene markers for clusters were identified using the FindMarkers function and the MAST test (Finak et al., 2015). Genes with an absolute expression change of at least 1.5-fold and a BH-adjusted  $p$ -value [?] 0.05 were deemed significant.

### *Trajectory analysis and enrichment analysis*

To understand the changes in cellular response along infection, we utilized the 10 hpi timepoint to conduct trajectory analysis within the Monocle3 package v.1.2.9 (Trapnell et al., 2014). First, the processed 10 hpi Seurat object was transformed to a Monocle object using the appropriate function within the SeuratWrappers package v.0.3.0. Next, clusters were identified using the Leiden community detection and a resolution of 0.002. A trajectory is learned, and cells ordered according to pseudotime. Finally, genes correlated with the progression of cells along the trajectory were identified. Genes were deemed significant if the  $q$ -value was  $< 0.01$  and the average expression of the respective gene in the Seurat object was  $> 0.5$ .

Overrepresentation of gene ontology (GO) terms within the gene list found to have a significant difference along pseudotime were identified using ClusterProfiler v4.0.5 (Yu et al., 2012) and a GO term reference list established previously (Virgolini et al., 2022). Terms with an adjusted  $p$ -value  $< 0.5$  were deemed significant.

## Results

### AAV production

AAV were produced by co-infecting Sf9 insect cells with rBAC-GFP and rBAC-REP/CAP (MOI of 0.05 for each virus) in a 2 L stirred tank bioreactor (**Figure 1a**). Cell growth kinetics after infection (**Figure 1b**) followed a typical profile for a low MOI infection, indicated by one cell population doubling until growth arrest at 24 hpi comparable to previous studies (Correia et al., 2022; Pais et al., 2019; Pais et al., 2020 and Virgolini et al., 2022). The production yielded  $2.5 \times 10^{10}$  VG.mL<sup>-1</sup> intracellular genome-containing AAV particles at 72 hpi (**Figure 1c**). As one bioreactor run generated enough cells for single-cell RNA-seq analysis and the cell and AAV profiles were consistent with those of previous runs from our group, no further replicates were conducted.

## scRNA-seq data processing and quality control

Samples for scRNA-seq were acquired at 0 hpi (prior to infection), 10 hpi and 24 hpi (**Figure 1b**). Following Illumina sequencing, an average of 309  $\pm$  21 million reads were acquired for each of the three samples (**Table S1**). The BD Rhapsody WTA analysis pipeline from Seven Bridges Genomics was utilized to process sequencing data, perform genome mapping, and generate RMCE-adjusted molecule UMI count matrices. An average of 5,309  $\pm$  234 barcodes per sample passed the initial quality control step (e.g., read quality filtering, RSEC adjustment, cell label filtering), yielding an average of 21,647  $\pm$  1,397 UMIs and resulting in the detection of 3,405  $\pm$  299 genes per cell (**Table S1**). Retained barcodes were imported into Seurat and further assessed using established quality parameters, such as the proportion of mitochondrial UMIs (**Figure 1d**) and the detected UMIs per cell (**Figure S1a**). For quality control, only the proportion of UMIs originating from the mitochondrial genes were considered and barcodes where this proportion was below 5% were retained, resulting in a total of 5,288, 4,750 and 5,159 cells for 0, 10 and 24 hpi, respectively, to be considered for further analysis (**Figure 1e**). While the total number of genes identified per cell showed a similar median value along infection, a bimodal distribution can be observed at 24 hpi, with 21% of cells having less than 2,000 identified genes, because of a high percentage of baculovirus mRNA molecules in cells at this timepoint (**Figure 1f** overall results of baculovirus UMIs per cell in **Figure S1b**).

## Sf9 cell heterogeneity

To assess variations in the transcriptome of Sf9 associated with infection, each timepoint was combined into a single Seurat object for analysis. Merged samples were log-normalized and scaled, regressing out variations caused by different total UMIs per cell. For principal component analysis, 2,000 variable features were considered, and 20 principal components were used to construct a UMAP and perform clustering analysis. A total of 11 clusters were identified, of which 5 were present prior infection (**Figure 2a**), indicating an inherent degree of heterogeneity in this Sf9 cell line. One cluster (Cluster 7) was clearly separated at 0 hpi. Using the FindMarkers function with the MAST differential expression test revealed that cells in Cluster 7 had significant overexpression of stress response genes (e.g., *heat shock protein 68* and *protein lethal(2)essential for life*) and ribosomal RNA, suggesting increased cell stress and/or early signs of cell death (**Table S2**).

As cell cycle has been shown to contribute to heterogeneity in scRNA-seq datasets (Tirosh et al., 2016; Tzani et al., 2021), classification of cells to cell cycle phases was attempted using the Cell Cycle Scoring method in Seurat. For this 25 G2/M and 30 S phase genes (**Table S3**) were used to calculate a score representing the probability that a certain cell is in a particular phase of cell cycle. At early stages of infection (0 and 10 hpi), the predicted classification of cells seems to corroborate observed data, as can be observed by comparing the distribution of cells associated to each cell cycle phase (**Figure S2a-b**) with the relative expression of 3 selected cell cycle-related genes (**Figure S2c**). At later stages of infection (24 hpi) this is no longer valid, with the increase in the amount of baculovirus transcripts in infected cells hindering effective association of cells to a correct cell cycle phase, as demonstrated by the overall low expression of the G1 cell cycle-related gene (**Figure S2c**) but still association with G1 cell cycle phase (**Figure S2b**). As a result, cell cycle regression was not considered for further analysis.

Following visualization of the 3 cell populations (**Figure 2a**), a clear difference in the position of each population can be observed suggesting an increased impact on the cell transcriptome as infection progresses. This impact can be associated with enhanced expression of baculovirus genes, with the furthest clusters (Clusters 5 and 8) showing an average of 92  $\pm$  3% of baculovirus mRNA per cell (**Figure 2b**) and the lowest number of detected genes, averaging 783  $\pm$  331, while total UMIs per cell stayed similar to other clusters (**Figure S3**). To support this assumption, an established early gene (*proliferating cell nuclear antigen*) and a late baculovirus gene (*viral cathepsin*) were selected and their expression mapped to the identified clusters (**Figure 2c**). While the *proliferating cell nuclear antigen* (early gene) showed elevated expression in clusters (0, 4, and 6) associated to early infected cells (i.e., 10-24 hpi), the *viral cathepsin* (late

gene) only shows high expression in clusters (5, 8, and 10) associated with more advanced stages of infection (i.e., 24 hpi).

## Expression of AAV transgenes

Following the evaluation of cell heterogeneity, cells were categorized throughout infection according to the expression of each target transgene (*rep*, *cap*, and *gfp* - **Figure 3a**). Baculovirus and transgenes were not detected in cells prior to infection (0 hpi); this number increased to 36.2 % at 10 hpi, and reached close to 100 % at 24 hpi (all cells were infected) (**Figure 3b**).

At 24 hpi, 29.4 % of cells expressed transgenes from both rBACs (i.e., expressed *gfp* and *rep/cap*) and are therefore capable of producing packaged AAV. Moreover, 59.0 % of cells expressed *gfp* only, 3.3 % of cells expressed *rep/cap*, and 8.3% of cells had still no AAV transgene expressed at this timepoint (**Figures 3a-b**). In addition, there is a poor correlation between expression of *gfp* and *rep* (**Figure 3c**); for *rep* and *cap*, their expression is similar, although *rep* seems to be expressed at higher levels than *cap* (**Figure 3d**).

## Identification of cell marker genes associated to infection

Gene expression changes associated to baculovirus infection were identified using only the 10 hpi sample. This timepoint was deemed preferential over the 24 hpi one for two reasons: (1) it contains infected, non-infected and presumably cells in intermediate states; (2) the number of UMIs acquired from the host cell transcriptome at 24 hpi timepoint was extremely low. Given the high number of non-infected cells (63.8%) at 10 hpi, merging with the 0 hpi datasets was deemed unnecessary.

The 2,000 most variable genes were retained for PCA, with the first 20 principal components to establish UMAPs. Overall, 9 clusters were identified, separated into two major cluster groups (**Figure 4a**). Clusters 3, 4, 6, 8 were identified as clusters of infected cells, with 4 and 8 being most advanced in the infection process according to the percentage of baculovirus transcripts, and clusters 0, 1, 2 as clusters of non-infected cells (**Figure 4b**). Cluster 7 diverged from the remaining clusters of non-infected cells, showing fewer total UMIs and genes per cell compared to others (**Figure S4a-b**) and thus was not included in the differential expression analysis. To identify marker genes of infected cells, the FindMarkers function was used with the MAST test. Overall, 291 differentially expressed genes between infected and non-infected cells were identified, of which 134 were upregulated and 157 downregulated in infected cells clusters (**Table S4**). Genes significantly upregulated derived mostly from the baculovirus, with only 9 identified as host cell genes; these 9 genes were associated to stress response (e.g., *heat shock protein 68*, *peroxiredoxin-6*) and microtubule mobility (*dynein regulatory complex protein-8*) (**Figure 4c**). Genes downregulated in infected cells were associated to metabolic processes (e.g., mitochondrial *glutamate dehydrogenase* and *L-lactate dehydrogenase* and *aminomethyltransferase*), ion transport (e.g., *innexin-3*) as well as apoptosis inhibitors (*baculoviral IAP repeat-containing protein 5*) (**Figure 4c**).

## Assessing significantly enriched gene ontology terms along infection

To further assess transcriptional changes along infection, non-differential expression analysis was conducted. Monocle3 was used to perform trajectory analysis and thus evaluate which genes vary along pseudotime in cells at 10 hpi. The 10 hpi sample was again deemed preferential due to the before mentioned reasons. A trajectory graph was drawn, from which it becomes evident that clusters associated to infection were furthest along pseudotime (**Figure 5a**). Monocle3 identified 5,704 genes varying along pseudotime ( $q$ -value < 0.01), of which 2616 were deemed significant (**Table S5**). Gene ontology analysis of respective genes revealed biological processes such as stress response to infection (e.g., cellular response to virus), cell growth (e.g., cell division, DNA replication and transcription), cell cycle and protein folding significantly vary along infection (**Figure 5b**).

## Discussion

The production of viral vectors for gene therapy, such as recombinant AAV, is complex and usually demands coordinated expression of multiple genes within a cell to produce a packaged and functional virus (Srivastava et al., 2021). Typically, multiple plasmids and/or adenovirus infection (mammalian systems), or various baculoviruses (insect cell system) are needed, resulting in significant levels of defective product (e.g., empty or non-infectious particles) and other process-related contaminants (e.g., helper virus) (reviewed in Merten, 2016; Penaud-Budloo et al., 2018). Insect cells with the dual baculovirus expression vector system, herein used, have shown high titers for different AAV serotypes (Cecchini et al., 2011; Pais et al., 2019; Smith et al., 2009), although implications for final product quality remain a concern (e.g. genome packaging, ratio of capsid proteins and their correct folding). The host cell transcriptional response to virus infection, as well as the heterogeneity reported in clonal CHO cells (Tzani et al., 2021) and virus populations (Sun et al., 2020), could suggest that such production processes are highly heterogeneous and can potentially impact product titers and/or quality. Thus, to assess the characteristics of the dual baculovirus system, as well as how Sf9 cells responds to baculovirus infection by transcriptome alterations, single-cell RNA-sequencing was employed.

In this study, we observed heterogeneity in non-infected Sf9 insect cells, similar to what has been reported in clonally derived cell lines (Tzani et al., 2021). While this highlights the individuality of cells within the population, the most dominant influence observed herein was associated to cell cycle, with cells identified in G2/M phase showing distinct clustering compared to others, as reported elsewhere (Tirosh et al., 2016; Tzani et al., 2021). Moreover, baculovirus infection has been hypothesized to arrest cells in a “pseudo” S phase (Rohrmann, 2019), which we could confirm in our study, i.e. we observed an increase in S phase association to infected cells. The high association to G1 phase in late infection samples was however unexpected. Upon further evaluation of the cell cycle scoring, it was observed that cell cycle phase association was biased, as high baculovirus gene expression at later infection stages masks cell cycle genes. In this scenario, cells cannot be associated to either S or G2M phases and thus are attributed to G1 by default. Despite being useful to decipher heterogeneity in non-infected samples, cell cycle scoring was not considered as a correct approach to describe heterogeneity in samples in which the host cell transcriptome is overwhelmed by foreign virus replication. An additional source of variation observed in a small sub-population of non-infected cells was correlated to the activation of stress response mechanism. While more stringent quality control parameters might exclude this cluster, this would not be possible in this system as the impact of baculovirus on the transcriptome limits the regression of some quality control parameters (i.e., number of detected genes).

Baculovirus infection has been shown to follow a random Poisson distribution (Palomares et al., 2002). However, in a dual baculovirus system this tends to be more complex, as interference and synergistic effects of both viruses can be observed (Mena et al., 2007), as well as differences in virus replication and/or infection kinetics could occur (Galibert et al., 2021). Indeed, along infection cells had higher *gfp* expression when compared to *rep* and/or *cap*, highlighting possible differences in infection kinetics between both rBACs. While this could arise from the promoters regulating the expression of each transgene (*cmv* is an earlier promoter than the later viral promoters *polh/p10*), other possible sources of asynchronous infection and replication of both baculoviruses include titer determination and random variations in infection kinetics, emphasizing a potential need for customized infection strategies.

The heterogeneity of infected cells was linked to the overexpression of baculovirus genes, infection progression and host cell transcriptome responses. However, it was also probably augmented because of the low MOI infection strategy employed here. While high MOI processes are less desirable due to challenges associated with generation of master virus stocks (e.g., larger production scale are needed), these could prove more useful if a synchronous infection between two baculoviruses is desired. Another limitation of low MOI, dual-baculovirus based processes is the fact that in case one of these baculoviruses infects and replicates faster than the other, cells infected with the more replicative virus might be unable to receive the other, in a process called super-infection exclusion (i.e., previously infected cells cannot be re-infected) (Beperet et al., 2014; Folimonova, 2012). The possibility of baculovirus reinfection has however been reported (Gotoh et

al., 2008; Mena et al., 2007), nevertheless it is still unknown how long infected cells are susceptible to new viruses entry and how efficient is the expression of the newly arrived transgene (Sokolenko et al., 2012). A recent report showed the limitations of re-infection, as a maximum of 40% of cells were found infected with both rBACs in a dual baculovirus system using similar conditions (Galibert et al., 2021), highlighting the need to assess possibilities to increase this number to improve product quality and titer.

The production of fully packed AAV particles is dependent on the presence of both recombinant baculoviruses carrying their respective transgenes in the same cell. In our study, at 24 hpi, although all cells are infected, only 29.4% of cells were shown to have all the necessary transgenes expressed to produce packaged AAV particles. While this does not necessarily correlate to subsequent protein expression levels, as AAV proteins have been shown to undergo post-transcriptional and translational regulation (Virag et al., 2009), this data raises the possibility of a potential production bottleneck. Nevertheless, the number of cells showing transcriptional capacity for producing packed AAVs could have been underestimated here, since both *rep* and *cap* transgenes are expressed using late viral promoters and thus expression of these genes in cells that have been infected in the second infection cycle (occurring between 18 and 24 hpi) might not yet be detectable at 24 hpi.

Baculovirus infection has been shown to significantly impact the host cell machinery, activating stress response, cell cycle arrest and reorganization of the cytoskeleton and cell nucleus, while shutting down cell growth and protein folding capacities among others (as reviewed in Monteiro et al., 2012 and Rohrmann, 2013). Here, similar biological processes to those previously reported were shown to vary along infection, such as stress response mechanisms (e.g., *heat shock protein 68*) (Chen et al., 2014; Koczka et al., 2018). This might arise due to the response to unfolded proteins, as folding capacity has been found impacted (Koczka et al., 2018) and could result in reduced product quality along infection. Additionally, energy metabolism alterations along infection have been reported (Bernal et al., 2009; Bernal et al., 2010; Monteiro et al., 2017) and also be associated with mitochondrial function (Chen et al., 2014; Xue et al., 2012), which were also found altered along infection here. Alteration of the *glutamate dehydrogenase* and *glutamate synthase* genes predicted to encode proteins involved in the ammonia recycling system (Bernal et al., 2009; Doverskog et al., 2000), identified herein, has been shown in bulk analysis (Virgolini et al., 2022) and further confirms the impact of infection on host cell machinery.

Overall, 75% of enriched gene ontology terms identified herein were also found in our previous bulk transcriptomics analysis (Virgolini et al., 2022). Nonetheless, the majority of the GO biological processes identified in previous bulk RNA-seq data were only found at later stages of infection, suggesting that transcriptomic alterations of a sub-population of cells might be masked in this analysis. This further underlines the added benefit of single-cell analysis in heterogeneous production systems, as it is able to dissect early transcriptional alterations due to stress and/or infection in a sub-population of cells, thus having the potential to predict subsequent population response.

## Conclusion

This work reports, to our knowledge, the first single-cell RNA-sequencing analysis of Sf9 insect cells during the production of AAV using a low MOI, dual-baculovirus system. The findings of our work highlights production bottlenecks of this system, suggesting the need to reduce cell heterogeneity (e.g., by applying synchronization strategies) and differences in transgene expression (e.g., by applying similar promoters), as well as enhance and/or prolong the production process (e.g., by manipulation of specific pathways) for its improvement. Overall, this study lays the foundation for further single-cell transcriptomics studies in insect cell-based bioprocesses and further elevates our knowledge on the underlying biological mechanism of multi-vector production processes.

## Authors contributions

Conceptualization, NV, IAI, AR, CC; Investigation, NV, MS, RCC; Data curation, NV, MSI; Formal analysis, NV, RH; Supervision, IAI, AR, PMA, CC; Funding acquisition, PMA, AR, IAI, CC; Writing – original draft, NV; Writing – review & editing, NV, IAI, CC, AR, PMA, RC, MS, RH.

## Acknowledgments

This authors gratefully acknowledge funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813453, and from Fundacao para a Ciencia e a Tecnologia/Ministerio da Ciencia, Tecnologia e Ensino Superior (FCT/MCTES, Portugal) through national funds to iNOVA4Health (UIDB/04462/2020 and UIDP/04462/2020) and the Associate Laboratory LS4FUTURE (LA/P/0087/2020). The authors would also like to thank Genethon for kindly providing the rBAC-GFP baculovirus, Marcos Sousa (iBET) for his support in running the bioreactors and Siobhan Cashman and Helen Smith (both BD Biosciences) for their support on the Rhapsody platform.

## Conflicts of interest

The authors declare no conflict of interest.

## Data availability

Raw sequencing data has been deposited to the Sequence Read Archive (SRA) with the accession number PRJNA892787.

## Supporting information

Additional supporting information can be found online in the supporting information section at the end of this article.

## References

- Ailor E, Betenbaugh MJ. 1999. Modifying secretion and post-translational processing in insect cells. *Current Opinion in Biotechnology* **10** :142–145.
- Ayres MD, Howard SC, Kuzio J, Lopez-Ferber M, Possee RD. 1994. The Complete DNA Sequence of Autographa californica Nuclear Polyhedrosis Virus. *Virology* **202** :586–605.
- Beperet I, Irons SL, Simon O, King LA, Williams T, Possee RD, Lopez-Ferber M, Caballero P. 2014. Superinfection Exclusion in Alphabaculovirus Infections Is Concomitant with Actin Reorganization. *J Virol* **88** :3548–3556.
- Bernal V, Carinhas N, Yokomizo AY, Carrondo MJT, Alves PM. 2009. Cell density effect in the baculovirus-insect cells system: A quantitative analysis of energetic metabolism. *Biotechnology and Bioengineering* **104** :162–180.

- Bernal V, Monteiro F, Carinhas N, Ambrosio R, Alves PM. 2010. An integrated analysis of enzyme activities, cofactor pools and metabolic fluxes in baculovirus-infected *Spodoptera frugiperda* Sf9 cells. *Journal of Biotechnology* **150** :332–342.
- Carinhas N, Bernal V, Monteiro F, Carrondo MJT, Oliveira R, Alves PM. 2010. Improving baculovirus production at high cell density through manipulation of energy metabolism. *Metabolic Engineering* **12** :39–52.
- Carinhas N, Robitaille AM, Moes S, Carrondo MJT, Jenoe P, Oliveira R, Alves PM. 2011. Quantitative Proteomics of *Spodoptera frugiperda* Cells during Growth and Baculovirus Infection. *PLOS ONE* **6** :e26444.
- Cecchini S, Virag T, Kotin RM. 2011. Reproducible High Yields of Recombinant Adeno-Associated Virus Produced Using Invertebrate Cells in 0.02- to 200-Liter Cultures. *Hum Gene Ther* **22** :1021–1030.
- Chen W, Yang X, Tetreau G, Song X, Coutu C, Hegedus D, Blissard G, Fei Z, Wang P. 2019. A high-quality chromosome-level genome assembly of a generalist herbivore, *Trichoplusia ni*. *Molecular Ecology Resources* **19** :485–496.
- Chen Y-R, Zhong S, Fei Z, Gao S, Zhang S, Li Z, Wang P, Blissard GW. 2014. Transcriptome Responses of the Host *Trichoplusia ni* to Infection by the Baculovirus *Autographa californica* Multiple Nucleopolyhedrovirus. *Journal of Virology* **88** :13781–13797.
- Correia R, Fernandes B, Alves PM, Carrondo MJT, Roldao A. 2020. Improving Influenza HA-Vlps Production in Insect High Five Cells via Adaptive Laboratory Evolution. *Vaccines* **8** :589.
- Correia R, Fernandes B, Castro R, Nagaoka H, Takashima E, Tsuboi T, Fukushima A, Viebig NK, Depraetere H, Alves PM, Roldao A. 2022. Asexual Blood-Stage Malaria Vaccine Candidate PfRipr5: Enhanced Production in Insect Cells. *Frontiers in Bioengineering and Biotechnology* **10** . <https://www.frontiersin.org/articles/10.3389/fbioe.2022.908509>.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29** :15–21.
- Doverskog M, Jacobsson U, Chapman BE, Kuchel PW, Haggstrom L. 2000. Determination of NADH-dependent glutamate synthase (GOGAT) in *Spodoptera frugiperda* (Sf9) insect cells by a selective <sup>1</sup>H/<sup>15</sup>N NMR in vitro assay. *Journal of Biotechnology* **79** :87–97.
- Fernandes B, Castro R, Bhoelan F, Bemelman D, Gomes RA, Costa J, Gomes-Alves P, Stegmann T, Amacker M, Alves PM, Fleury S, Roldao A. 2022a. Insect Cells for High-Yield Production of SARS-CoV-2 Spike Protein: Building a Virosome-Based COVID-19 Vaccine Candidate. *Pharmaceutics* **14** :854.
- Fernandes B, Sousa M, Castro R, Schafer A, Hauser J, Schulze K, Amacker M, Tamborrini M, Pluschke G, Alves PM, Fleury S, Roldao A. 2022b. Scalable Process for High-Yield Production of PfCyRPA Using Insect Cells for Inclusion in a Malaria Virosome-Based Vaccine Candidate. *Frontiers in Bioengineering and Biotechnology* **10** . <https://www.frontiersin.org/articles/10.3389/fbioe.2022.879078>.
- Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, Linsley PS, Gottardo R. 2015. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16** :278.
- Folimonova SY. 2012. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *Journal of Virology* **86** :5554–5561.
- Galibert L, Jacob A, Savy A, Dickx Y, Bonnin D, Lecomte C, Rivollet L, Sanatine P, Boutin Fontaine M, Le Bec C, Merten O-W. 2021. Monobac System—A Single Baculovirus for the Production of rAAV. *Microorganisms* **9** :1799.
- Gotoh T, Ando N, Kikuchi K-I. 2008. Re-Infection Profile of Baculoviruses to Sf-9 Insect Cells that Have Already Been Infected: Virus Binding and Recombinant Protein Production. *Journal of Chemical Engineering*

of *Japan* **41** :804–808.

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. 2021. Integrated analysis of multimodal single-cell data. *Cell* **0** . [https://www.cell.com/cell/abstract/S0092-8674\(21\)00583-3](https://www.cell.com/cell/abstract/S0092-8674(21)00583-3).

Ke M, Elshenawy B, Sheldon H, Arora A, Buffa FM. 2022. Single cell RNA-sequencing: A powerful yet still challenging technology to study cellular heterogeneity. *BioEssays* **44** :2200084.

Koczka K, Peters P, Ernst W, Himmelbauer H, Nika L, Grabherr R. 2018. Comparative transcriptome analysis of a *Trichoplusia ni* cell line reveals distinct host responses to intracellular and secreted protein products expressed by recombinant baculoviruses. *Journal of Biotechnology* **270** :61–69.

Mena JA, Ramirez OT, Palomares LA. 2003. Titration of Non-Occcluded Baculovirus Using a Cell Viability Assay. *BioTechniques* **34** :260–264.

Mena JA, Ramirez OT, Palomares LA. 2007. Population kinetics during simultaneous infection of insect cells with two different recombinant baculoviruses for the production of rotavirus-like particles. *BMC Biotechnology* **7** :39.

Merten O-W. 2016. AAV vector production: state of the art developments and remaining challenges. In: .

Monteiro F, Bernal V, Alves P. 2016. Metabolic drivers of IC-BEVS productivity: Tackling the production of enveloped viral particles. *Vaccine Technology VI* . [https://dc.engconfintl.org/vaccine\\_vi/110](https://dc.engconfintl.org/vaccine_vi/110).

Monteiro F, Bernal V, Alves PM. 2017. The role of host cell physiology in the productivity of the baculovirus-insect cell system: Fluxome analysis of *Trichoplusia ni* and *Spodoptera frugiperda* cell lines. *Biotechnology and Bioengineering* **114** :674–684.

Monteiro F, Carinhas N, Carrondo MJT, Bernal V, Alves PM. 2012. Toward system-level understanding of baculovirus–host cell interactions: from molecular fundamental studies to large-scale proteomics approaches. *Front. Microbiol.* **3** . <https://www.frontiersin.org/articles/10.3389/fmicb.2012.00391/full>.

Nayyar N, Kaur I, Malhotra P, Bhatnagar RK. 2017. Quantitative proteomics of Sf21 cells during Baculovirus infection reveals progressive host proteome changes and its regulation by viral miRNA. *Sci Rep* **7** :10902.

Pais DAM, Galrao PRS, Kryzhanska A, Barbau J, Isidro IA, Alves PM. 2020. Holographic Imaging of Insect Cell Cultures: Online Non-Invasive Monitoring of Adeno-Associated Virus Production and Cell Concentration. *Processes* **8** :487.

Pais DAM, Portela RMC, Carrondo MJT, Isidro IA, Alves PM. 2019. Enabling PAT in insect cell bioprocesses: In situ monitoring of recombinant adeno-associated virus production by fluorescence spectroscopy. *Biotechnology and Bioengineering* **116** :2803–2814.

Palomares LA, Lopez S, Ramirez OT. 2002. Strategies for manipulating the relative concentration of recombinant rotavirus structural proteins during simultaneous production by insect cells. *Biotechnology and Bioengineering* **78** :635–644.

Penaud-Budloo M, Francois A, Clement N, Ayuso E. 2018. Pharmacology of Recombinant Adeno-associated Virus Production. *Mol Ther Methods Clin Dev* **8** :166–180.

Rohrmann GF. 2013. *Baculovirus Molecular Biology* 3rd ed. Bethesda (MD): National Center for Biotechnology Information (US). <http://www.ncbi.nlm.nih.gov/books/NBK114593/>.

Rohrmann GF. 2019. *Baculovirus infection: The cell cycle and apoptosis. Baculovirus Molecular Biology [Internet]. 4th edition* . National Center for Biotechnology Information (US). <https://www.ncbi.nlm.nih.gov/books/NBK543456/>.

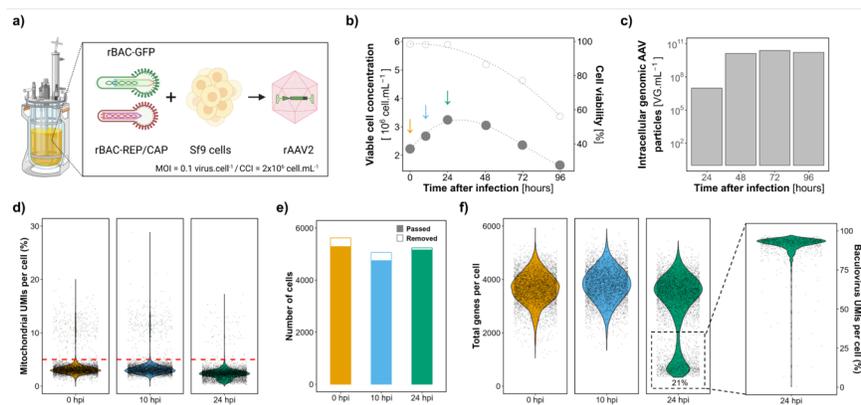
- Roldao A, Oliveira R, Carrondo MJT, Alves PM. 2009. Error assessment in recombinant baculovirus titration: Evaluation of different methods. *Journal of Virological Methods* **159** :69–80.
- Russell AB, Trapnell C, Bloom JD. 2018. Extreme heterogeneity of influenza virus infection in single cells. Ed. Arup K Chakraborty. *eLife* **7** :e32303.
- Silvano M, Correia R, Virgolini N, Clarke C, Alves PM, Isidro IA, Roldao A. 2022. Gene Expression Analysis of Adapted Insect Cells during Influenza VLP Production Using RNA-Sequencing. *Viruses* **14** :2238.
- Smith GE, Summers MD, Fraser MJ. 1983. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol Cell Biol* **3** :2156–2165.
- Smith RH, Levy JR, Kotin RM. 2009. A simplified baculovirus-AAV expression vector system coupled with one-step affinity purification yields high-titer rAAV stocks from insect cells. *Mol Ther* **17** :1888–1896.
- Sokolenko S, George S, Wagner A, Tuladhar A, Andrich JMS, Aucoin MG. 2012. Co-expression vs. co-infection using baculovirus expression vectors in insect cell culture: Benefits and drawbacks. *Biotechnology Advances* **30** :766–781.
- Srivastava A, Mallela KMG, Deorkar N, Brophy G. 2021. Manufacturing Challenges and Rational Formulation Development for AAV Viral Vectors. *Journal of Pharmaceutical Sciences* **110** :2609–2624.
- Sun J, Vera JC, Drnevich J, Lin YT, Ke R, Brooke CB. 2020. Single cell heterogeneity in influenza A virus gene expression shapes the innate antiviral response to infection. *PLoS Pathog* **16** :e1008671.
- Tennant JR. 1964. EVALUATION OF THE TRYPAN BLUE TECHNIQUE FOR DETERMINATION OF CELL VIABILITY. *Transplantation* **2** :685–694.
- Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, Fallahi-Sichani M, Dutton-Regester K, Lin J-R, Cohen O, Shah P, Lu D, Genshaft AS, Hughes TK, Ziegler CGK, Kazer SW, Gaillard A, Kolb KE, Villani A-C, Johannessen CM, Andreev AY, Allen EMV, Bertagnolli M, Sorger PK, Sullivan RJ, Flaherty KT, Frederick DT, Jane-Valbuena J, Yoon CH, Rozenblatt-Rosen O, Shalek AK, Regev A, Garraway LA. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352** :189–196.
- Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. 2014. Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat Biotechnol* **32** :381–386.
- Tzani I, Herrmann N, Carillo S, Spargo CA, Hagan R, Barron N, Bones J, Dillmore WS, Clarke C. 2021. Tracing production instability in a clonally derived CHO cell line using single-cell transcriptomics. *Biotechnology and Bioengineering* **118** :2016–2030.
- Virag T, Cecchini S, Kotin RM. 2009. Producing Recombinant Adeno-Associated Virus in Foster Cells: Overcoming Production Limitations Using a Baculovirus–Insect Cell Expression Strategy. *Hum Gene Ther* **20** :807–817.
- Virgolini N, Hagan R, Correia R, Silvano M, Fernandes S, Alves PM, Clarke C, Roldao A, Isidro IA. 2022. Transcriptome analysis of Sf9 insect cells during production of recombinant Adeno-associated virus. *Biotechnology Journal* **n/a** :2200466.
- Wei L, Cao L, Miao Y, Wu S, Xu S, Wang R, Du J, Liang A, Fu Y. 2017. Transcriptome analysis of *Spodoptera frugiperda* 9 (Sf9) cells infected with baculovirus, AcMNPV or AcMNPV-BmK IT. *Biotechnol Lett* **39** :1129–1139.
- Xiao H, Ye X, Xu H, Mei Y, Yang Y, Chen X, Yang Y, Liu T, Yu Y, Yang W, Lu Z, Li F. 2020. The genetic adaptations of fall armyworm *Spodoptera frugiperda* facilitated its rapid global dispersal and invasion. *Molecular Ecology Resources* **20** :1050–1068.

Xue J, Qiao N, Zhang W, Cheng R-L, Zhang X-Q, Bao Y-Y, Xu Y-P, Gu L-Z, Han J-DJ, Zhang C-X. 2012. Dynamic Interactions between *Bombyx mori* Nucleopolyhedrovirus and Its Host Cells Revealed by Transcriptome Analysis. *Journal of Virology* **86** :7345–7359.

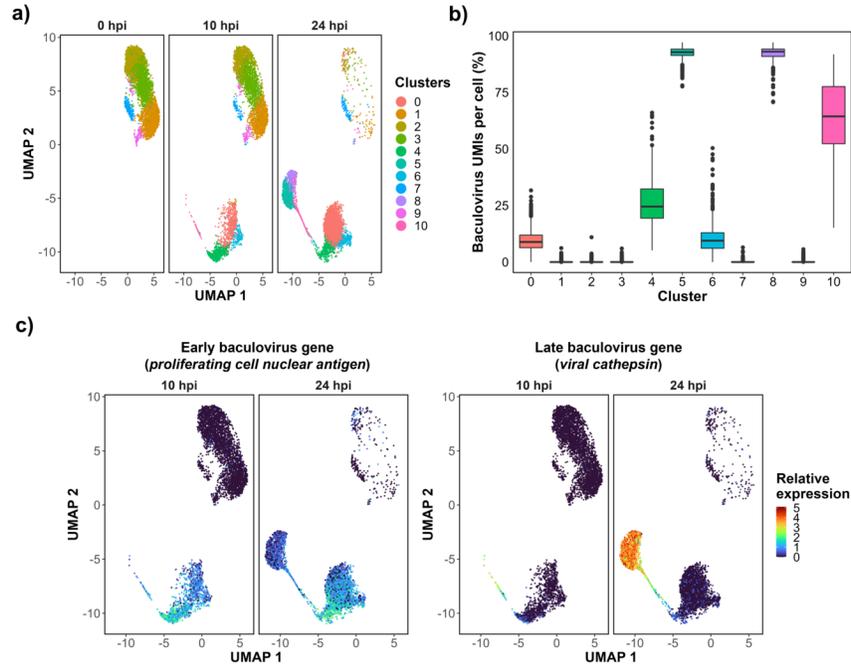
Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **16** :284–287.

Yu Q, Xiong Y, Liu J, Wang Q, Qiu Y, Wen D. 2016. Comparative proteomics analysis of apoptotic *Spodoptera frugiperda* cells during p35 knockout *Autographa californica* multiple nucleopolyhedrovirus infection. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **18** :21–29.

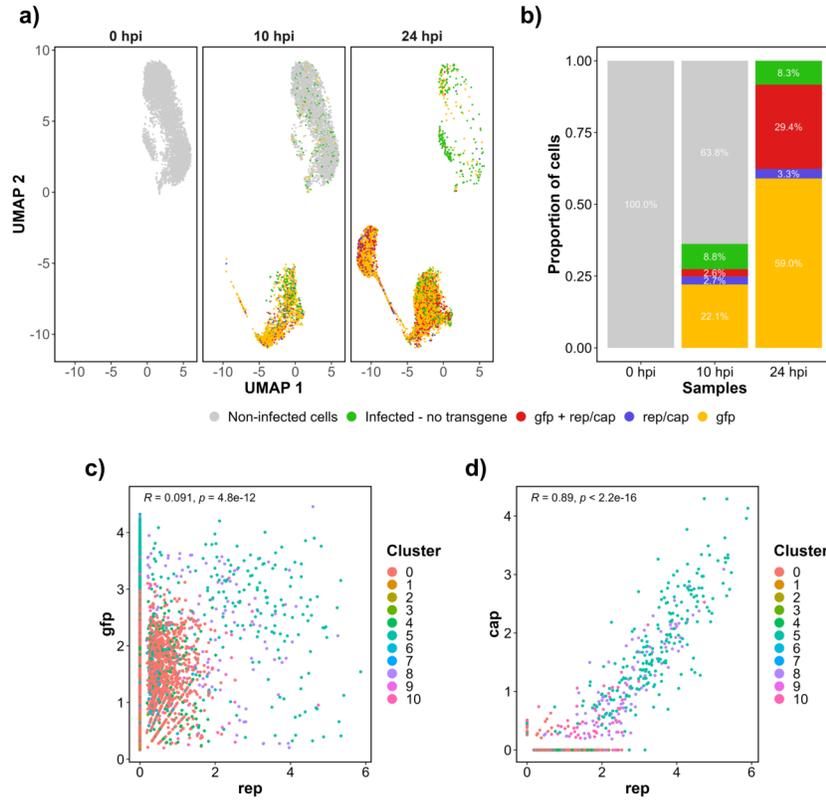
## Tables and Figures



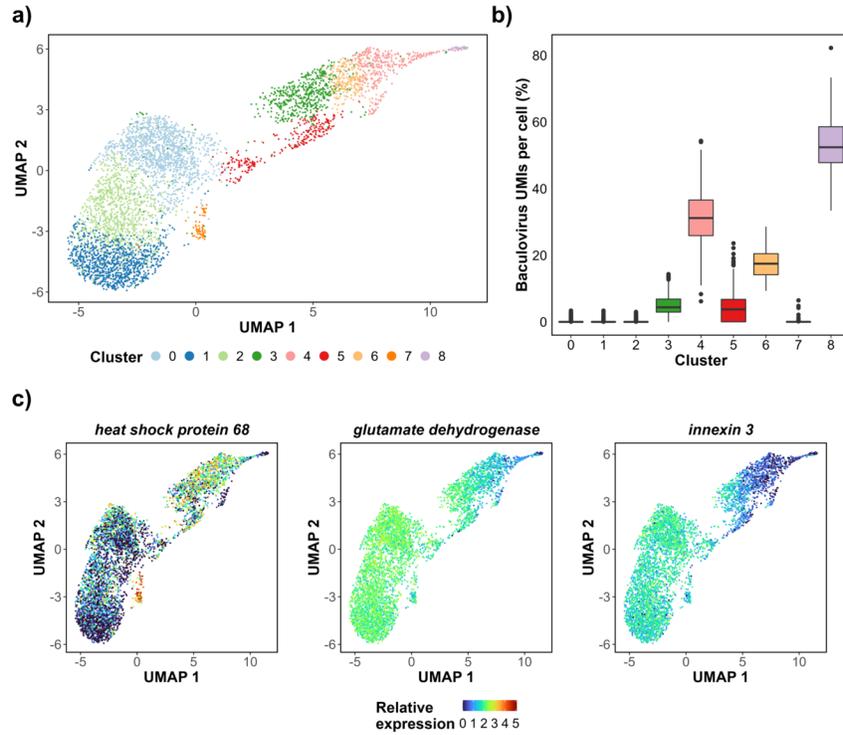
**Figure 1:** Utilisation of single-cell RNA-seq to study baculovirus-infected Sf9 insect cells during AAV2 production. a) Overview of the cell culture and infection strategy. b) Viable cell concentration (\*) and cell viability (o) throughout infection. The arrows show timepoints captured for scRNA-seq analysis. c) Intracellular genomic AAV particles ( $\text{VG}\cdot\text{mL}^{-1}$ ) along infection ( $n=1$ ). d) Percentage of mitochondrial UMIs per cell for all samples. The red dashed line indicates the quality filter threshold applied in this study (percentage of mitochondrial genes  $< 5\%$  considered for further analysis). e) Number of cells that have passed (i.e., considered valid) or have been removed in the quality filtering step in each sample. f) Total genes detected per cell. A bimodal distribution can be observed for the 24 hpi sample, with 21% of cells having less than 2000 genes detected while being associated with a high percentage of baculovirus UMIs.



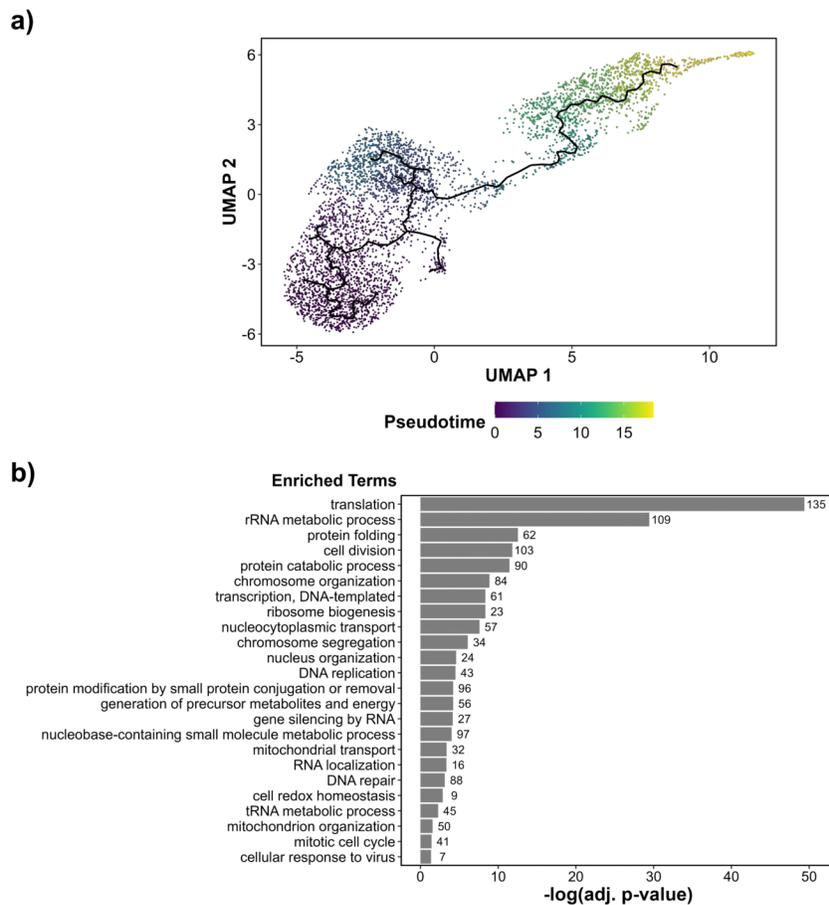
**Figure 2:** Cell clustering assessment along infection. a) UMAP of merged samples (0, 10 and 24 hpi) identifying 11 distinct clusters. b) Percentage of baculovirus mRNA molecules per cluster, across all timepoints. c) Relative expression of an early (proliferating cell nuclear antigen) and late (viral cathepsin) baculovirus gene.



**Figure 3:** Transgene expression of across cells in merged analysis. a) UMAP of cells categorized according to infection and transgene expression. Cells with a percentage of baculovirus UMIs [?] 0.1 % were deemed non-infected. Infected cells were distinguished between no transgene expression (relative expression of *gfp*, *rep* and *cap* [?] 0), *gfp* expression (*gfp* > 0 and *rep* and *cap* [?] 0), *rep/cap* expression (*gfp* [?] 0 and *rep* and/or *cap* > 0), and *gfp*+ *rep/cap* expression (*gfp* > 0, and *rep* and/or *cap* > 0). b) Proportion of cells in each category. c) Correlation between *rep* and *gfp* d) Correlation between *rep* and *cap* expression. Pearson correlation coefficient and p-value is shown on top left corner of the plot area.



**Figure 4:** Identification of host cell genes driving response to baculovirus infection (assessed through single sample analysis of 10 hpi, unmerged timepoint). a) UMAP highlighting two major cluster groups (Clusters 0, 1, 2 and Clusters 3, 4, 6, 8). b) Baculovirus mRNA molecules per cell for each cluster. c) Relative expression of 3 examples of differentially expressed genes (*heat shock protein 68*, *L-lactate dehydrogenase* and *innexin 3*).



**Figure 5:** Trajectory analysis and associated enriched gene ontology (GO) terms for the 10 hpi timepoint (unmerged dataset) (a) Trajectory analysis drawn from non-infected to infected cells. b) Enriched gene ontology terms (Biological Processes – BP) identified from variable genes along pseudotime.