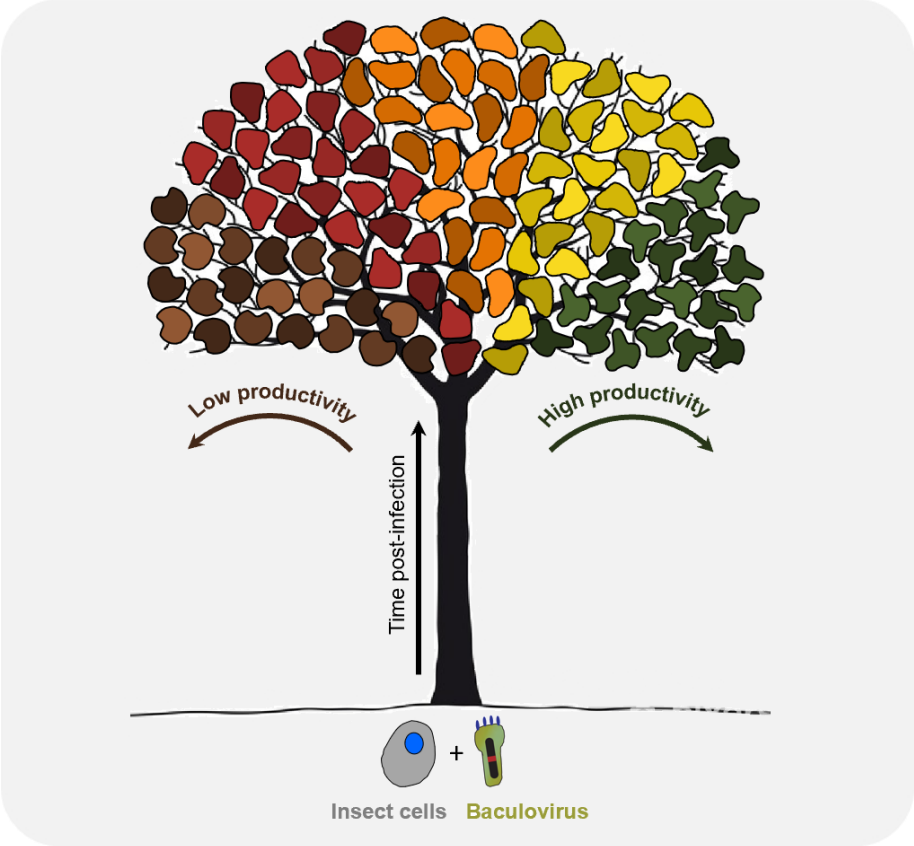
2023

Single-cell RNA sequencing for deciphering key biological mechanisms of insect cells during influenza HA-VLPs production

Marco Silvano



Dissertation presented to obtain the **Ph.D degree in Molecular Biosciences**

Oeiras, April, 2023

Single-cell RNA sequencing for deciphering key biological mechanisms of insect cells during influenza HA-VLPs production

Marco Silvano

Dissertation presented to obtain the Ph.D degree in Molecular Biosciences

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras, April 2023

Single-cell RNA sequencing for deciphering key biological mechanisms of insect cells during influenza HA-VLPs production

Marco Silvano

The work developed in this thesis was supervised by:

- **Professor Paula Alves**, PhD, Principal Investigator at Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa (ITQB NOVA) and Instituto de Biologia Experimental e Tecnológica (iBET), Invited Associate Professor at FCT-NOVA, and CEO of iBET (Supervisor)
- **António Roldão,** PhD, Head of Cell-based Vaccine Development Lab, Animal Cell Technology Unit (ACT Unit), ITQB NOVA and iBET (Co-Supervisor)

From left to right: Dr. Gabriel António Amaro Monteiro, Dr. Paula Alves, Dr. Francisca Sarreira Simões Horta Monteiro, Dr. António Roldão, Dr. Marco Silvano, Dr. Inês Isidro, Dr. Cristina Silva Pereira. On the screen: Dr. Niall Barron, Dr. Veronique Chotteau

Supervisors

- **Paula Alves**, PhD, Principal Investigator at ITQB NOVA and iBET, Invited Associate Professor at FCT-NOVA, and CEO of iBET, Oeiras, Portugal
- **António Roldão,** PhD, Head of Cell-based Vaccine Development Lab, ACT Unit, ITQB NOVA and iBET, Oeiras, Portugal

Jury

- **Veronique Chotteau**, PhD, Associate Professor and Principal Investigator of the Cell Technology Group at KTH Royal Institute of Technology and Director of VINNOVA Competence Centre for Advanced BioProduction by Continuous Processing, AdBIOPRO, Stockholm, Sweden.
- **Gabriel António Amaro Monteiro**, PhD, Associated Professor at Instituto Superior Técnico, Lisbon, Portugal
- **Niall Barron**, PhD, Professor of Biochemical Engineering in the School of Chemical and Bioprocess Engineering in University College Dublin and Principal Investigator at the National Institute for Bioprocessing Research & Training, Dublin, Ireland
- **Francisca Sarreira Simões Horta Monteiro**, PhD, Postdoctoral Fellow at University of Lisbon, Center for Ecology, Evolution and Environmental Changes, Lisbon, Portugal

Financial support from:

Marie Sklodowska-Curie Action of Horizon2020

PhD fellowship granted by STACCATO European Industrial Doctorate (grant agreement No 813453).

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 813453, and from Fundação para a Ciência e a Tecnologia/Ministério da Ciência, 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).

Single-cell RNA sequencing for deciphering key biological mechanisms of insect cells during influenza HA-VLPs production

Instituto Tecnologia Química e Biológica António Xavier

Universidade Nova de Lisboa

To all those who are driven by curiosity to know and to experiment, who are animated by insatiable doubts and who care more about the desire to listen openly and to compare rather than unshakable certainties

Vittorio Bo

Presidente Codice Idee per la cultura

Acknowledgements

The thesis comes with an important personal and professional growth, I would like to thank those who have supported me in these years of work.

To Prof. Paula Alves, my supervisor, for providing me the best environment to carry out my studies. Her leadership and experience are admirable, her words of great teaching.

To Dr. António Roldão, my co-supervisor and mentor. He was the one who most shaped me along this path. His dedication is example to me, I thank him for pushing me every time to do and give more. I learned a lot, his supervision was the added value.

To Dr. Inês Isidro, for inspiring and guiding me during this journey. The constructive discussions, the insights, the ideas provided have been of great help.

To Prof. Manuel Carrondo, for inspiring me through his unparallel experience and dedication in making iBET a centre of excellence.

Many thanks to my thesis committee members, Dr. Monique van Oers and Dr. Colin Clarke, for the fruitful discussions and guidance to best directions and better achievements.

To all my current and former colleagues from the Cell-based Vaccines Development Lab, I want to thank you for the constant share of ideas and feedback.

A special mention to Nikolaus, colleague and good friend: I am grateful to have shared this experience with you, it was special. I learned a lot from you.

To Dr. Ricardo Correia, for his work as a tutor and his constant availability to listen and help. His dedication and his support were fundamental to achieving this thesis.

To my lifelong friends, Team Animali and in particular Camera 24+1. Thank you our friendship throughout these years, for being the best friends one could have.

To Madeline for instilling me her immense strength. Thank you for being there, and for all the time you have dedicated to me.

To Valeria for her great patience. I thank God for giving me a sister like you. I can always count on her, you will always be in my life.

I infinitely thank my parents who have always supported me in the realization of my projects. I will never stop thanking you for allowing me to get here.

Abstract

The use of the Insect Cell - Baculovirus Expression Vector System (IC-BEVS) as alternative to other cell-based expression systems has increased thanks to its continued development, both as research tool and manufacturing platform. However, titers achieved are occasionally low, thus bioprocess engineering is critical to improve process performance. The comprehensive knowledge of the host cell transcriptome, mainly during infection phase, can provide clues about which molecular signatures are correlated with highly producing cells, ultimately aiding the development of rational cell line optimization strategies.

In this thesis, a thorough transcriptome analysis and characterization of the IC-BEVS system was done to decipher key biological mechanisms of insect High Five cells during influenza virus-like particles (VLPs) production, highlighting the physiological determinants that contribute to increased productivity.

In **Chapter 1**, the state of the art on existing 'omics strategies to improve cell line performance is detailed, with a particular focus on bulk and single-cell RNA-sequencing (scRNA-seq) technologies. IC-BEVS and one of the most studied biologics derived from it, i.e. influenza virus-like particles (VLPs), are introduced, and the transcriptomics tools used to date to investigate the insect cell - baculovirus interaction highlighted.

In **Chapter 2**, bulk RNA-seq was used to study the gene expression profile of a novel, high-producer insect High Five cell line established in our lab by adaption to neutral pH (and compare it to that of its non-adapted counterpart) during expression of influenza VLPs using BEVS, aiming to find the biological mechanisms associated with increased productivity of adapted cells. All differentially expressed baculovirus genes were found down-regulated in adapted cells (compared to non-adapted cells), and pathways such as ribosome biosynthesis, carbohydrate and lipid metabolism were found enriched. This

study may assist the identification of potential engineering targets to increase recombinant protein production.

The outcome of virus-based cell culture processes is governed by the complex, dynamic interplay between viruses and the host cell population. Today, it is clear that cell populations are highly heterogeneous, but little is known about how such heterogeneity impacts product expression.

In **Chapter 3**, single-cell RNA-seq was used to provide insights on heterogeneity of insect High Five cells during production of influenza haemagglutinin (HA)-VLPs using BEVS. Cell heterogeneity was observed before baculovirus infection, and it was found further amplified upon infection as consequence of viral DNA expression and cell cycle arrest. Timing and level of baculovirus gene expression was shown to drive the clustering of infected insect cells, and the relative expression of VLP transgenes (M1 and HA) were found to be comparable regardless of the cell cluster. Specific pathways (e.g., endocytosis, protein sorting, folding and degradation) were identified as being those which vary the most during insect cell infection and production of influenza HA-VLPs. This study lays the ground for future application of scRNA-seq in IC-BEVS, with identification of pathways involved in energy metabolism (i.e., oxidative phosphorylation and citrate acid cycle) and protein folding, sorting and degradation (e.g., protein processing in endoplasmic reticulum) as potential target for rational cell and process engineering towards improved HA-VLP production.

In **Chapter 4**, the findings of the previous chapters are summarized and discussed. Special emphasis is given to: (i) the benefits of correctly annotating non-model organism genomes, such as those of insect cells, as a critical step to better decipher complex biological mechanisms; (ii) the challenging task of using transcriptomics data to improve bioprocesses; and (iii) the need to further

investigate cell-to-cell heterogeneity in order to design strategies to mitigate it and thus improve process robustness and reproducibility.

In summary, this thesis contributes to extend the knowledge on the impact of baculovirus infection on insect (High Five) cells, in particular at the gene expression level, providing new insights for rational cell/virus and process engineering towards improved VLP production.

Resumo

A utilização de células de insecto com o sistema de expressão por vectores de baculovirus (IC-BEVS) para a expressão de proteínas tem vindo a aumentar como alternativa a outros sistemas de expressão, graças ao seu desenvolvimento continuado, e tem sido bastante utilizado como ferramenta de investigação e como plataforma de produção. No entanto, a produtividade deste sistema IC-BEVS ainda é considerada moderada em alguns casos. Assim, a engenharia do bioprocesso é um fator crítico a ser otimizado de modo a aumentar a sua performance. O estudo compreensivo do transcriptoma das células de insecto, em particular durante a fase de infeção com os baculovirus, poderá fornecer pistas sobre quais os fenótipos associados a uma maior produtividade, ajudando ao desenvolvimento racional de linhas celulares otimizadas.

Nesta tese, é realizado uma minuciosa análise e caracterização do transcriptoma do sistema IC-BEVS para encontrar e compreender mecanismos biológicos chave durante a produção de "virus-like particles (VLPs)" de influenza em células de insecto High Five. Esta análise identifica mecanismos fisiológicos determinantes para a produtividade de VLPs, bem como os requisitos para uma elevada qualidade do produto.

No **Capítulo 1**, é descrito o estado da arte das várias estratégias de transcriptómica utilizadas para melhorar a performance de linhas celulares, focando-se em particular nas tecnologias de sequenciação de RNA, em massa (bulk RNA-seq) e a nível unicelular (scRNA-seq). O sistema IC-BEVS e as VLPs (um dos produtos biológicos mais produzidos neste sistema), são descritos de uma forma geral, assim como as ferramentas de transcriptómica utilizadas para investigar a interação entre baculovírus e células de insecto.

No **Capítulo 2**, foi utilizada a técnica de bulk RNA-seq para estudar os padrões de expressão genética desta nova linha celular, em comparação com a linha celular original High Five que não se encontra adaptada a valores de pH neutros. Esta comparação dos padrões de expressão genética foi realizada durante a produção de VLPs de influenza, utilizando o sistema IC-BEVS, tendo como objetivo identificar os mecanismos biológicos associados a uma maior produtividade das células adaptadas aos valores de pH neutro. Todos os genes de baculovirus com uma expressão diferenciada encontravam-se sub-expressados nas células adaptadas (em comparação com as células não adaptadas ao pH neutro), enquanto que vias biológicas como a biossíntese de ribossomas, o metabolismo de hidratos de carbono e de lípidos estavam significativamente expressas diferencialmente. Este estudo identifica potenciais alvos a serem modificados e otimizados de modo a contribuir para um aumento da produção de proteínas recombinantes.

O resultado final da produção de proteínas associada à cultura celular e sua infeção utilizando vetores virais é diretamente influenciado pela complexa dinâmica de interações entre os vírus e a respetiva população de células alvo. Atualmente, é claro que as populações celulares são bastante heterogéneas. No entanto, o impacto desta heterogeneidade na produtividade é pouco estudado.

No **Capítulo 3**, a técnica de scRNA-seq é utilizada para explorar e avaliar o impacto da heterogeneidade da população de células High Five na produção de VLPs de influenza através da utilização do sistema IC-BEVS. A heterogeneidade celular observada antes da infeção das células deve-se sobretudo às diferentes fases do ciclo celular. Esta heterogeneidade aumentou ainda mais após a infeção com os baculovirus, como consequência da expressão do DNA viral e da resultante paragem do ciclo celular. O momento de infeção e o nível da expressão dos genes de baculovirus foram identificados como os parâmetros que levaram à diferenciação e agrupamento das células infetadas, sendo os

níveis de expressão dos transgenes das VLPs (M1 e HA) comparáveis, independentemente do grupo de células. Vias biológicas específicas (como por exemplo a endocitose, a formação e degradação de proteínas) foram identificadas como aquelas que apresentam uma maior variabilidade durante a infeção das células de inseto e produção das VLPs de influenza. Este estudo estabelece as bases para futuras aplicações de scRNA-seq no sistema IC-BEVS, identificando as vias biológicas envolvidas no metabolismo (como por exemplo a fosforilação oxidativa e o ciclo dos ácidos tricarboxílicos) e na formação da estrutura de proteínas e sua degradação (como por exemplo o processamento proteico no retículo endoplasmático) como potenciais alvos para um desenvolvimento racional de um processo melhorado de produção de VLP.

No **Capítulo 4** é realizado um resumo e discussão das principais conclusões dos capítulos anteriores. A discussão coloca particularmente em foco: (i) os benefícios de uma correta anotação de genomas de organismos não-modelo, como as células de inseto, sendo este um passo crítico para identificar e explorar mecanismo biológicos complexos; (ii) a tarefa desafiante de utilizar dados de transcriptómica para melhorar processos biológicos; e (iii) a necessidade de uma investigação mais profunda sobre a heterogeneidade das populações celulares, definindo estratégias para a mitigar de modo a melhorar a robustez do processo de produção e aumentar a sua reprodutibilidade.

Em resumo, esta tese contribui para um mais extenso e profundo conhecimento do impacto da infeção de baculovirus em células de insecto (High Five), particularmente ao nível da expressão genética, revelando novo conhecimento útil para o desenvolveemnto e optimização de processos biológicos que permitam uma produção de VLPs mais eficiente.

Thesis Publications

Silvano, M., Correia, R., Virgolini, N., Clarke, C., Alves P.M., Isidro I.A., Roldão, A. 2022. Gene Expression Analysis of Adapted Insect Cells during Influenza VLP Production Using RNA-Sequencing. Viruses. 14, 2238.

doi.org/10.3390/v14102238

Silvano, M., Virgolini, N., Correia, R., Clarke, C., Isidro I.A., Alves P.M., Roldão, A. (2023). Dissecting insect cell heterogeneity during influenza VLP production using single-cell transcriptomics. Frontiers in Biotechnology and Bioengineering.

doi: 10.3389/fbioe.2023.1143255

Additional Publications

Virgolini, N., Hagan, R., Correia, R., **Silvano, M.**, Fernandes, S., Alves, P.M., Clarke, C., Roldão, A., Isidro, I.A. 2022. Transcriptome analysis of Sf9 insect cells during production of recombinant Adeno-associated virus. Biotechnology Journal.

doi:10.1002/biot.202200466

Virgolini, N., **Silvano, M.**, Hagan, R., Correia, R., Alves, P.M., Clarke, C., Roldão, A., Isidro, I.A. 2022. Impact of dual baculovirus Infection on the Sf9 Insect cell transcriptome during AAV production using single-cell RNA-seq.

doi: 10.22541/au.167121909.92795915/v1

Table of Contents

List of Abbreviations

AAV: adeno-associated virus

***Ac*MNPV:** *Autographa californica* multiple nucleopolyhedrovirus

ALE: adaptive laboratory evolution

BmNPV: *Bombyx mori* nuclear polyhedrosis virus

CCI: cell concentration at infection

CHO: Chinese Hamster Ovary

DIP: defective interfering particles

DNA: desoxyribonucleic acid

ER: endoplasmic reticulum

GOI: gene of interest

HA: influenza hemagglutinin protein

HPI: hours post-infection

IC-BEVS: insect cell - baculovirus expression vector system

KEGG: Kyoto encyclopedia of genes and genomes

mRNA: messenger RNA

MOI: multiplicity of infection

NA: influenza neuraminidase protein

NGS: next-generation sequencing

PCA: principal component analysis

PCR: polymerase chain reaction

rBAC: recombinant baculovirus

RNA: ribonucelic acid

scRNA-seq: single-cell RNA-sequencing

TCA: tricarboxylic acid cycle

UMAP: uniform manifold approximation and projection for dimension reduction

UMI: unique molecular identifier

VCP: valosin-containing protein

VLP: virus-like particle

WTA: whole transcriptome analysis

Chapter 1

Introduction

Author contribution

Marco Silvano wrote the chapter based on the referred bibliography

TABLE OF CONTENTS

1. Bioprocess engineering strategies for recombinant protein production

Animal cells are the preferred “cell factories” for the production of recombinant proteins for prophylactic, therapeutic or diagnostic applications (Butler, 2005). Recent developments in bioprocessing engineering have increased production yields while cutting down costs, allowing a number of processes to reach industrial scale and thus opening the door for the treatment of multiple diseases and disorders (O’Flaherty et al., 2020). Nonetheless, animal cell culture processes capable of delivering sufficient quantities of high-quality recombinant biologics are still limited, being one of the major concerns of the biopharmaceutical industry (Hunter et al., 2019).

Different factors can play a role on the performance of an expression system. Productivity and/or quality of the product can be enhanced through a bioprocess approach (e.g. fine-tune culture and expression conditions, tailor-made supplementation regimens, optimization of bioreactor setups and designs, process intensification and integration) or a biological approach (e.g. engineering of cells and expression vectors), for example. Regarding the latter, classical methods to obtain high-producer cell clones have not changed significantly in the past two decades. These include gene selection and amplification techniques where the primary focus is on generating stable producer cell clones in shorter timeframes rather than on improvements in cell specific production rates (O’Callaghan & James, 2008). Most of the efforts undertaken to increase specific production rates have relied on empirical experience and hypothesis-driven studies rather than on comprehensive knowledge of how cell functions when in culture (Carvalhal et al., 2011; Shen & Kamen, 2012). Establishment of such knowledge would be key for rational cell engineering, ultimately streamlining process development (T. Lai et al., 2013).

With progress in the 'omics technologies, the molecular traits associated with high-producer cells are coming to light. Systems biology has played an important role in organizing the data generated from these large-scale cell state characterizations, allowing cell physiology to be fully characterized.

1.1. Application of 'omics for process development

Traditionally, process development relies on empirical, labour-intensive and time-consuming optimization activities (Sharfstein, 2008). While improvements have been attained, little fundamental understanding on how or why specific conditions have led to improvements can be extracted (Courtes et al., 2013; Selvarasu et al., 2012). It is then evident that future progress will require a transition from purely empirical approaches to knowledge-based ones (Schaub et al., 2010, 2012; Wuest et al., 2012; Young, 2013). ‘Omics technologies may assist this paradigm shift, providing comprehensive information on how molecular changes shape cellular responses (Lewis et al., 2016). Four primary ‘omics areas have emerged, namely genomics, transcriptomics, proteomics, and metabolomics, which respectively examine genes, messenger RNA (mRNA), proteins, and metabolites present in a particular cellular environment. In addition to these, researchers have also applied lipidomics (Budge et al., 2020; Yeo et al., 2018; Y. Zhang et al., 2017), epigenomics (Feichtinger et al., 2016), and glycomics (Sumit et al., 2019) to investigate host cell biology and reveal novel engineering targets. Importantly, none of these ‘omics tools alone provide a complete description of complex biological systems, thus an integration of 'omics data is essential to comprehensively understand an organism at molecular level (J. H. Wang et al., 2010) and link genotype to phenotype (Subramanian et al., 2020).

The application of 'omics technologies to optimize animal cell culture processes has lagged the development witnessed for industrial microorganisms (Otero & Nielsen, 2010; Stephanopoulos et al., 2004). This is mostly because of the unavailability of databases and/or complete genome sequences and annotations for biotechnologically relevant animal cell lines. Accumulated knowledge on pathway and enzymatic is also scarce. Nonetheless, the use of ‘omics technologies continues to gain interest amongst those working with mammalian cells in academic and industrial communities, leading to the maturation (i.e. quality and quantity) of available data (Samoudi et al., 2021).

In comparison to metabolites and proteins, genes are less chemically heterogeneous and therefore it is analytically less challenging to perform genomics and transcriptomics when compared to proteomics and metabolomics (Aizat et al., 2018). In addition, revolutionary developments in sequencing instruments have helped to acquire genome and transcriptome data.

Metabolomics is often the preferred tool for rational media design, as depletion of key nutrients can be measured directly. Metabolite profiling of a Chinese Hamster Ovary (CHO)-GS IgG4-producer was conducted during a batch campaign in chemically defined media (Sellick et al., 2011). By assessing both intra and extracellular metabolites, researchers identified several media components depleted by early stationary phase including aspartate, asparagine, glutamate, and pyruvate. This information was used to design a simple feed strategy that increased cell biomass 35% and antibody titer by 75%.

Next to process improvements through media formulation, 'omics experiments can often suggest promising genetic targets for improved process characteristics. Transcriptome analysis helped to identify 32 genes that were consistently upregulated in high producing CHO cell clones (Berger et al., 2020), suggesting that these genes play a role in mediating high productivity. An

ontology analysis revealed that these candidate genes were predominantly involved in signalling, protein folding, cytoskeleton organization, and cell survival functions. Several targets showed beneficial effects, most notably Foxa1 that upon overexpression resulted in increased cell density, viability, and yield of difficult-to-express proteins (Berger et al., 2020). Research conducted at Pfizer compared four mAb-secreting cell lines, with two considered fast growing, and two slow growing in a batch model (Doolan et al., 2010). Transcriptome and proteome profiling identified differentially expressed genes correlated with growth rate. Collectively, 21 genetic targets with differential expression at both the mRNA and protein levels were identified. In order to validate findings, researchers selected five targets for siRNA knockdown, including valosin-containing protein (VCP). siRNA knockdown of VCP resulted in more than 40% reduced cell viability by day 3, and transient over-expression resulted in 1.2–2.1-fold improved cell growth by day 5.

These examples demonstrate how ‘omics data can identify critical genetic targets, and modifications to those targets can result in improved cellular phenotypes. There are only a handful of studies demonstrating rational bioprocess improvement using ‘omics technology, nonetheless, cases continue to emerge and demonstrate that a rational approach has both time and resource saving benefits.

2. Transcriptome analysis using next-generation sequencing

Expression of mRNA has been traditionally measured using microarrays or real-time PCR. While the first method does not have much sensitivity, the latter can't be used for a genome-wide survey of gene expression (Mardis, 2008). Currently, RNA-seq with next-generation sequencing (NGS) is the preferred method for studying the transcriptome, capable of measuring the expression levels of thousands of genes simultaneously and providing insights into functional pathways and the regulatory networks in biological systems (Levy & Myers, 2016). In addition, RNA-seq can provide novel insights into alternative splicing (Schreiber et al., 2015), unannotated exons and novel transcripts (Mutz et al., 2013).

Several sequencing platforms are available today, all of them use short fragments (so-called reads) to investigate genome sequences (Korbel et al., 2007), chromatin immunoprecipitation or mapping of DNA methylation (Mardis, 2008). NGS analyses millions of short DNA fragments during one sequencing run. The read lengths of those fragments depend on the type of NGS platform and can be in the range of 25 – 450 base pairs (Mutz et al., 2013).

The analysis pipeline of RNA-Seq consists of four fundamental analysis steps, providing that an already sequenced reference genome or transcriptome is available for the reviewed organism (Figure 1).

Figure 1 General workflow for carrying out RNA-seq experiment.

In Step 1, raw image data has to be converted into short read sequences. The raw data generated by NGS consists of fluorescence signals, which have to be converted into base sequences by platform specific base calling-algorithms provided by the manufacturer. Additionally, a quality score for each base is calculated, which indicates the reliability of each base call. While the output of all NGS platforms is stored in the standard FASTQ format, the calculation of quality score differs from manufacturer to manufacturer.

In Step 2, short read sequences are aligned to the reference genome or transcriptome. Read mapping algorithms require reference sequences to which they can align the short reads. Read mapping algorithms in contrast to conventional alignment algorithms use indexing strategies, which enable them to align millions of short reads in an appropriate period of time. As mapping tools have to balance between sensitivity and speed, no aligner can be best suited for all applications (P. J. Park, 2009). Most mapping tools allow few mismatches during the sequence alignment considering sequencing errors, single nucleotide polymorphisms or mutations but they do not allow any large gaps. The exact number of allowed mismatches has to be chosen depending on the read length (Cullum et al., 2011; P. J. Park, 2009). As this length is determined by the used sequencing platform, a certain read mapping tool is never suitable for all sequencing technologies. Moreover, every biological application demands for special mapping algorithms: for RNA-seq analysis of eukaryotes, cDNA reads are preferably aligned to a reference transcriptome. Since these data are unfortunately rarely available, researchers have to use the genomic sequences instead. This requires spliced read mapping software, which considers the genomic intron–exon structure by splitting unmapped reads and aligning the read fragments independently (Pagani et al., 2012).

In Step 3, the amount of mapped reads is counted and the gene expression level is calculated by peak calling algorithms. To estimate the gene expression level

using RNA-seq, reads, which are mapped to a particular gene, have to be quantified. Therefore, bioinformatic tools count the number of reads in a window of defined size. By moving the window along the whole sequence, an expression profile is generated. Subsequently these expression scores have to be normalized because of inherent bias in read quantification: on one hand the gene length influences the number of reads mapped to it, and on the other hand this number also depends on the sequencing depth (total number of sequenced reads) (P. J. Park, 2009). Normalization of read counts enables the comparison of expression level between different genes as well as different experiments.

In Step 4, differential gene expression is determined using statistical tests. Many methods have been developed so far, their selection being user-dependent. Moreover, many comparison studies highlight that no single method outperforms others in all circumstances (Rapaport et al., 2013; Seyednasrollah et al., 2013; Soneson & Delorenzi, 2013; Z. H. Zhang et al., 2014). However, it is recommended the use of tools based on the negative binomial distribution (Gierliński et al., 2015). Examples of these tools include edgeR, DESeq, DESeq2, Cuffdiff, Cuffdiff 2, and baySeq. Some non-parametric methods can also be used as alternatives when the data does not seem to fit the negative binomial law; these methods are less often used and usually require a higher replicate number to perform equally well (Spies & Ciaudo, 2015).

During the past years several algorithms have been developed for each analysis step and afterwards adapted to specific applications. Today researchers can combine a variety of bioinformatic tools to obtain an appropriate analysis system optimized to their requirements (Pagani et al., 2012). One of the most commonly used open-source repositories of bioinformatics tools used in transcriptomics is the Bioconductor project (Gentleman et al., 2004; Huber et al., 2015). Bioconductor tools are written in the R statistical programming language and are freely available to download, install, and modify through an

open-source and open-development model supported by the use of the GitHub software management system. Several packages are currently available under Bioconductor, with the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway being one of the most used ones for understanding biological pathways and functions of cellular processes (Kanehisa, 2019).

1.1. Single-cell RNA sequencing technologies

Although bulk RNA-seq is widely used for transcriptome analysis, it encloses shortcomings since it only provides an average view of the expression of thousands of cells potentially ignoring important features such as host gene expression differences between infected and uninfected cells in virus-based processes. Hence, clustering cells into groups on the basis of their individual gene expression levels is an important challenge that can be approached using single-cell RNA-sequencing (scRNA-seq) (Kolodziejczyk et al., 2015). Notably, current bioprocess analytics (cell growth, productivity, metabolic rates, etc.) result in average values neglecting the suspected heterogeneity within the cell culture, thus often leading to unexpected outcomes or even process failure (Figure 2).

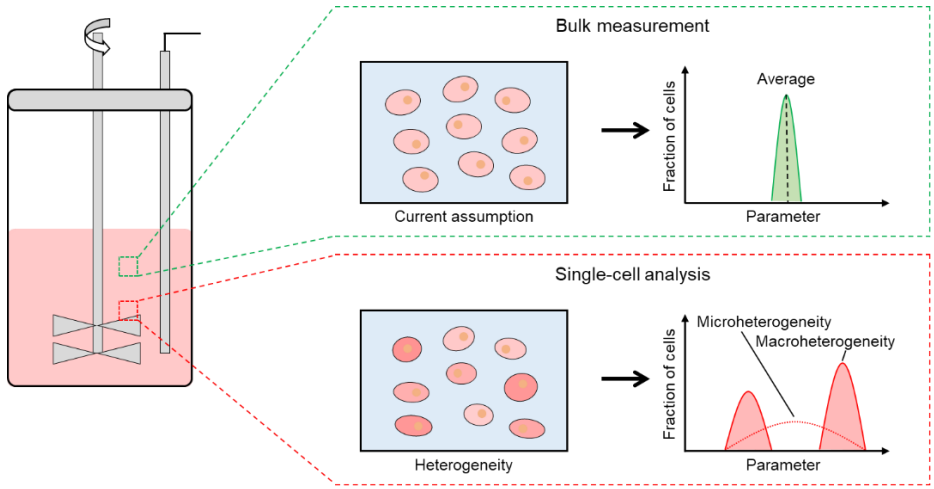


Figure 2 Bulk Measurement versus Single-Cell Analysis. Comparison of traditional average bulk measurements and single-cell resolution analysis contrasts the anticipated homogeneous behavior of an isogenic mammalian population with its real heterogeneous status.

The first scRNAseq protocol was published in 2009 (F. Tang et al., 2009), and since then many protocols and commercial platforms have been released (Svensson et al., 2017; Ziegenhain et al., 2017). The most common scRNA-seq approach is to use microscopic droplets or wells to isolate a large number of cells and then sequence the libraries relatively shallowly (J. Cao et al., 2019; Han et al., 2018). In order to identify from which cell a given transcript originates, these methods use cellular barcodes (short nucleotide tags attached to each read that are unique to a droplet or well). This high-throughput, low-depth model is typical for experiments using the popular 10× Chromium™ platform and BD Rhapsody™ Single-Cell Analysis system. An important advantage of this technology is that it supports unique molecular identifiers (UMIs). UMIs are short barcodes that are attached to transcripts before amplification, making it possible to remove polymerase chain reaction duplicates and to obtain more accurate estimates of expression levels. A major shortcoming of this approach is that platforms only allow for the 5′ or 3′ end of each mRNA to be sequenced.

Recent innovations in scRNA-seq technologies have eased the experimental workflow and thus provided a substantial reduction in cost per sample. Consequently, there has been an exponential growth in terms of the number of cells profiled (Svensson et al., 2017). Given the large volumes of data generated, efficient computational and statistical methods are required for single-cell data analysis. A central component of scRNA-seq analysis is the expression matrix, which represents the number of transcripts observed for each gene and cell. A strategy for reducing the negative effect of the high dimensionality of the expression matrix is to perform dimensionality reduction on the feature space. There are many methods available (S. Sun et al., 2019), but the most commonly used one involves principal component analysis (PCA), a linear transformation that preserves Euclidean distances between cells in the full PCA space, with the number of components retained for later analysis depending on the complexity of the dataset. When scRNA-seq datasets are so complex that their structure cannot be captured by two or three principal components, visualization algorithms can be used to create a two-dimensional plot summarizing an scRNA-seq dataset from a larger number of significant components. The current best-practice method is Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) (McInnes et al., 2018). This algorithm approximates the topology of the data using a cell–cell nearest-neighbour network and then estimates a lowdimensional embedding of the data that best preserves the structure.

Another key component of scRNA-seq analysis is cell clustering, the application of which depends on a case-by-case basis. For example, if the dataset represents a developmental process or is derived from a time-course experiment, then it is more appropriate to view the cells as drawn from a continuum. Such a continuous trajectory, which could represent spatial positions, chemical concentrations or time courses, is often referred to as ‘pseudotime’ with each

cell being assigned to a specific position. Most tools cannot determine the direction or speed that the cells are moving along the trajectory; this must be inferred using external information such as sampling time (for time-course experiments) or marker genes (for developmental trajectories). Two main approaches are used to perform 'pseudotime' analysis. The first approach is to use dimensionality reduction techniques to identify a low-dimensional 'manifold' that the cells lie upon and then use a cell–cell graph to describe the topology of the manifold. Popular methods using this strategy include Monocle (J. Cao et al., 2019) and DPT (Haghverdi et al., 2016). The second approach is to use unsupervised clustering to group cells before linking the clusters and projecting individual cells onto the branches (J. Chen et al., 2016; Ji & Ji, 2016). Cluster-based methods tend to be more accurate when there is an unequal density of cells through the trajectory. On the other hand, manifold-based methods perform best when there is an even sampling of cells across the transition and when examining details of singular transitions.

1.2. Heterogeneity on virus-based process

Single cell RNA-seq analysis has unambiguously demonstrated that biology is highly heterogeneous at the single-cell level (Finn et al., 2019; Spitzer & Nolan, 2016). In 2013, McWilliam Leitch and McLauchlan published the first scRNA-seq work in virology, aiming at exploring the heterogeneity of Hepatitis C virus quasi-species and assessing their fitness (McWilliam Leitch & McLauchlan, 2013). Since then, many studies have been published on the topic. For example, in a recent study on Human Immunodeficiency Virus, scRNA-Seq was used to identify the transcriptional signature of specific cell subsets, such as dendritic cells from elite controllers, highly permissive CD4+ T cells or inducible latently infected CD4+ T cells (Golumbeanu et al., 2018; Rato et al., 2017).

Single-cell RNA-seq has been also applied to influenza A virus (IAV) infection processes, aiming to understand if the high cell-to-cell variability in the production of infectious progeny as well as viral transcripts per cell can be explained by the presence of defective interfering particles (DIPs) in the inoculum (Kupke et al., 2020; C. Wang et al., 2020). A recent study shown substantial variability in viral transcription between human A549 cells and primary bronchial epithelial cells infected with IAV, including the accumulation of defective viral genomes that impact viral replication (C. Wang et al., 2020), ultimately demonstrating the intricate effects of defective viral genomes on host transcriptional responses and highlighting the importance of capturing host-virus interactions at the single-cell level.

Studies aim to dissect the interactions between influenza virus and host cell heterogeneity reveal the fraction of viral mRNAs compared to total mRNA in individual cells was anything from < 0.1% to over 50% at 8 hpi (Russell et al., 2018), or from <1% to about 90% at 16 hpi (J. Sun et al., 2020). Furthermore, Russell et al., 2018 found that less than 10% of influenza virus-infected cells contained over half of the scored viral transcripts at 8 hpi, thus underlying the importance of single-cell analyses over bulk assays.

As the aforementioned studies highlight, the infection process in a virus-based system encloses a significant degree of variability resulting in specific cellular environments that shape the generation of new viruses. Two non-mutually exclusive hypotheses can explain this: (i) virus-to-virus heterogeneity, i.e. mixture of viral particles displaying differences in infection ability (e.g. DIPs); and (ii) cell-to-cell heterogeneity, i.e. mixture of cells at different metabolic and or cell cycle states, for example. Single-cell analyses represent novel opportunities to identify specific cellular and molecular features promoting or restricting virus replication, thereby adding to the understanding of virus–host interactions and providing new targets to inhibit viral replication. In this context, scRNA-seq is a

powerful technology for virus-based processes optimization, the IC-BEVS being one example. This expression system, although extensively explored for production of recombinant proteins, titers (and cell specific productivities) still require improvement.

3. The insect-cell baculovirus expression system

Insect cells are widely used for the expression of recombinant proteins, the most common cell line being *Sf*9 (S2, *Sf*21, Tn-368, and High-Five™ cells are also used) (Contreras-Gómez et al., 2014). Insect cell lines can promote mammalian-like protein translational modification, such as glycosylation, phosphorylation, disulphide bond formation and further protein folding and processing. Despite not being able to synthesize mammalian-specific complex glycan structures such as processing of N-glycans to terminally sialylated complex-type structures by default, modified cells lines have been established for this end (van Oers et al., 2015).

The transient IC-BEVS allows to achieve considerably high production yields. It is based on the baculovirus propensity to infect insect cells, using this virus as vector for foreign gene expression; the most common being the *Autographa californica* multiple nucleopolyhedrovirus (*Ac*MNPV). This vector can be genetically modified to include the gene(s) coding for the protein(s) of interest in replacement of genes that are non-essential for virus' in vitro infection and replication. Most common recombinant baculovirus (rBAC) constructs involve replacement of the polyhedrin or p10 genes whose promoters allow very high gene expression levels (van Oers et al., 2015); other less conventional promoters have been explored, e.g. *IE1*, *39k* and *vp39* (Grose et al., 2021). A single rBAC vector can be used for heterologous production of one or multiple proteins of interest, enabling the generation of multi-component particle structures such

as VLPs (Mi et al., 2021). Different antigens of interest (targeting different diseases or different strains of the same pathogen) can be displayed on the surface of the same particle resulting in multivalent VLPs (Shima et al., 2016). This can be achieved by (i) infecting insect cells with a rBAC comprising multiple genes of interest (GOI) or (ii) co-infecting insect cells with as many rBAC as needed, each coding for one of the GOI. Stable insect cell lines have also been established for production of recombinant proteins; however, these often offer low productivities (Vidigal et al., 2018). Stable insect cell lines and IC-BEVS can be combined to produce VLPs with higher valency, overcoming the instability of rBAC derived from including high number of foreign GOI (Sequeira et al., 2018). Since they are unable to replicate and propagate within a human host, baculoviruses are seen as attractive and safer alternatives to classical mammalian-derived viruses. In addition, the concept of using the baculovirus itself as a vectored vaccine has been explored, further evidencing the plasticity of such vectors is serving a myriad of biotechnological and therapeutic applications (Fabre et al., 2019).

The IC-BEVS has been employed to produce VLPs as vaccine candidates against infection by influenza (Pushko et al., 2005), Ebola (Ye et al., 2006), Chikungunya (Metz et al., 2013), rabies (Bernardino et al., 2021), SARS-CoV-2 (Naskalska et al., 2021), MERS (C. Wang et al., 2017), human papillomavirus (Touzé et al., 1996), amongst many others. To date, four drug products produced using IC-BEVS were approved for human use: two vaccines (Cervarix®, GSK, against human papillomavirus; Flublok®, Protein Sciences, against influenza) and two therapeutics (Provenge®, Dendreon, for prostate cancer; Glybera®, uniQure, world's first gene therapy). IC-BEVS has also been used to produce seven veterinary vaccines (Cid & Bolívar, 2021; Felberbaum, 2015).

3.1. Baculovirus-host cell interaction

The baculovirus *Ac*MNPV has a large genome (~ 134 kbp) that encodes approximately 156 proteins. In addition to genes that regulate and/or mediate viral transcription, translation and DNA replication, the virus encodes a variety of genes that modify cellular physiology and architecture (Y.-R. Chen et al., 2014)

The infection process in IC-BEVS is completed in a relatively short time period (approximately 24 to 48 h), producing infectious budded viruses. Although the timing is variable and depends on experimental conditions, transcription of viral genome may be detected as early as 1 hour post infection (hpi), viral DNA replication may begin as early as 4 to 6 hpi, and viral late gene transcription begins simultaneously with or shortly after DNA replication (Schaly et al., 2021).

A number of cellular responses to baculovirus infection have been documented previously, and those that are observed by light microscopy are collectively referred to as the cytopathic effect. Examples of specific cellular responses to baculovirus infection include: (i) early rearrangement and induction of actin polymerization, (ii) triggering of apoptotic responses and DNA damage responses, (iii) modifications of the cell architecture such as expansion of nucleus, formation of virogenic stroma within the nucleus, and cell rounding, (iv) reduction of host protein synthesis, (v) plasma membrane ruffling, (vii) impact on cell cycle, and (viii) cell lysis (Monteiro et al., 2012). While many of the cellular responses to *Ac*MNPV infection have been described and documented, it is clear that these responses are complex and that our understanding of them is, at best, rudimentary.

3.1.1. The impact of baculovirus infection on cell metabolism

Viral infection claims an intensification of host cell biosynthetic activity in order to supply the building blocks needed for the biogenesis of membrane lipids and

for the synthesis of viral nucleic acids and proteins (Maynard et al., 2010). In this regard, the success of infection is highly dependent on the metabolic state of the cells at the moment of infection, jointly with the viral manipulation of energy metabolism to fit such needs (Carinhas et al., 2010, 2011)

Baculovirus infection provokes an important metabolic burden on insect cells, causing an enhancement of the fluxes through the major catabolic pathways, namely glycolysis and tricarboxylic acid cycle (TCA) (Bernal et al., 2009, 2010). A concomitant increase in the oxygen uptake rate is also observed, which accounts for a higher rate of respiration upon infection (Bernal et al., 2009; Kamen et al., 1996.; Palomares et al., 2004). Metabolic Flux Analysis was applied to quantify the fluxome distribution shifts in response to infection in High Five cells, and a comparison with *Sf*9 cells was conducted, highlighting the capacity of the virus to re-direct the cellular fluxome toward ATP production to support replication and progeny generation independently of the insect cell line (Monteiro et al., 2017).

Cell specific productivity decreases when insect cells are infected with baculovirus at high cell densities, the so called “cell density effect”. Bernal et al., 2009 succeeded in deciphering the metabolic basis of such phenomenon, during which *Sf*9 cells undergo a progressive inhibition of central metabolism. Since a successful viral infection strongly correlates with the energetic state of the cell, viral replication is impaired at high cell density cultures.

3.1.2. The impact of baculovirus infection on cell proteome

Carinhas et al., 2011 applied a SILAC approach for quantitative proteomics of *Sf*9 cells during growth and early baculovirus infection, contributing with the first comparative quantitative proteomic analysis of the response of *S. frugiperda* cells to infection. The authors found new differentially expressed proteins

related to energy metabolism, endoplasmatic reticulum and oxidative stress during *Ac*MNPV infection (Carinhas et al., 2011). In particular, the up-regulation of two metabolic enzymes, PDH-E3 and ALDH, was observed, which account for an increased efficiency of the coupling of glycolysis and the TCA cycle and for the anaplerotic feeding of carboxylic acids, respectively. These observations go along with the increased metabolic fluxes of central carbon metabolism during baculovirus infection (Bernal et al., 2009), emphasizing the importance of energetic metabolism during viral infection. Concerning the cellular stress response to infection, the authors observed a decrease in the levels of the chaperone ERp57 and the polypeptide transporter SRP57. Both proteins are effectors of the untranslated-protein response, and their downregulation envisages the baculovirus capacity in avoiding the deleterious effects of cellular stress response. Altogether, these results demonstrate that baculovirus efficient replication depends on its capacity to manipulate different cellular pathways and highlight a number of regulatory mechanisms that are driving the success of baculovirus infection.

3.1.3. The impact of baculovirus infection on cell transcriptome

Transcriptome analysis has been widely used to characterize mammalian cells responses to foreign gene expression or to stress conditions, enabling the identification of new routes to enhance production yields (Lewis et al., 2016). Its application to the IC-BEVS, however, has been held back by the lack of sequenced insect cell genomes, and the scarcity of curated databases. In fact, only five complete genome sequences exist and are publicly available, four of *Trichoplusia ni* (T. ni) cells and one of *Ac*MNPV. Nonetheless, recent studies have shown the importance of transcriptomics to assess the impact of baculovirus infection on insect cells and thus guide targeted cell and/or process engineering approaches capable of circumventing production bottlenecks. For example, Y.-

R. Chen et al., 2014 measured the expression levels of each transcript throughout a 48-h time course of *Ac*MNPV infection in *T.ni* cells. Using differential gene expression analysis, they found that the majority of host transcripts were downregulated after 6 hpi and throughout the remainder of the infection, correlating with the replication of the virus and a dramatic increase in viral transcripts. In addition, they examined the responses of genes belonging to a number of specific pathways of interest, including stress responses, apoptosis, immunity, and protein trafficking (Y.-R. Chen et al., 2014). In fact, the manipulation of genes involved in apoptosis can prolong cell survival and subsequently achieve higher product titers in IC-BEVS. The inhibition of pre-apoptotic (e.g., dronc) (Huang et al., 2013) and apoptotic (e.g., caspase-1) gene expression has already been shown, indicating slight improvements (Y. K. Lai et al., 2012; C.-C. Lin et al., 2007; Q. Wang et al., 2016; X. Zhang et al., 2018, 2021). Similarly, the strategy to delay the cell lysis was pursued by Steele et al., 2017 using cell lines constitutively expressing vankyrin or vankyrin-encoding baculovirus vectors.

In the same context, targeting the protein folding mechanisms could potentially contribute to maintain product quality throughout the infection process. The comparative transcriptome of two model systems, a cytosolic (mCherry) and a secreted (hemagglutinin) protein overexpressed in Tnms42 cell, revealed distinct host responses to the different products (Koczka et al., 2018). The most significantly regulated transcripts were identified and assigned to biochemical pathways and gene ontology categories related to protein processing, folding and response to unfolded protein, opening the road to design specific virus engineering concepts for improving the yield of proteins that are dependent on the secretory pathway.

In another example, Iwanaga and co-workers performed an exploratory analysis using subtractive hybridization in order to identify differentially expressed host

genes following *Bombyx mori* nuclear polyhedrosis virus (BmNPV) infection (Iwanaga et al., 2007). The authors paid special attention to the response of energy metabolism to infection, and reported the up-regulation of citrate synthase and ATP-dependent proteasome 26S homologous genes. Citrate synthase is the first enzyme of the TCA cycle, which plays a central role in aerobic energy production and metabolic interconversions in mitochondria (Holloszy et al., 1970). On the other hand, the proteasome-ubiquitin pathway plays an important role during baculovirus infection as mentioned previously (Katsuma et al., 2011). In a follow-up study, next generation sequencing and gene enrichment analysis showed that gene sets related to mitochondrial function were highly up-regulated during the course of BmNPV infection of *Bm*5 cells (Xue et al., 2012).

Recent advances were made by Virgolini et al., 2022 exploring the gene expression of *Sf*9 insect cells producing recombinant adeno-associated virus (AAV) through a dual baculovirus expression system, with low multiplicity of infection, by RNA-seq. Interestingly, the work provided new insights into cell and/or metabolic engineering targets that can be leveraged for rational engineering of AAV production processes using IC-BEVS.

4. Virus-like particles (VLPs)

VLPs are an alternative platform for developing effective vaccines against infectious diseases (Tariq et al., 2022). These are self-assembling complexes of proteins that mimic the overall structure of their parental virus, but voided of viral genetic material. VLPs are more immunogenic than other subunit vaccines as they present repetitive antigenic epitopes on their surface in a more authentic conformation. In addition, they can be modified either chemically or genetically to present higher stability and functionality (Qian et al., 2020). VLPs

have been synthesized in various expression systems, from bacteria and yeast to plants and mammalian and insect cells (Nooraei et al., 2021).

4.1. Production of influenza VLPs in insect cells

Influenza virus infections are the leading cause of chronic human respiratory symptoms, leading to severe public health outcomes about endemic and seasonal infections and even result in unpredictable pandemic outbreaks (C. C. Lai et al., 2019). Influenza virus is an enveloped, segmented, negative-sense RNA virus, which belongs to the family *Orthomyxoviridae*. The main surface glycoproteins are neuraminidase (NA) and hemagglutinin (HA) (To & Torres, 2019). Influenza virus attaches to the residues of sialic acid on the cell membrane surface via HA. NA appears to be less frequent on the viral surface as compared to HA, with a generally observed NA:HA ratio of 1:4. Its enzymatic function is critical in the cleavage of sialic acid, thereby facilitating viral release from the infected host cell surface. NA activation also enables the successful penetration by influenza to mucus through a mechanism involving cleavage of sialic (Buffin et al., 2019).

Influenza VLPs have been generated in insect cells using (i) single infection strategies, using one recombinant baculovirus expressing either four (HA, NA, M1 and M2) or three (HA, NA and M1) structural influenza genes, or (ii) co-infection strategies, using two recombinant baculoviruses each coding for HA and M1 genes (Quan et al., 2016). It is also possible to obtain influenza VLPs through pseudotyping approaches in which HA is displayed on the surface of a VLP core from another virus (Haynes et al., 2009). Efforts have been made to generate vaccines protecting from potentially pandemic strains. Sequeira et al., 2018 developed a modular system comprising stable and baculovirus-mediated expression in insect cells for production of multi-HA influenza enveloped VLPs. Other studies aimed to design multivalent VLP enclosing different HA subtypes

showing promising results as a vaccine candidate against virus strains with pandemic potential (Tretyakova et al., 2016). Electron microscopy images of influenza VLPs show particles of similar size to native virus, with 80–120 nm (Carvalho et al., 2022)

Different bioprocess engineering strategies have been employed to enhance cell specific influenza VLPs productivity (Charlton Hume et al., 2019); adaptive laboratory evolution (ALE) of cell lines is one of those approaches. The use of ALE, i.e. adaptation of cells to efficiently grow under non-standard culture conditions through consecutive sub-culturing, allows for the selection of cell populations with enhanced fitness. ALE has been suggested as an approach to maximize recombinant protein titers in both prokaryotes and animal cells (Mundhada et al., 2017; Sunley et al., 2008). The first indication in IC-BEVS for possible success came from work by J. M. Wagner et al., 2014 who showed that a novel insect cell variant derived by exposure of *Sf*21 to elevated culture pH for a prolonged period of time was capable of maintaining normal cell growth into the typical mammalian cell culture consecutive sub-culturing under these and produced 11-fold higher Chikungunya VLP yields compared to the parental *Sf* cell line (J. M. Wagner et al., 2014). Correia et al., 2020 used a similar approach to adapt T. ni insect cells to grow at a neutral culture pH (7.0) resulting in improved production of influenza HA-displaying VLPs. The cell-specific HA productivity was increased three-fold and volumetric HA titer of up to four-fold as compared to non-adapted cells, whereas a pH shift alone did not improve yield. Fernandes et al., 2020 used ALE to improve the production of HIV-Gag VLPs in stable *Sf*9 and T. ni cell lines. Cells were cultured under controlled hypothermic conditions (22 C instead of standard 27 C) for a prolonged period of time (over 3 months), which allowed the selection of a population of cells with an improved phenotype. Adapted cells expressed up to 26-fold (*Sf* cells) and 10-fold (T. ni cells) more Gag VLPs than non-adapted cells cultured at

standard conditions (Fernandes et al., 2020). Dissecting the mechanisms driving the better phenotype of adapted cells, for example using scRNA-seq analysis, would be key for rational cell/virus engineering towards improved VLPs production.

5. Concluding remarks

Viral infection induces a multi-level response in the host, during which a vast number of molecular changes occur that ultimately dictates the performance of an expression system. A deeper understanding of such phenomena, for example through transcriptomic analysis, will certainly allow for the rational design of strategies towards bioprocess optimization. In this context, the application of bulk and scRNA-seq approaches to IC-BEVS may assist in uncovering the fundamental intricacies of this system, revealing targets for genetic engineering towards improved process performance.

6. Aim of the thesis

The IC-BEVS is a powerful platform for recombinant protein production. Nevertheless, the lack of comprehensive understanding about the impact of baculovirus infection on the host cell machinery still poses major challenges to further improve this expression system. The work developed during this PhD thesis aimed at unveiling the key biological processes of insect cells that are impacted during infection by baculovirus. The knowledge herein extracted could be used to identify relevant target genes to design specific host and/or virus engineering concepts for improving protein yields.

In the first part of the work (**Chapter 2**), the transcriptome of insect High Five cells adapted to neutral pH and producing influenza VLPs using IC-BEVS, was

assessed by bulk RNA-seq and compared to those of parental, non-adapted insect High Five cells in order to gain an understanding of the mechanisms driving higher productivity in adapted cells. In the second part (**Chapter 3**), single-cell RNA-seq was used to assess the variability of (parental) insect cell populations and gene expression profiles along infection, providing insight for rational cell and process engineering towards improved VLP production in IC-BEVS.

This PhD thesis advances our understanding of the IC-BEVS, constituting a reference study of the application of transcriptomics to a relevant virus-based process of biotechnological interest. Moreover, it provides clues on the impact of baculovirus infection in insect cells, adding new insights for the complex study of baculovirus-host cell interactions.

7. References

Aizat, W. M., Ismail, I., & Noor, N. M. (2018). Recent Development in Omics Studies. *Advances in experimental medicine and biology, 1102*, 1–9. https://doi.org/10.1007/978-3-319-98758-3_1

Berger, A., le Fourn, V., Masternak, J., Regamey, A., Bodenmann, I., Girod, P. A., & Mermod, N. (2020). Overexpression of transcription factor Foxa1 and target genes remediate therapeutic protein production bottlenecks in Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *117*(4), 1101–1116. https://doi.org/10.1002/bit.27274

Bernal, V., Carinhas, N., Yokomizo, A. Y., Carrondo, M. J. T., & Alves, P. M. (2009). Cell density effect in the baculovirus-insect cells system: A quantitative analysis of energetic metabolism. *Biotechnology and Bioengineering*, *104*(1), 162–180. https://doi.org/10.1002/bit.22364

Bernal, V., Monteiro, F., Carinhas, N., Ambrósio, R., & Alves, P. M. (2010). An integrated analysis of enzyme activities, cofactor pools and metabolic fluxes in baculovirus-infected Spodoptera frugiperda Sf9 cells. *Journal of Biotechnology*, *150*(3), 332–342. https://doi.org/10.1016/j.jbiotec.2010.09.958

Bernardino, T. C., Astray, R. M., Pereira, C. A., Boldorini, V. L., Antoniazzi, M. M., Jared, S. G. S., Núñez, E. G. F., & Jorge, S. A. C. (2021). Production of Rabies VLPs in Insect Cells by Two Monocistronic Baculoviruses Approach. *Molecular Biotechnology*. https://doi.org/10.1007/s12033-021-00366-z

Budge, J. D., Knight, T. J., Povey, J., Roobol, J., Brown, I. R., Singh, G., Dean, A., Turner, S., Jaques, C. M., Young, R. J., Racher, A. J., & Smales, C. M. (2020). Engineering of Chinese hamster ovary cell lipid metabolism results in an expanded ER and enhanced recombinant biotherapeutic protein production. *Metabolic Engineering*, *57*, 203–216. https://doi.org/10.1016/j.ymben.2019.11.007

Buffin, S., Peubez, I., Barrière, F., Nicolaï, M. C., Tapia, T., Dhir, V., Forma, E., Sève, N., & Legastelois, I. (2019). Influenza A and B virus-like particles produced in mammalian cells are highly immunogenic and induce functional antibodies. *Vaccine*, *37*(46), 6857–6867. https://doi.org/10.1016/j.vaccine.2019.09.057

Butler M. (2005). Animal cell cultures: recent achievements and perspectives in the production of biopharmaceuticals. *Applied microbiology and biotechnology, 68*(3), 283–291. https://doi.org/10.1007/s00253-005-1980-8

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C., & Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*(7745), 496–502. https://doi.org/10.1038/s41586-019-0969-x

Carinhas, N., Bernal, V., Monteiro, F., Carrondo, M. J. T., Oliveira, R., & Alves, P. M. (2010). Improving baculovirus production at high cell density through manipulation of energy metabolism. *Metabolic Engineering*, *12*(1), 39–52. https://doi.org/10.1016/j.ymben.2009.08.008

Carinhas, N., Bernal, V., Teixeira, A. P., Carrondo, M. J. T., Alves, P. M., & Oliveira, R. (2011). Hybrid metabolic flux analysis: Combining stoichiometric and statistical constraints to model the formation of complex recombinant products. *BMC Systems Biology*, *5*, 34. https://doi.org/10.1186/1752-0509-5-34

Carinhas, N., Robitaille, A. M., Moes, S., Carrondo, M. J. T., Jenoe, P., Oliveira, R., & Alves, P. M. (2011). Quantitative proteomics of Spodoptera frugiperda cells during growth and baculovirus infection. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0026444

Carvalhal, A. v., Santos, S. S., & Carrondo, M. J. T. (2011). Extracellular purine and pyrimidine catabolism in cell culture. *Biotechnology Progress*, *27*(5), 1373–1382. https://doi.org/10.1002/btpr.656

Carvalho, S. B., Silva, R. J. S., Sousa, M. F. Q., Peixoto, C., Roldão, A., Carrondo, M. J. T., & Alves, P. M. (2022). Bioanalytics for Influenza Virus-Like Particle Characterization and Process Monitoring. *Frontiers in Bioengineering and Biotechnology*, *10*. https://doi.org/10.3389/fbioe.2022.805176

Charlton Hume, H. K., Vidigal, J., Carrondo, M. J. T., Middelberg, A. P. J., Roldão, A., & Lua, L. H. L. (2019). Synthetic biology for bioengineering virus-like particle vaccines. *Biotechnology and bioengineering, 116*(4), 919–935. https://doi.org/10.1002/bit.26890

Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F., & Poidinger, M. (2016). Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11988

Chen, Y.-R., Zhong, S., Fei, Z., Gao, S., Zhang, S., Li, Z., Wang, P., & Blissard, G. W. (2014). Transcriptome Responses of the Host Trichoplusia ni to Infection by the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus. *Journal of Virology*, *88*(23), 13781–13797. https://doi.org/10.1128/jvi.02243-14

Cid, R., & Bolívar, J. (2021). Platforms for Production of Protein-Based Vaccines: From Classical to Next-Generation Strategies. *Biomolecules, 11*(8), 1072. https://doi.org/10.3390/biom11081072

Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E., & Chisti, Y. (2014). Protein production using the baculovirus-insect cell expression system. *Biotechnology Progress*, *30*(1), 1–18. https://doi.org/10.1002/btpr.1842

Correia, R., Fernandes, B., Alves, P. M., Carrondo, M. J. T., & Roldão, A. (2020). Improving influenza HA-VLPs production in insect high five cells via adaptive laboratory evolution. *Vaccines*, *8*(4), 1–15. https://doi.org/10.3390/vaccines8040589

Courtes, F. C., Lin, J., Lim, H. L., Ng, S. W., Wong, N. S. C., Koh, G., Vardy, L., Yap, M. G. S., Loo, B., & Lee, D. Y. (2013). Translatome analysis of CHO cells to identify key growth genes. *Journal of Biotechnology*, *167*(3), 215–224. https://doi.org/10.1016/j.jbiotec.2013.07.010

Cullum, R., Alder, O., & Hoodless, P. A. (2011). The next generation: Using new sequencing technologies to analyse gene regulation. *Respirology*, *16*(2), 210–222. https://doi.org/10.1111/j.1440-1843.2010.01899.x

Doolan, P., Meleady, P., Barron, N., Henry, M., Gallagher, R., Gammell, P., Melville, M., Sinacore, M., McCarthy, K., Leonard, M., Charlebois, T., & Clynes, M. (2010). Microarray and proteomics expression profiling identifies several candidates, including the Valosin-Containing Protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. *Biotechnology and Bioengineering*, *106*(1), 42–56. https://doi.org/10.1002/bit.22670

Fabre, M. L., Arrías, P. N., Masson, T., Pidre, M. L., & Romanowski, V. (2020). Baculovirus-Derived Vectors for Immunization and Therapeutic Applications. *Emerging and Reemerging Viral Pathogens*, 197–224. https://doi.org/10.1016/B978-0-12-814966-9.00011-1

Feichtinger, J., Hernández, I., Fischer, C., Hanscho, M., Auer, N., Hackl, M., Jadhav, V., Baumann, M., Krempl, P. M., Schmidl, C., Farlik, M., Schuster, M., Merkel, A., Sommer, A., Heath, S., Rico, D., Bock, C., Thallinger, G. G., & Borth, N. (2016). Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. *Biotechnology and bioengineering, 113*(10), 2241–2253. https://doi.org/10.1002/bit.25990

Felberbaum R. S. (2015). The baculovirus expression vector system: A commercial manufacturing platform for viral vaccines and gene therapy vectors. *Biotechnology journal, 10*(5), 702–714. https://doi.org/10.1002/biot.201400438

Fernandes, B., Vidigal, J., Correia, R., Carrondo, M. J. T., Alves, P. M., Teixeira, A. P., & Roldão, A. (2020). Adaptive laboratory evolution of stable insect cell lines for improved HIV-Gag VLPs production. *Journal of Biotechnology*, *307*, 139–147. https://doi.org/10.1016/j.jbiotec.2019.10.004

Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A. L., Oomen, M. E., Dekker, J., Mirny, L., & Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell*, *176*(6), 1502-1515.e10. https://doi.org/10.1016/j.cell.2019.01.020

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., … Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology, 5*(10), R80. https://doi.org/10.1186/gb-2004-5-10-r80

Gierliński, M., Cole, C., Schofield, P., Schurch, N. J., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2015). Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. *Bioinformatics*, *31*(22), 3625–3630. https://doi.org/10.1093/bioinformatics/btv425

Golumbeanu, M., Cristinelli, S., Rato, S., Munoz, M., Cavassini, M., Beerenwinkel, N., & Ciuffi, A. (2018). Single-Cell RNA-Seq Reveals Transcriptional Heterogeneity in Latent and Reactivated HIV-Infected Cells. *Cell Reports*, *23*(4), 942–950. https://doi.org/10.1016/j.celrep.2018.03.102

Grose, C., Putman, Z., & Esposito, D. (2021). A review of alternative promoters for optimal recombinant protein expression in baculovirus-infected insect cells. *Protein expression and purification, 186*, 105924. https://doi.org/10.1016/j.pep.2021.105924

Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F., & Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods*, *13*(10), 845–848. https://doi.org/10.1038/nmeth.3971

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., Huang, D., Xu, Y., Huang, W., Jiang, M., Jiang, X., Mao, J., Chen, Y., Lu, C., Xie, J., … Guo, G. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*, *172*(5), 1091-1107.e17. https://doi.org/10.1016/j.cell.2018.02.001

Haynes, J. R., Dokken, L., Wiley, J. A., Cawthon, A. G., Bigger, J., Harmsen, A. G., & Richardson, C. (2009). Influenza-pseudotyped Gag virus-like particle vaccines provide broad protection against highly pathogenic avian influenza challenge. *Vaccine*, *27*(4), 530–541. https://doi.org/10.1016/j.vaccine.2008.11.011

Holloszy, J. 0., Oscai, L. B., Don, I. J., Molé, P. A. (1970). Mitochondrial citric acid cycle and related enzymes: Adaptive response to exercise. *Biochemical and Biophysical Research Communications, 40* (6), 1368-1373. https://doi.org/10.1016/0006-291x(70)90017-3

Huang, N., Civciristov, S., Hawkins, C. J., & Clem, R. J. (2013). SfDronc, an initiator caspase involved in apoptosis in the fall armyworm Spodoptera frugiperda. *Insect Biochemistry and Molecular Biology*, *43*(5), 444–454. https://doi.org/10.1016/j.ibmb.2013.02.005

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MaCdonald, J., Obenchain, V., Oleš, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

Hunter, M., Yuan, P., Vavilala, D., & Fox, M. (2019). Optimization of Protein Expression in Mammalian Cells. *Current Protocols in Protein Science*, *95*(1). https://doi.org/10.1002/cpps.77

Iwanaga, M., Shimada, T., Kobayashi, M., & Kang, W. K. (2007). Identification of differentially expressed host genes in Bombyx mori nucleopolyhedrovirus infected cells by using subtractive hybridization. *Applied Entomology and Zoology*, *42*(1), 151–159. https://doi.org/10.1303/aez.2007.151

Ji, Z., & Ji, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Research*, *44*(13), e117. https://doi.org/10.1093/nar/gkw430

Kamen, A. A., Bedard, C., Tom, R., Perret, S., & Jardin, B. (1996). On-Line monitoring of respiration in recombinant-baculovirus infected and uninfected insect cell bioreactor cultures. *Biotechnology and bioengineering, 50*(1), 36–48. https://doi.org/10.1002/(SICI)1097-0290(19960405)50:1<36::AID-BIT5>3.0.CO;2-2

Kanehisa M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein science : a publication of the Protein Society, 28*(11), 1947–1951. https://doi.org/10.1002/pro.3715

Katsuma, S., Tsuchida, A., Matsuda-Imai, N., Kang, W. K., & Shimada, T. (2011). Role of the ubiquitin-proteasome system in Bombyx mori nucleopolyhedrovirus infection. *Journal of General Virology*, *92*(3), 699–705. https://doi.org/10.1099/vir.0.027573-0

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. https://doi.org/10.1016/j.jbiotec.2018.02.001

Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The technology and biology of single-cell RNA sequencing. Molecular cell, 58(4), 610–620. https://doi.org/10.1016/j.molcel.2015.04.005

Korbel, J. O., Urban, A. E., Grubert, F., Du, J., Royce, T. E., Starr, P., Zhong, G., Emanuel, B. S., Weissman, S. M., Snyder, M., & Gerstein, M. B. (2007). Systematic prediction and validation of breakpoints associated with copy-number variants in the human

genome. *Proceedings of the National Academy of Sciences of the United States of America, 104*(24), 10110–10115. https://doi.org/10.1073/pnas.0703834104

Kupke, S. Y., Ly, L. H., Thomas Börno, S., Ruff, A., Timmermann, B., Vingron, M., Haas, S., & Reichl, U. (2020). Single-cell analysis uncovers a vast diversity in intracellular viral defective interfering RNA content affecting the large cell-to-cell heterogeneity in influenza A virus replication. *Viruses*, *12*(1). https://doi.org/10.3390/v12010071

Lai, C. C., Cheng, Y. C., Chen, P. W., Lin, T. H., Tzeng, T. T., Lu, C. C., Lee, M. S., & Hu, A. Y. C. (2019). Process development for pandemic influenza VLP vaccine production using a baculovirus expression system. *Journal of Biological Engineering*, *13*(1). https://doi.org/10.1186/s13036-019-0206-z

Lai, T.; Yang, Y.; Ng, S.K. Advances in Mammalian Cell Line Development Technologies for Recombinant Protein Production. *Pharmaceuticals 2013, 6*, 579-603. https://doi.org/10.3390/ph6050579

Lai, Y. K., Hsu, J. T., Chu, C. C., Chang, T. Y., Pan, K. L., & Lin, C. C. (2012). Enhanced recombinant protein production and differential expression of molecular chaperones in sf-caspase-1-repressed stable cells after baculovirus infection. *BMC biotechnology, 12*, 83. https://doi.org/10.1186/1472-6750-12-83

Levy, S. E., & Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annual review of genomics and human genetics, 17*, 95–115. https://doi.org/10.1146/annurev-genom-083115-022413

Lewis, A. M., Abu-Absi, N. R., Borys, M. C., & Li, Z. J. (2016). The use of 'Omics technology to rationally improve industrial mammalian cell line performance. *Biotechnology and bioengineering, 113*(1), 26–38. https://doi.org/10.1002/bit.25673

Lin, C.-C., Hsu, J. T.-A., Huang, K.-L., Tang, H.-K., Shu, C.-W., & Lai, Y.-K. (2007). Sf-Caspase-1-repressed stable cells: resistance to apoptosis and augmentation of recombinant protein production. *Biotechnology and Applied Biochemistry*, *48*(1), 11. https://doi.org/10.1042/ba20070044

Mardis E. R. (2008). Next-generation DNA sequencing methods. *Annual review of genomics and human genetics, 9*, 387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359

Mardis E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in genetics : TIG, 24*(3), 133–141. https://doi.org/10.1016/j.tig.2007.12.007

Maynard, N. D., Gutschow, M. V., Birch, E. W., & Covert, M. W. (2010). The virus as metabolic engineer. *Biotechnology journal, 5*(7), 686–694. https://doi.org/10.1002/biot.201000080

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. http://arxiv.org/abs/1802.03426

McWilliam Leitch, E. C., & McLauchlan, J. (2013). Determining the Cellular Diversity of Hepatitis C Virus Quasispecies by Single-Cell Viral Sequencing. *Journal of Virology*, *87*(23), 12648–12655. https://doi.org/10.1128/jvi.01602-13

Metz, S. W., Gardner, J., Geertsema, C., Le, T. T., Goh, L., Vlak, J. M., Suhrbier, A., & Pijlman, G. P. (2013). Effective Chikungunya Virus-like Particle Vaccine Produced in Insect Cells. *PLoS Neglected Tropical Diseases*, *7*(3). https://doi.org/10.1371/journal.pntd.0002124

Mi, Y., Xie, T., Zhu, B., Tan, J., Li, X., Luo, Y., Li, F., Niu, H., Han, J., Lv, W., & Wang, J. (2021). Production of sars-cov-2 virus-like particles in insect cells. *Vaccines*, *9*(6). https://doi.org/10.3390/vaccines9060554

Monteiro, F., Bernal, V., & Alves, P. M. (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*(3), 674–684. https://doi.org/10.1002/bit.26089

Monteiro, F., Carinhas, N., Carrondo, M. J., Bernal, V., & Alves, P. M. (2012). Toward system-level understanding of baculovirus-host cell interactions: from molecular fundamental studies to large-scale proteomics approaches. *Frontiers in microbiology, 3*, 391. https://doi.org/10.3389/fmicb.2012.00391

Mundhada, H., Seoane, J. M., Schneider, K., Koza, A., Christensen, H. B., Klein, T., Phaneuf, P. v., Herrgard, M., Feist, A. M., & Nielsen, A. T. (2017). Increased production of L-serine in Escherichia coli through Adaptive Laboratory Evolution. *Metabolic Engineering*, *39*, 141–150. https://doi.org/10.1016/j.ymben.2016.11.008

Mutz, K. O., Heilkenbrinker, A., Lönne, M., Walter, J. G., & Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology, 24*(1), 22-30. https://doi.org/10.1016/j.copbio.2012.09.004

Naskalska, A., Dabrowska, A., Szczepanski, A., Jasik, K. P., Gromadzka, B., & Pyrc, K. (2021). Functional Severe Acute Respiratory Syndrome Coronavirus 2 Virus-Like Particles From Insect Cells. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.732998

Nooraei, S., Bahrulolum, H., Hoseini, Z. S., Katalani, C., Hajizade, A., Easton, A. J., & Ahmadian, G. (2021). Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. *Journal of nanobiotechnology, 19*(1), 59. https://doi.org/10.1186/s12951-021-00806-7

O'Callaghan, P. M., & James, D. C. (2008). Systems biotechnology of mammalian cell factories. *Briefings in functional genomics & proteomics, 7*(2), 95–110. https://doi.org/10.1093/bfgp/eln012

O'Flaherty, R., Bergin, A., Flampouri, E., Mota, L. M., Obaidi, I., Quigley, A., Xie, Y., & Butler, M. (2020). Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. *Biotechnology advances, 43*, 107552. https://doi.org/10.1016/j.biotechadv.2020.107552

Otero, J. M., & Nielsen, J. (2010). Industrial systems biology. *Biotechnology and bioengineering, 105*(3), 439–460. https://doi.org/10.1002/bit.22592

Pagani, I., Liolios, K., Jansson, J., Chen, I. M., Smirnova, T., Nosrat, B., Markowitz, V. M., & Kyrpides, N. C. (2012). The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic acids research, 40*(Database issue), D571–D579. https://doi.org/10.1093/nar/gkr1100

Palomares, L. A., López, S., & Ramírez, O. T. (2004). Utilization of oxygen uptake rate to assess the role of glucose and glutamine in the metabolism of infected insect cell cultures. *Biochemical Engineering Journal*, *19*(1), 87–93. https://doi.org/10.1016/j.bej.2003.12.002

Park P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews. Genetics, 10*(10), 669–680. https://doi.org/10.1038/nrg2641

Pushko, P., Tumpey, T. M., Bu, F., Knell, J., Robinson, R., & Smith, G. (2005). Influenza virus-like particles comprised of the HA, NA, and M1 proteins of H9N2 influenza virus induce protective immune responses in BALB/c mice. *Vaccine*, *23*(50), 5751–5759. https://doi.org/10.1016/j.vaccine.2005.07.098

Qian, C., Liu, X., Xu, Q., Wang, Z., Chen, J., Li, T., Zheng, Q., Yu, H., Gu, Y., Li, S., & Xia, N. (2020). Recent Progress on the Versatility of Virus-Like Particles. *Vaccines, 8*(1), 139. https://doi.org/10.3390/vaccines8010139

Quan, F. S., Lee, Y. T., Kim, K. H., Kim, M. C., & Kang, S. M. (2016). Progress in developing virus-like particle influenza vaccines. *Expert review of vaccines, 15*(10), 1281–1293. https://doi.org/10.1080/14760584.2016.1175942

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C. E., Socci, N. D., & Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome biology, 14*(9), R95. https://doi.org/10.1186/gb-2013-14-9-r95

Rato, S., Rausell, A., Muñoz, M., Telenti, A., & Ciuffi, A. (2017). Single-cell analysis identifies cellular markers of the HIV permissive cell. *PLoS Pathogens*, *13*(10). https://doi.org/10.1371/journal.ppat.1006678

Russell, A. B., Trapnell, C., & Bloom, J. D. (2018). Extreme heterogeneity of influenza virus infection in single cells. *eLife, 7*, e32303. https://doi.org/10.7554/eLife.32303

Samoudi, M., Masson, H. O., Kuo, C. C., Robinson, C. M., & Lewis, N. E. (2021). From omics to cellular mechanisms in mammalian cell factory development. *Current Opinion in Chemical Engineering, 32*. https://doi.org/10.1016/j.coche.2021.100688

Schaly, S., Ghebretatios, M., & Prakash, S. (2021). Baculoviruses in Gene Therapy and Personalized Medicine. *Biologics : targets & therapy, 15*, 115–132. https://doi.org/10.2147/BTT.S292692

Schaub, J., Clemens, C., Kaufmann, H., & Schulz, T. W. (2012). Advancing biopharmaceutical process development by system-level data analysis and integration of omics data. *Advances in Biochemical Engineering/Biotechnology*, *127*, 133–163. https://doi.org/10.1007/10_2010_98

Schaub, J., Clemens, C., Schorn, P., Hildebrandt, T., Rust, W., Mennerich, D., Kaufmann, H., & Schulz, T. W. (2010). CHO gene expression profiling in biopharmaceutical process analysis and design. *Biotechnology and Bioengineering*, *105*(2), 431–438. https://doi.org/10.1002/bit.22549

Schreiber, K., Csaba, G., Haslbeck, M., & Zimmer, R. (2015). Alternative splicing in next generation sequencing data of saccharomyces cerevisiae. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0140487

Sellick, C. A., Croxford, A. S., Maqsood, A. R., Stephens, G., Westerhoff, H. v., Goodacre, R., & Dickson, A. J. (2011). Metabolite profiling of recombinant CHO cells: Designing tailored feeding regimes that enhance recombinant antibody production. *Biotechnology and Bioengineering*, *108*(12), 3025–3031. https://doi.org/10.1002/bit.23269

Selvarasu, S., Ho, Y. S., Chong, W. P. K., Wong, N. S. C., Yusufi, F. N. K., Lee, Y. Y., Yap, M. G. S., & Lee, D. Y. (2012). Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnology and Bioengineering*, *109*(6), 1415–1429. https://doi.org/10.1002/bit.24445

Sequeira, D. P., Correia, R., Carrondo, M. J. T., Roldão, A., Teixeira, A. P., & Alves, P. M. (2018). Combining stable insect cell lines with baculovirus-mediated expression for multi-HA influenza VLP production. *Vaccine*, *36*(22), 3112–3123. https://doi.org/10.1016/j.vaccine.2017.02.043

Seyednasrollah, F., Laiho, A., & Elo, L. L. (2013). Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in Bioinformatics*, *16*(1), 59–70. https://doi.org/10.1093/bib/bbt086

Sharfstein, S. T. (2008). Advances in cell culture process development: Tools and techniques for improving cell line development and process optimization. *Biotechnology Progress*, *24*(3), 727–734. https://doi.org/10.1021/bp070471q

Shen, C. F., & Kamen, A. (2012). Hyperosmotic pressure on HEK 293 cells during the growth phase, but not the production phase, improves adenovirus production. *Journal of Biotechnology*, *157*(1), 228–236. https://doi.org/10.1016/j.jbiotec.2011.11.016

Shima, R., Li, T. C., Sendai, Y., Kataoka, C., Mori, Y., Abe, T., Takeda, N., Okamoto, T., & Matsuura, Y. (2016). Production of hepatitis E virus-like particles presenting multiple foreign epitopes by co-infection of recombinant baculoviruses. *Scientific Reports*, *6*. https://doi.org/10.1038/srep21638

Soneson, C., & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, *14,* 91. https://doi.org/10.1186/1471-2105-14-91

Spies, D., & Ciaudo, C. (2015). Dynamics in Transcriptomics: Advancements in RNA-seq Time Course and Downstream Analysis. *Computational and structural biotechnology journal, 13*, 469–477. https://doi.org/10.1016/j.csbj.2015.08.004

Spitzer, M. H., & Nolan, G. P. (2016). Mass Cytometry: Single Cells, Many Features. *Cell, 165*(4), 780–791. https://doi.org/10.1016/j.cell.2016.04.019

Steele, K. H., Stone, B. J., Franklin, K. M., Fath-Goodin, A., Zhang, X., Jiang, H., Webb, B. A., & Geisler, C. (2017). Improving the baculovirus expression vector system with vankyrin-enhanced technology. *Biotechnology Progress*, *33*(6), 1496–1507. https://doi.org/10.1002/btpr.2516

Stephanopoulos, G., Alper, H., & Moxley, J. (2004). Exploiting biological complexity for strain improvement through systems biology. *Nature biotechnology, 22*(10), 1261–1267. https://doi.org/10.1038/nbt1016

Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and biology insights, 14*, 1177932219899051. https://doi.org/10.1177/1177932219899051

Sumit, M., Dolatshahi, S., Chu, A. H. A., Cote, K., Scarcelli, J. J., Marshall, J. K., Cornell, R. J., Weiss, R., Lauffenburger, D. A., Mulukutla, B. C., & Figueroa, B. (2019). Dissecting N-Glycosylation Dynamics in Chinese Hamster Ovary Cells Fed-batch Cultures using Time Course Omics Analyses. *IScience*, *12*, 102–120. https://doi.org/10.1016/j.isci.2019.01.006

Sun, J., Vera, J. C., Drnevich, J., Lin, Y. T., Ke, R., & Brooke, C. B. (2020). Single cell heterogeneity in influenza A virus gene expression shapes the innate antiviral response to infection. *PLoS Pathogens*, *16*(7). https://doi.org/10.1371/journal.ppat.1008671

Sun, S., Zhu, J., Ma, Y., & Zhou, X. (2019). Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1898-6

Sunley, K., Tharmalingam, T., & Butler, M. (2008). CHO cells adapted to hypothermic growth produce high yields of recombinant beta-interferon. *Biotechnology progress, 24*(4), 898–906. https://doi.org/10.1002/btpr.9

Svensson, V., Natarajan, K. N., Ly, L. H., Miragaia, R. J., Labalette, C., Macaulay, I. C., Cvejic, A., & Teichmann, S. A. (2017). Power analysis of single-cell rnA-sequencing experiments. *Nature Methods*, *14*(4), 381–387. https://doi.org/10.1038/nmeth.4220

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*(5), 377–382. https://doi.org/10.1038/nmeth.1315

Tariq, H., Batool, S., Asif, S., Ali, M., & Abbasi, B. H. (2022). Virus-Like Particles: Revolutionary Platforms for Developing Vaccines Against Emerging Infectious Diseases. *Frontiers in microbiology, 12*, 790121. https://doi.org/10.3389/fmicb.2021.790121

To, J., & Torres, J. (2019). Viroporins in the Influenza Virus. *Cells, 8*(7), 654. https://doi.org/10.3390/cells8070654

Touzé, A., Dupuy, C., Chabaud, M., Le Cann, P., & Coursaget, P. (1996). Production of human papillomavirus type 45 virus-like particles in insect cells using a recombinant baculovirus. *FEMS microbiology letters, 141*(1), 111–116. https://doi.org/10.1111/j.1574-6968.1996.tb08371.x

Tretyakova, I., Hidajat, R., Hamilton, G., Horn, N., Nickols, B., Prather, R. O., Tumpey, T. M., & Pushko, P. (2016). Preparation of quadri-subtype influenza virus-like particles using bovine immunodeficiency virus gag protein. *Virology*, *487*, 163–171. https://doi.org/10.1016/j.virol.2015.10.007

van Oers, M. M., Pijlman, G. P., & Vlak, J. M. (2015). Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *The Journal of general virology, 96*(Pt 1), 6–23. https://doi.org/10.1099/vir.0.067108-0

Vidigal, J., Fernandes, B., Dias, M. M., Patrone, M., Roldão, A., Carrondo, M. J. T., Alves, P. M., & Teixeira, A. P. (2018). RMCE-based insect cell platform to produce membrane proteins captured on HIV-1 Gag virus-like particles. *Applied Microbiology and Biotechnology*, *102*(2), 655–666. https://doi.org/10.1007/s00253-017-8628-3

Virgolini, N., Hagan, R., Correia, R., Silvano, M., Fernandes, S., Alves, P. M., Clarke, C., Roldão, A., & Isidro, I. A. (2023). Transcriptome analysis of Sf9 insect cells during

production of recombinant Adeno-associated virus. *Biotechnology journal, 18*(2), e2200466. https://doi.org/10.1002/biot.202200466

Wagner, J. M., Pajerowski, J. D., Daniels, C. L., McHugh, P. M., Flynn, J. A., Balliet, J. W., Casimiro, D. R., & Subramanian, S. (2014). Enhanced production of Chikungunya virus-like particles using a high-pH adapted Spodoptera frugiperda insect cell line. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0094401

Wang, C., Forst, C. v., Chou, T. W., Geber, A., Wang, M., Hamou, W., Smith, M., Sebra, R., Zhang, B., Zhou, B., & Ghedin, E. (2020). Cell-to-cell variation in defective virus expression and effects on host responses during influenza virus infection. *MBio*, *11*(1). https://doi.org/10.1128/mBio.02880-19

Wang, C., Zheng, X., Gai, W., Zhao, Y., Wang, H., Wang, H., Feng, N., Chi, H., Qiu, B., Li, N., Wang, T., Gao, Y., Yang, S., & Xia, X. (2017). MERS-CoV virus-like particles produced in insect cells induce specific humoural and cellular imminity in rhesus macaques. *Oncotarget, 8*(8), 12686–12694. https://doi.org/10.18632/oncotarget.8475

Wang, J. H., Byun, J., & Pennathur, S. (2010). Analytical approaches to metabolomics and applications to systems biology. *Seminars in Nephrology*, *30*(5), 500–511. https://doi.org/10.1016/j.semnephrol.2010.07.007

Wang, Q., Zhou, Y., Chen, K., & Ju, X. (2016). Suppression of Bm-Caspase-1 Expression in BmN Cells Enhances Recombinant Protein Production in a Baculovirus Expression Vector System. *Molecular Biotechnology*, *58*(5), 319–327. https://doi.org/10.1007/s12033-016-9931-4

Wuest, D. M., Harcum, S. W., & Lee, K. H. (2012). Genomics in mammalian cell culture bioprocessing. *Biotechnology advances, 30*(3), 629–638. https://doi.org/10.1016/j.biotechadv.2011.10.010

Xue, J., Qiao, N., Zhang, W., Cheng, R. L., Zhang, X. Q., Bao, Y. Y., Xu, Y. P., Gu, L. Z., Han, J. D., & Zhang, C. X. (2012). Dynamic interactions between Bombyx mori nucleopolyhedrovirus and its host cells revealed by transcriptome analysis. *Journal of virology, 86*(13), 7345–7359. https://doi.org/10.1128/JVI.07217-12

Ye, L., Lin, J., Sun, Y., Bennouna, S., Lo, M., Wu, Q., Bu, Z., Pulendran, B., Compans, R. W., & Yang, C. (2006). Ebola virus-like particles produced in insect cells exhibit dendritic cell stimulating activity and induce neutralizing antibodies. *Virology*, *351*(2), 260–270. https://doi.org/10.1016/j.virol.2006.03.021

Yeo, H. C., Chen, S., Ho, Y. S., & Lee, D. Y. (2018). An LC–MS-based lipidomics pre-processing framework underpins rapid hypothesis generation towards CHO systems biotechnology. *Metabolomics*, *14*(7). https://doi.org/10.1007/s11306-018-1394-0

Young J. D. (2013). Metabolic flux rewiring in mammalian cell cultures. *Current opinion in biotechnology, 24*(6), 1108–1115. https://doi.org/10.1016/j.copbio.2013.04.01

Zhang, X., Xu, K., Ou, Y., Xu, X., & Chen, H. (2018). Development of a baculovirus vector carrying a small hairpin RNA for suppression of sf-caspase-1 expression and improvement of recombinant protein production. *BMC Biotechnology*, *18*(1). https://doi.org/10.1186/s12896-018-0434-1

Zhang, X., Zhao, K., Lan, L., Shi, N., Nan, H., Shi, Y., Xu, X., & Chen, H. (2021). Improvement of protein production by engineering a novel antiapoptotic baculovirus vector to suppress the expression of Sf-caspase-1 and Tn-caspase-1. *Biotechnology and Bioengineering*, *118*(8), 2977–2989. https://doi.org/10.1002/bit.27807

Zhang, Y., Baycin-Hizal, D., Kumar, A., Priola, J., Bahri, M., Heffner, K. M., Wang, M., Han, X., Bowen, M. A., & Betenbaugh, M. J. (2017). High-Throughput Lipidomic and Transcriptomic Analysis to Compare SP2/0, CHO, and HEK-293 Mammalian Cell Lines. *Analytical Chemistry*, *89*(3), 1477–1485. https://doi.org/10.1021/acs.analchem.6b02984

Zhang, Z. H., Jhaveri, D. J., Marshall, V. M., Bauer, D. C., Edson, J., Narayanan, R. K., Robinson, G. J., Lundberg, A. E., Bartlett, P. F., Wray, N. R., & Zhao, Q. Y. (2014). A comparative study of techniques for differential expression analysis on RNA-seq data. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0103207

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., & Enard, W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell*, *65*(4), 631-643.e4. https://doi.org/10.1016/j.molcel.2017.01.023

Chapter 2

Gene expression analysis of higher-producer insect cells using bulk RNA-seq

This chapter was based on the following manuscript:

Silvano, M., Correia, R., Virgolini, N., Clarke, C., Alves P.M., Isidro I.A., Roldão, A. **2022**. Gene Expression Analysis of Adapted Insect Cells during Influenza VLP Production Using RNA-Sequencing. *Viruses*. 14, 2238.

Author contribution

Marco Silvano design and performed the experiments, analysed the data and wrote the chapter.

Abstract

Adaptive laboratory evolution has been used to improve production of influenza hemagglutinin (HA)-displaying virus-like particles (VLPs) in insect cells. However, little is known about the underlying biological mechanisms promoting higher HA-VLPs expression in such adapted cell lines. In this article we present a study of gene expression patterns associated with high-producer insect High Five cells adapted to neutral pH, in comparison to non-adapted cells, during expression of influenza HA-VLPs. RNA-seq shows a decrease in the amount of reads mapping to host cell genome along infection, and an increase in those mapping to baculovirus and transgenes. A total of 1 742 host cell genes were found differentially expressed between adapted and non-adapted cells throughout infection, 474 of those being either up- or down-regulated at both time points evaluated (12 and 24 hours post-infection). Interestingly, while host cell genes were found up- and down-regulated in approximately a 1:1 ratio, all baculovirus differentially expressed genes were found down-regulated in infected adapted cells. Pathway analysis of differentially expressed genes revealed enrichment of ribosome biosynthesis and carbohydrate, amino acid, and lipid metabolism. In addition, oxidative phosphorylation and protein folding, sorting and degradation pathways were also found to be overrepresented. These findings contribute to our knowledge of biological mechanisms of insect cells during baculovirus-mediated transient expression and will assist the identification of potential engineering targets to increase recombinant protein production in the future.

TABLE OF CONTENT

1. Introduction

The insect cell-baculovirus expression vector system (IC-BEVS) relies on infection of insect cells with a recombinant baculovirus genetically modified to include a nucleic acid sequence encoding a gene of interest, commonly under the transcriptional control of very-late, strong baculovirus promoters such as *polh* and *p10* (van Oers et al., 2015). The IC-BEVS system permits the expression recombinant protein at high titers and, importantly, glycosylation patterns similar to those of higher eukaryotes (Kost et al., 2005). This expression system is widely used in the production of enzymes, membrane proteins, viral capsids, and envelope proteins for use as vaccines or for analytical purposes (Kost & Kemp, 2016). The success of the human papilloma virus vaccine (Cervarix®) and the hemagglutinin-based influenza vaccine (Flublok®) illustrates the utility of the system for biopharmaceutical manufacturing.

In recent years, academic and industrial research groups have sought to improve the knowledge of the molecular characteristics underpinning the efficient production of biopharmaceuticals using IC-BEVS (Monteiro et al., 2012). Understanding the host cell's metabolic regulation during expression of a foreign gene also facilitates the design and upscaling of a production process (Abaandou et al., 2021; J. K. Hong et al., 2018). Next-generation sequencing technologies have accelerated the development of better expression systems to produce recombinant protein (Babar et al., 2018). Transcriptomics was applied to analyze the transcriptional changes of both *Autographa californica* multiple nucleopolyhedrovirus (*Ac*MNPV) (Y.-R. Chen et al., 2013) and alphanodavirus-free High Five cells (Tnms42) (Y.-R. Chen et al., 2014) during protein expression using IC-BEVS. It has been shown that the number of viral transcripts increased significantly after the first 6 hours of infection, concomitantly with a decrease in expression of host cell transcripts due to global shut-off of the host protein synthesis (X. Du & Thiem, 1997). More recently, comparative transcriptome

analysis was conducted to study the differences in Tnms42 cells response upon expression of intracellular or secreted protein products using IC-BEVS, identifying key proteins as promising targets for achieving higher yields of protein secretion (Koczka et al., 2018).

While genetic engineering of insect cells and/or baculovirus has been used to improve production yields (Martínez-Solís et al., 2019), strategies such as shifts in culture parameters to non-physiological values have been shown to impact the growth performance and recombinant protein production (Lin et al., 2015). The use of atypical culture conditions in insect cells has also been recently addressed through adaptive laboratory evolution approaches, i.e., by adapting cells to grow at such non-standard culture conditions, allowing higher recombinant protein yields both in stable cell lines (Fernandes et al., 2020) and IC-BEVS (Correia et al., 2020). In the later, adaptation of insect High Five cells to grow at neutral pH allowed the improvement of cell-specific production rate of influenza virus-like particles (VLPs) by 3-fold. However, relatively little is known about the mechanisms underlying the higher recombinant protein productivity achieved with this adapted cell line.

In this study, the transcriptome of High Five cells adapted to neutral pH (established in Correia et al., 2020), producing influenza VLPs using IC-BEVS, were assessed by RNA-seq and compared to those of parental, non-adapted insect High Five cells in order to gain an understanding of the mechanisms behind the higher productivity of the adapted cell line.

2. Materials and methods

2.1. Cell lines and culture media

Insect High Five cells (Invitrogen), hereon referred as non-adapted cells, and High Five cells adapted to neutral pH (Correia et al., 2020), hereon referred as adapted cells were routinely sub-cultured to 0.3-0.5×10^6 cell.mL^{-1} every 2-3 days when cell concentration reached 2-3×10^6 cell.mL^{-1}. Non-adapted cells were cultured in Insect-XPRESS™ medium (Sartorius); Adapted cells were cultured in cell culture medium composed of a 1:1 mixture of Insect-XPRESS™ and chemically defined solution as previously reported (Correia et al., 2020). Both cell lines were cultured in 125-500 mL shake flasks (10% working volume) and maintained at 27 ºC in a Inova 44R shaking incubator (Eppendorf) set to 100 RPM and with an orbital motion diameter of 2.54 cm.

2.2. Baculovirus amplification and storage

Recombinant baculoviruses enclosing influenza capsid M1 from A/California/06/2009 H1N1 strain and hemagglutinin (HA) from A/Brisbane/59/2007 strain genes were kindly provided by Redbiotec AG (Schlieren, Switzerland). Amplification of baculovirus stocks was performed as described elsewhere (Vieira et al., 2005).

2.3. Production of influenza HA-VLPs

Influenza HA-VLPs were produced in 500 mL shake flasks (10% working volume) by infecting non-adapted cells and adapted cells at a cell concentration at infection (CCI) of 2×106 cell.mL-1 with baculovirus using a multiplicity of infection (MOI) of 1 pfu/cell (n=3, number of replicates). Samples were taken daily for the determination of cell concentration and viability, metabolites

concentration, and detection/relative quantification of M1 and HA proteins; for RNA-seq , samples were taken before infection, and at 12 and 24 hours post-infection (hpi) (further details in **Supplementary Figure S1**).

2.4. Analytics

2.4.1. Cell concentration and viability

Cell counting was performed in a Fuchs–Rosenthal hemacytometer chamber (Brand) and viability was assessed using the trypan-blue exclusion method (J R Tennant, 1964).

2.4.2. Metabolites concentration

For metabolites quantification, cell culture samples were centrifuged (300 ×g, 4 °C, 5 min) and supernatant collected and stored at -20 °C. Metabolite quantification was performed using Cedex Bio Analyzer 7100 (Roche).

2.4.3. Baculovirus titration

Baculovirus titers were determined using the MTT assay as described elsewhere (Mena et al., 2003; Roldão et al., 2009)

2.4.4. Western blot

For M1 and HA detection/relative quantification, cell culture samples were centrifuged (300 ×g, 4 °C, 5 min) and supernatant was collected and stored at 4 °C. Western blot analysis was performed as reported elsewhere (Correia et al., 2020). Briefly, for HA identification, a mouse monoclonal antibody (IRR, Manassas, VA, USA, FR-494—mouse monoclonal antibody to recombinant H1

HA from influenza A/Brisbane/59/2007 (H1N1)) was used at a dilution of 1:2000, and M1 protein was identified using a goat polyclonal antibody (Abcam, Cambridge, UK, Cat# ab20910) at a dilution of 1:2000. Secondary anti-mouse or anti-goat IgG antibodies conjugated with alkaline phosphatase were used at a dilution of 1:2000 for identification of HA and M1, respectively. Densitometry analysis of Western blot membranes (to assess relative productivity) was performed using the FIJI software (Schindelin et al., 2012).

2.5. RNA sequencing and data analysis

2.5.1. RNA isolation and library preparation

For RNA sequencing analysis, cell culture samples were centrifuged (300 ×g, 4°C, 5 min) and pellets collected for RNA extraction using 1 mL of Trizol (Invitrogen) and the Directzol RNA mini prep kit (Zymo Research) according to manufacturer's instructions. RNA purity and quality was assessed by spectrophotometry (mySPEC equipment, VWR) and fragment analysis (Agilent).

Library preparation and sequencing was performed elsewhere (Genewiz). In short, Poly(A) selection on total RNA (NEBNext® Poly(A) mRNA Magnetic Isolation Module) was performed prior strand-specific library preparation (NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina®). Sequencing libraries were quality checked using Qubit (Invitrogen) and a fragment analyser (Agilent), and loaded on the Illumina NovaSeq 6000 system configured to yield a minimum of 25 million, 2 × 150bp Paired-End (PE) reads per sample.

2.5.2. Data processing, alignment and counting

Trimmomatic v0.36 was used to remove adapters and to perform quality trimming (Bolger et al., 2014) of the raw RNA-seq reads. The reads were

subsequently aligned to a hybrid reference, comprising the insect *Trichoplusia ni* cell genome (Tnl; RefSeq assembly accession: GCF_003590095.1) (Fu et al., 2018), the baculovirus, i.e., *Ac*MNPV (RefSeq assembly accession: GCF_000838485.1, ViralProj14023) (Ayres et al., 1994), and the transgenes (*M1* and *HA*) sequences using STAR v2.7.3a (Dobin et al., 2013). Those reads mapping to annotated genes were counted using HTSeq (Putri et al., 2022).

2.5.3. Differential expression analysis

The edgeR Bioconductor package (Robinson et al., 2009) was used to determine the number of differentially expressed genes between infected adapted cells and non-adapted cells. Count data were normalized to account for variation in the number of sequenced reads in each sample using the TMM method (Robinson & Oshlack, 2010). To assess differential gene expression, the Fisher´s exact test used to identify statistically significant differences in gene expression between selected groups (Robinson & Smyth, 2008). Genes with expression changes of at least 1.5-fold and with a false discovery rate (FDR)-adjusted p-value < 0.05 were considered to be differentially expressed.

2.5.4. Functional annotation and pathway enrichment analysis

For gene annotation, the amino acid sequence of protein-coding genes with at least one read aligned was used as a query. Blastp search was applied in the NCBI nr protein database using Blast2GO OmicsBox software (Götz et al., 2008). No taxonomy filter was applied, and the E-value cutoff was set to 1.0×10^{-3}.

Blast2GO was used to perform pathway enrichment analysis with Fisher's exact test and the Gene Set Enrichment Analysis (GSEA) method (Kanehisa & Goto, 2000). These tests were used to identify statistically significant over-represented biological processes for each differentially expressed gene list. FDR

was applied as multiple test correction method with a cut-off of 0.05 (Benjamini & Hochberg, 1995).

3. Results

3.1. Production of influenza HA-VLPs using adapted insect High Five cells

Adapted and non-adapted cells were infected at the optimum conditions (CCI of 2×10^6 cell.mL^{-1} and MOI of 1 pfu.cell^{-1}) previously identified in our laboratory (Correia et al., 2020), and infection kinetics and HA expression were assessed throughout. Adapted cells maintained higher cell concentration and viability upon infection, with the onset of cell viability drop having a 24-hour delay in comparison with non-adapted cells (**Figure 1A**). The specific consumption or production rates for glucose (Glc), glutamine (Gln), and lactate (Lac) were similar in both cell lines, with the highest variation observed for glucose (**Figure 1B**). Importantly, M1 and HA proteins were identified by Western blot (**Supplementary Figure S2)**, with relative band intensities suggesting that expression of HA was approximately 2.5-fold higher in adapted cells while M1 expression remained similar in both cell lines (**Figure 1C**), in agreement with previously reported data. Despite not having been addressed in this work, the produced HA-VLPs were shown the typical size and morphology of influenza VLPs, thus confirming the null impact of the adaptation process and neutral culture pH on the quality of the HA-VLPs produced (Correia et al., 2020).

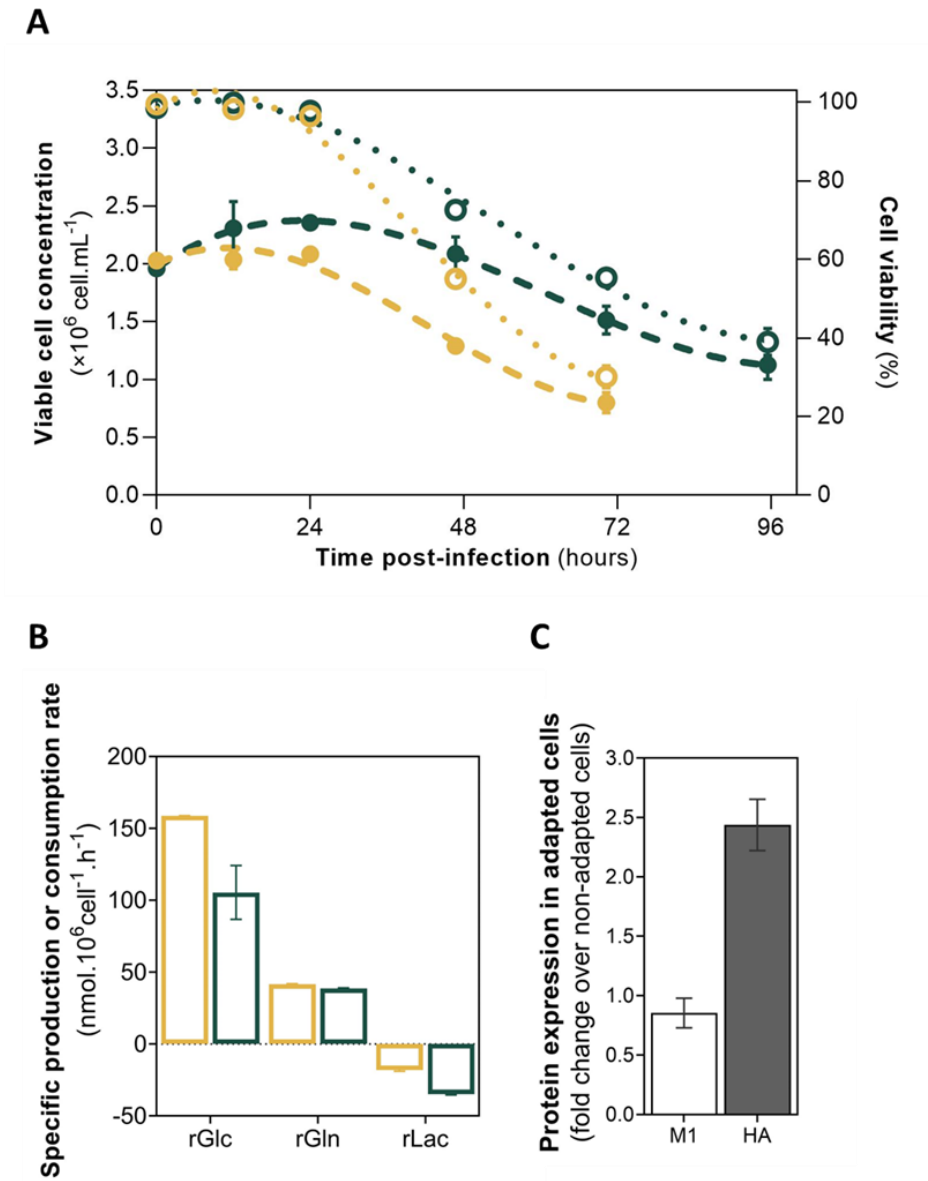


Figure 1. HA-VLPs production in not-adapted and adapted High Five insect cells using the baculovirus expression vector system. A) Viable cell concentration (full circles) and cell viability (empty circles) after infection. B) Specific glucose (rGlc) and glutamine (rGln) consumption and lactate (rLac) production (hence shown as negative) rates, estimated by linearization of metabolite concentration and integral of total cells, during infection. C) Fold-change (adapted/non-adapted) in expression of HA and M1 proteins assessed by densitometry analysis of Western blot images (i.e. relative band intensity). Color code: yellow represents non-adapted cells, green represents adapted cells. Data are expressed as mean ± standard deviation of three culture replicates (n = 3).

3.2. Gene expression profiling of infected adapted vs non-adapted cells

To reveal the differences at transcriptional level between adapted and non-adapted cells upon infection, quality-filtered reads were mapped to the reference genome and sequences. A total of 10 850 and 10 987 reads (at 12 hpi), and 10 082 and 10 236 reads (at 24 hpi) were mapped in samples from adapted and non-adapted cells, respectively; from these, 10 389 (at 12 hpi) and 9 620 (at 24 hpi) reads are coincident in both cell lines.

A reduction in the proportion of reads mapping to host cell genome was observed between 12 and 24 hpi, with a concomitant increase in reads mapping to baculovirus and transgenes; at 24 hpi, 51-60% of the total reads were mapped to baculovirus genome and 9-14 % of the total reads were mapped to transgenes (HA+M1) (**Figure 2A**). Differential gene expression analysis revealed that a total of 1 742 host cell genes were differentially expressed between adapted and non-adapted cells throughout infection, 474 of those being differentially expressed at both time points evaluated (**Figure 2B**). The top-20 most differentially expressed host cell genes at both 12 and 24 hpi are detailed in **Tables 1-2**. The most up-regulated genes encode for transmembrane transports (e.g. *organic cationic transports protein-like*, 93-fold at 24 hpi) and proteins involved in proteolysis (e.g. *xxa-Pro aminopeptidase 1*, 756-fold at 24 hpi), and lipase activity (e.g. *lipase member H-like*, 112-fold at 12 hpi), whereas the most down-regulated genes encode for proteins involved in programmed cell death (e.g. *homeobox protein abdominal-A homolog isoform X1*, 120-fold at 24 hpi), lipid transport (e.g. *apolipophorins*, 50-fold at 24 hpi), innate immune response (e.g. *protein toll-like*, 37-fold at 12 hpi), and oxidoreductase (e.g. *cytochrome P450 9e2-like,* 163-fold at 12 hpi). Protein-coding genes involved in regulation of transcription and signaling were found both up- and down-regulated. While host cell genes were found up- and down-regulated in an approx. 1:1 ratio regardless

of infection time, all baculovirus differentially expressed genes were found down-regulated in infected adapted cells (**Figure 2C**).

Table 1. List of TOP-20 up-regulated genes in adapted cells vs. non-adapted cells.

Gene ID	Gene name	hpi	FC	logCPM	Biological Process (P) or Molecular Function (M)
LOC113500835	*xaa-Pro aminopeptidase 1*	12	273.6	3.1	hydrolase activity (M)
		24	756.1	1.6	
LOC113493104	*potassium channel subfamily K member 18-like*	12	102.1	1.7	transmembrane transport (P)
		24	433.4	0.8	
LOC113505620	*serine protease HP21 precursor*	12	104.1	-0.1	proteolysis (P)
		24	116.2	-0.9	
LOC113496461	*lipase member H-like*	12	111.6	-1.6	lipid metabolic process (P)
		24	79.4	-1.3	
LOC113506781	*organic cation transporter protein-like*	12	91.9	-0.3	transmembrane transport (P)
		24	92.8	-1.1	
LOC113504337	*GATA zinc finger domain-containing protein 4-like*	12	79.8	-0.5	regulation of transcription (P)
		24	74.3	-1.3	
LOC113500515	*alkylglycerol monooxygenase-like*	12	97.8	1.0	lipid metabolic process (P)
		24	46.6	-0.3	
LOC113501396	*transcription factor glial cells missing 2-like*	12	108.6	-1.7	regulation of transcription (P)
		24	29.1	-2.2	
LOC113501938	*monocarboxylate transporter 1-like*	12	86.2	-1.8	transmembrane transport (P)
		24	36.5	-1.9	
LOC113494762	*octopamine receptor Oamb isoform X1*	12	65.3	1.1	regulation of transcription (P)
		24	18.4	0.7	
LOC113495982	uncharacterized protein	12	7.7	-1.0	n.a.
		24	74.8	-1.3	
LOC113501619	*organic cation transporter protein-like*	12	43.4	-2.5	transmembrane transport (P)
		24	24.1	-2.3	
LOC113506861	uncharacterized protein	12	37.4	-2.6	n.a.
		24	24.0	-2.3	
LOC113492056	*glucose dehydrogenase [FAD, quinone]-like*	12	34.6	-2.7	oxidoreductase activity (M)
		24	26.4	-2.2	
LOC113495496	*thyrotropin-releasing hormone receptor-like isoform X1*	12	3.8	-0.2	signaling (P)
		24	51.7	1-6	
LOC113500144	uncharacterized protein	12	39.2	-0.2	n.a.
		24	14.2	-1.6	
LOC113506931	*proton-coupled amino acid transporter-like protein pathetic*	12	36.3	-1.4	transmembrane transport (P)
		24	6.1	-1.2	
LOC113500825	*acid sphingomyelinase-like phosphodiesterase 3a*	12	4.5	-0.9	hydrolase activity (M)
		24	34.2	-2.1	
LOC113498260	*mitochondrial import receptor subunit TOM40 homolog*	12	29.8	0.3	transmembrane transport (P)
		24	7.7	-0.6	
LOC113500192	*glutamate receptor ionotropic, kainate 2-like*	12	18.8	3.0	regulation of transcription (P)
		24	18.4	2.2	

Table 2. List of TOP-20 down-regulated genes in adapted cells vs. non-adapted cells.

Gene ID	Gene name	hpi	FC	logCPM	Biological Process (P) or Molecular Function (M)
LOC113496224	*cuticle protein 16.5-like*	12	168.3	-1.0	structural activity (M)
		24	84.3	-1.4	
LOC113492802	*juvenile hormone esterase-like*	12	11.5	-0.1	hydrolase activity (M)
		24	165.3	-0.5	
LOC113495404	*cytochrome P450 9e2-like*	12	163.0	-1.1	oxidoreductase activity (M)
		24	13.6	-1.3	
LOC113496454	*homeobox protein abdominal-A homolog isoform X1*	12	41.7	0.7	programmed cell death (P)
		24	120.1	-0.9	
LOC113508341	uncharacterized protein	12	31.2	1.2	n.a.
		24	83.5	0.2	
LOC113498352	*apolipophorins isoform X2*	12	16.2	1.3	lipid transport (P)
		24	49.9	0.2	
LOC113493958	*solute carrier family 15 member 1-like*	12	36.2	-1.3	transmembrane transport (P)
		24	26.3	-2.3	
LOC113496890	*acanthoscurrin-1-like*	12	25.3	-1.7	n.a.
		24	34.5	-2.1	
LOC113501789	uncharacterized protein	12	19.4	2.7	structure development (P)
		24	31.2	1.3	
LOC113499613	*gloverin-like*	12	21.3	2.9	n.a.
		24	26.4	1.5	
LOC113503497	*chondroitin proteoglycan 2-like*	12	14.2	-0.7	chitin binding (M)
		24	31.3	-1.0	
LOC113496839	*protein toll-like*	12	36.6	0.3	immune system process (P)
		24	8.8	-1.3	
LOC113492804	*esterase FE4-like*	12	29.1	1.3	hydrolase activity (M)
		24	15.1	0.5	
LOC113494776	*myb-like protein D*	12	7.0	-2.3	n.a.
		24	34.2	-2.1	
LOC113494175	*alpha-tocopherol transfer protein-like*	12	12.7	-1.0	n.a.
		24	25.3	-2.3	
LOC113503288	*acetylcholine receptor subunit alpha-like 1*	12	23.4	-0.3	signaling (P)
		24	14.4	-0.8	
LOC113507593	*probable E3 ubiquitin-protein ligase bre1 isoform X1*	12	6.1	-1.0	n.a.
		24	30.2	-1.0	
LOC113503500	*odorant receptor 67c-like*	12	4.7	-1.8	signaling (P)
		24	31.3	-2.2	
LOC113505115	*irregular chiasm C-roughest protein-like isoform X1*	12	16.5	-0.3	programmed cell death (P)
		24	15.3	-1.7	
LOC113506962	*neural retina-specific leucine zipper protein-like*	12	25.1	2.3	regulation of transcription (P)
		24	5.5	1.1	

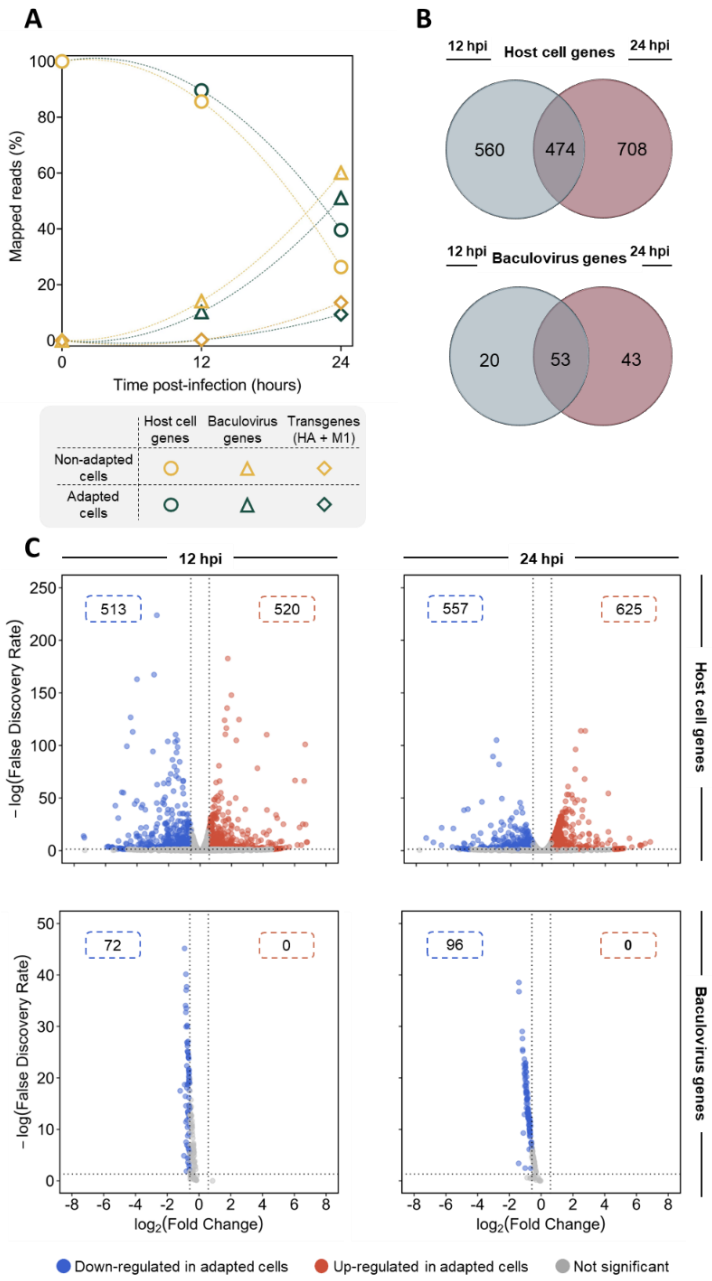


Figure 2. Differential gene expression analysis. A) Percentage of reads mapping to host cell (circles), baculovirus (triangles) and transgenes (diamonds) throughout infection. B) Venn diagram showing the number of host cell and baculovirus genes being differentially expressed (|log2(Fold-Change| > 0.58 and FDR < 0.05) between adapted and non-adapted cells, at 12 hpi, 24 hpi, and both timepoints simultaneously. C) Volcano plot showing the distribution of differentially expressed genes with higher (in red) or lower (in blue) expression in adapted cells, comparing to non-adapted cells. Number in boxes: number of differentially expressed genes.

3.3. Pathway enrichment analysis

To further understand the biological mechanisms behind adapted cells' higher productivity, pathway enrichment analysis was performed. KEGG analysis using Fisher´s exact test revealed that 14 pathways are enriched, two being upregulated in adapted cells (i.e., neuroactive ligand-receptor interaction, ribosome) while remaining 12 were down-regulated (including metabolism of xenobiotic and endogenous compounds, carbohydrates, amino acids, vitamins and lipids) (**Figure 3A**). Additional pathways were found enriched using the GSEA method, including those playing a role in protein processing (including proteasome and ubiquitin mediated proteolysis) and oxidative phosphorylation (**Supplementary Figure S3**). For the pathways found to be enriched, heatmaps with hierarchical clustering of all genes (differentially expressed and not) were generated. Results suggest that there is no common trend in adapted and non-adapted samples clustering as infection progresses, as reflected in the different dendrograms obtained - see two examples in **Figure 3B**.

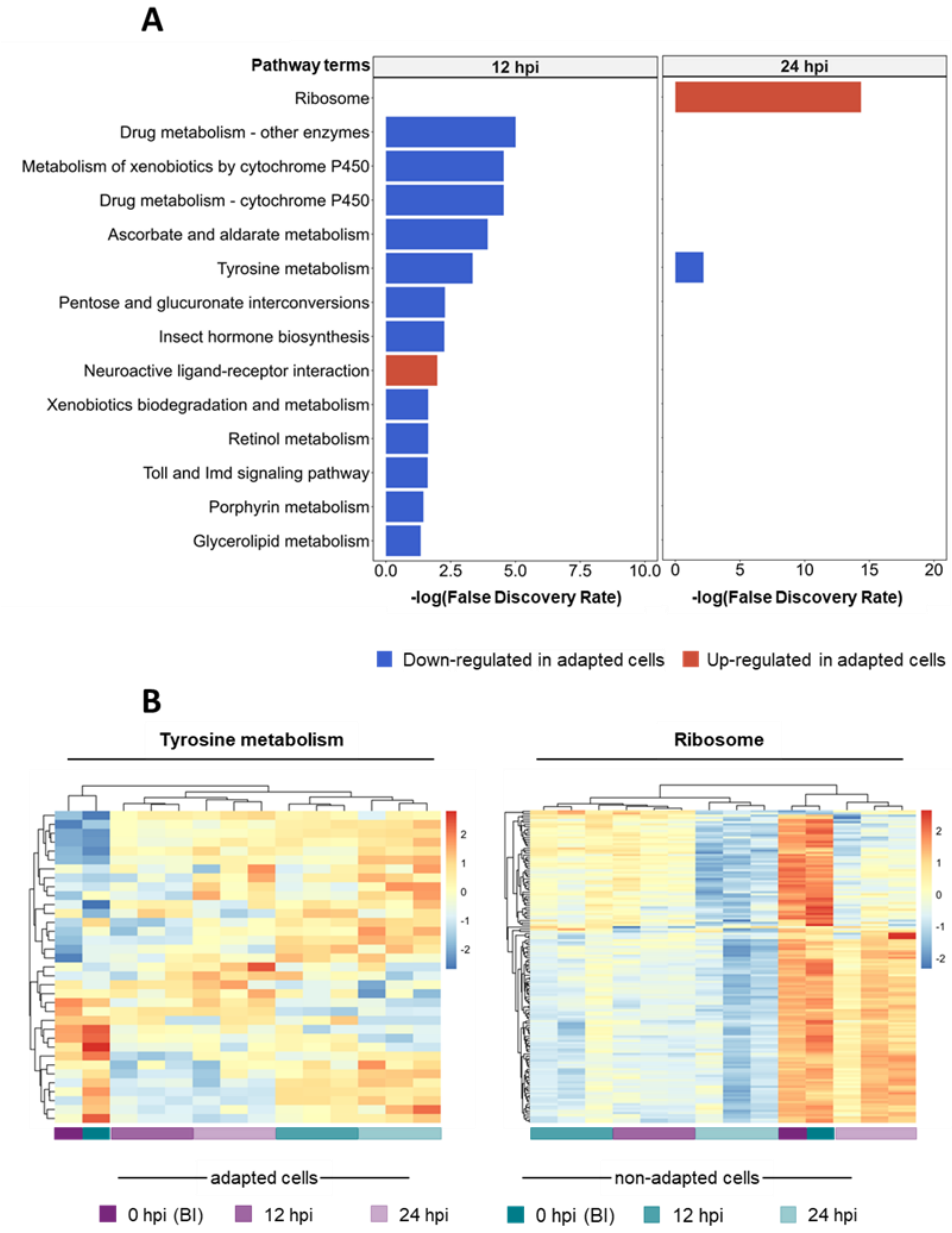


Figure 3. Pathway enrichment analysis. A) Pathway enrichment analysis of differentially expressed genes using Fisher´s exact test. Bar plots indicate enriched terms at 12 and 24 hpi (i.e. longer bars denote pathways more significantly enriched). Color code: red identifies pathways found up-regulated and blue identifies pathways found down-regulated in adapted cells, over non-adapted cells. B) Heat map of tyrosine metabolism pathway and genes encoding for ribosome subunits; the z-score (gradient color code) is defined for all the genes being found involved in these specific pathways (differentially expressed or not). BI denotes before infection.

4. Discussion

In this work, we assessed the gene expression profile of insect High Five cells adapted to neutral pH during production of influenza HA-VLPs using IC-BEVS, and compared it to that of non-adapted cells.

During infection, the number of reads mapping to the host cell genome decreased whereas those aligned to baculovirus (*Ac*MNPV) and transgenes (M1 and HA) sequences increased. Such trend is a consequence of the global takeover of the cellular transcription machinery by baculoviruses towards overproduction of viral proteins during infection (Monteiro et al., 2012). The percentages of reads aligned to host cells and virus genome throughout the course of infection is similar to those reported previously in other studies (Y.-R. Chen et al., 2014; Koczka et al., 2018). Differential expression analysis revealed that all baculovirus-derived genes were found down-regulated in adapted cells. Together with distinct onset cell viability drop, this suggests that adapted and non-adapted cells have different susceptibility to infection.

Adapted and non-adapted cells respond differently to baculovirus infection as reflected by the major differences observed in their gene expression profiles. For instance, genes involved in lipid metabolism processes (i.e., biosynthesis, hydrolysis and transport) were found either up- or down-regulated in adapted cells at both 12 and 24 hpi. Insect cells are known to have a limited lipid metabolism, reflected by their limited capacity in synthesizing, desaturating and elongating fatty acids (Drugmand et al., 2012a), and lipid deprivation is linked to cell degeneration and impairment in the production of baculovirus (Goodwin, 1991). Moreover, lipids such as cholesterol are especially important during the production of enveloped viral particles in mammalian culture systems (Chan et al., 2010) and IC-BEVS (Monteiro et al., 2016), such as the case of the influenza HA-VLPs herein produced. Therefore, distinct lipid metabolism could be associated to different productivity of adapted and non-adapted cells.

Pathway enrichment analysis revealed that genes encoding ribosome subunits were up-regulated in adapted cells at 24 hpi. Since baculovirus is known to promote the shutdown of host cell protein synthesis upon infection for overproduction of viral proteins (X. Du & Thiem, 1997; Nayyar et al., 2017), this result suggests that the host cell translation machinery was less impaired in adapted cells, thereby leading to higher protein biosynthesis (including of HA-VLPs). The oxidative phosphorylation pathway was also found up-regulated in adapted cells at 24 hpi, allowing lower substrate consumption when compared to non-adapted cells. In contrast, pathways associated to drug, carbohydrate and amino acid metabolism were down-regulated in adapted cells. Interestingly, the set of genes driving such enrichment is shared between most pathways; these code for proteins such as UPD-glucuronosyltransferases-like and glutathione S-transferase-like proteins, enzymes associated to cellular protection and resistance to oxidative stress (Kalthoff et al., 2010). In addition to these, genes involved in oxidative metabolism were also down-regulated in adapted cells, some of which have been already identified as differentially expressed in insect cells resistant to harmine and fungi (e.g. *cytochromes 450*) (Cui et al., 2020; Xing et al., 2017). Pathways associated to baculovirus infection such as immune response, protein processing in endoplasmic reticulum, proteosome and ubiquitin mediated proteolysis, which are known to be up-regulated upon infection (Marques & Imler, 2016; Monteiro et al., 2012), were also found down-regulated in adapted cells. Taken together, these results suggest that adapted cells better cope with the stress induced by baculovirus infection when compared to non-adapted cells.

The major *Ac*MNPV fusion protein, GP64, plays an essential role in mediating virus-receptor binding, internalization, and membrane fusion during virus entry into both mammalian and insect cells (Kataoka et al., 2012). The fusogenicity of GP64 is low-pH dependent, and fusion of viral envelope with the endosomal

membrane is triggered in the acidic endosomal lumen (Blissard et al., 1992). Virus fusion was shown to be impaired at relatively high-pH conditions in mammalian cells (Hu et al., 2019); thus, baculovirus entry could be less efficient in adapted cells. In this regard, the baculovirus gene *ACNVgp64* was found significantly down-regulated in adapted cells at 12 hpi (1.7-fold) and 24 hpi (2.0-fold). While the amount of differentially expressed host cell genes in adapted cells was equally distributed between those being up- or down-regulated, all the differentially expressed baculovirus-derived genes herein identified were shown to be down-regulated in adapted cells, hence showing that virus transcripts are produced earlier (or at higher quantity) in non-adapted cells. Whether this outcome is a consequence of less efficient virus entry in adapted cells still remains unknown, but such fact could be behind the different cell growth kinetics observed for adapted cells, i.e., slight increase in cell concentration and delayed onset of cell viability drop. The apparent lower burden caused to adapted cells at initial stages of infection may have been the key to achieve prolonged infection and consequently higher productivity.

5. Conclusion

In this study, comparative transcriptome analysis revealed significant differences between adapted and non-adapted insect High Five cells during production of influenza HA-VLPs using IC-BEVS. Differential gene expression analysis showed baculovirus genes being down-regulated in adapted cells, revealing less susceptibility to infection. Several pathways were found enriched and differently regulated, such as those associated to protein synthesis, metabolism of xenobiotic and endogenous compounds, carbohydrates and amino acids. The gene expression signatures herein identified can be exploited for rational genetic engineering of insect cells and/or baculovirus to further improve production yields. Furthermore, single-cell RNA sequencing could help

to further understand the molecular signatures playing a role on systems' productivity, as well as to conclude on adapted cell population heterogeneity.

6. Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 813453 and by Fundação para a Ciência e a Tecnologia/Ministério da Ciência, 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 like to acknowledge Pedro Barros for the support in genome annotation using OmicsBox.

7. Supplementary material

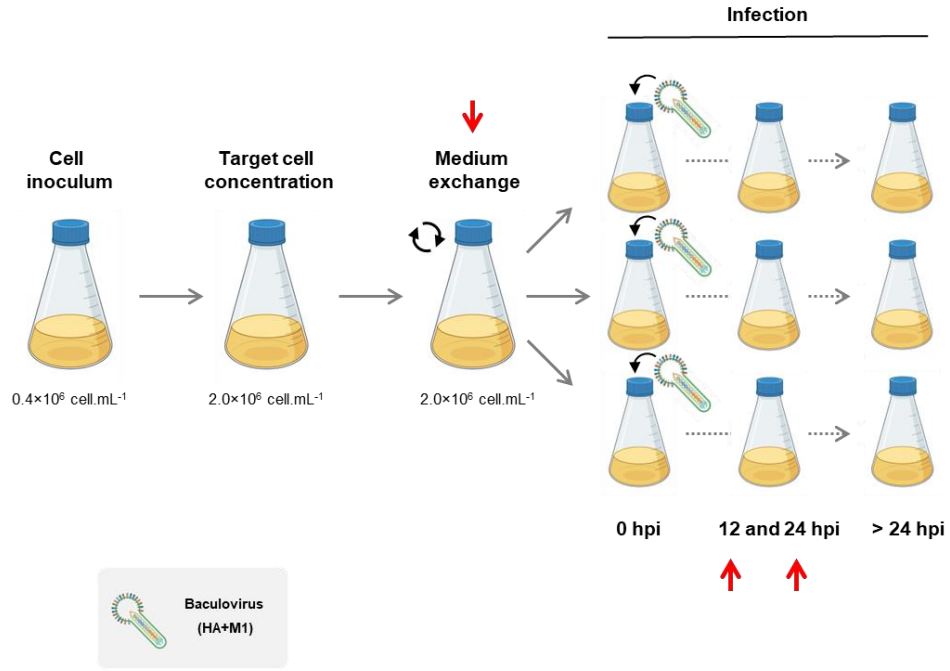


Figure S1. Experimental design. Cells were inoculated at 0.4×10^6 cell.mL^{-1} and grown until the target cell concentration of 2.0×10^6 cell.mL^{-1}. Medium exchange was performed before the infection (using centrifugation), and cells divided in three shake flasks. The red arrows highlight time-points at which samples were collected for total RNA extraction. hpi denoted hours post-infection.

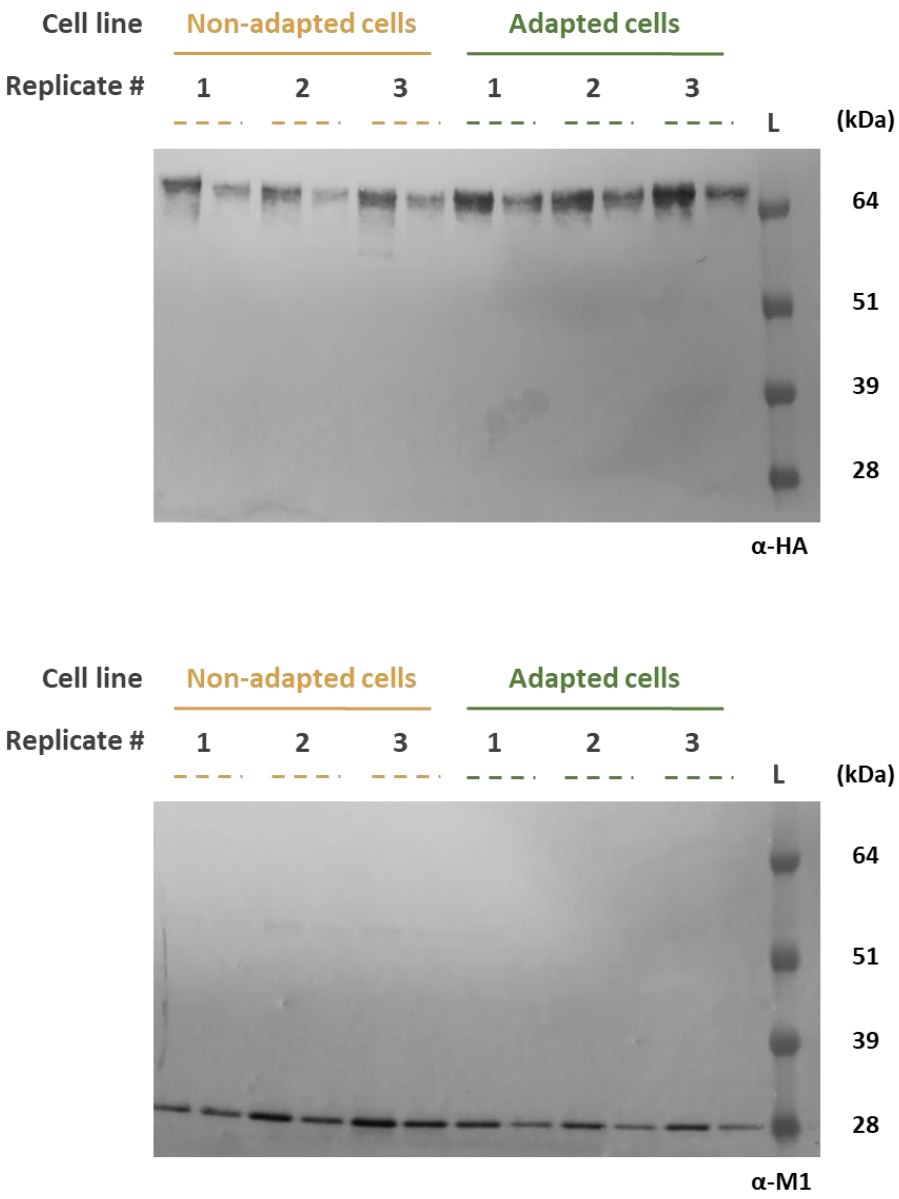


Figure S2. Identification of HA and M1 proteins by western blot. Two different dilutions of in-process samples collected from each cell culture (in triplicate) at the time of harvest are represented.

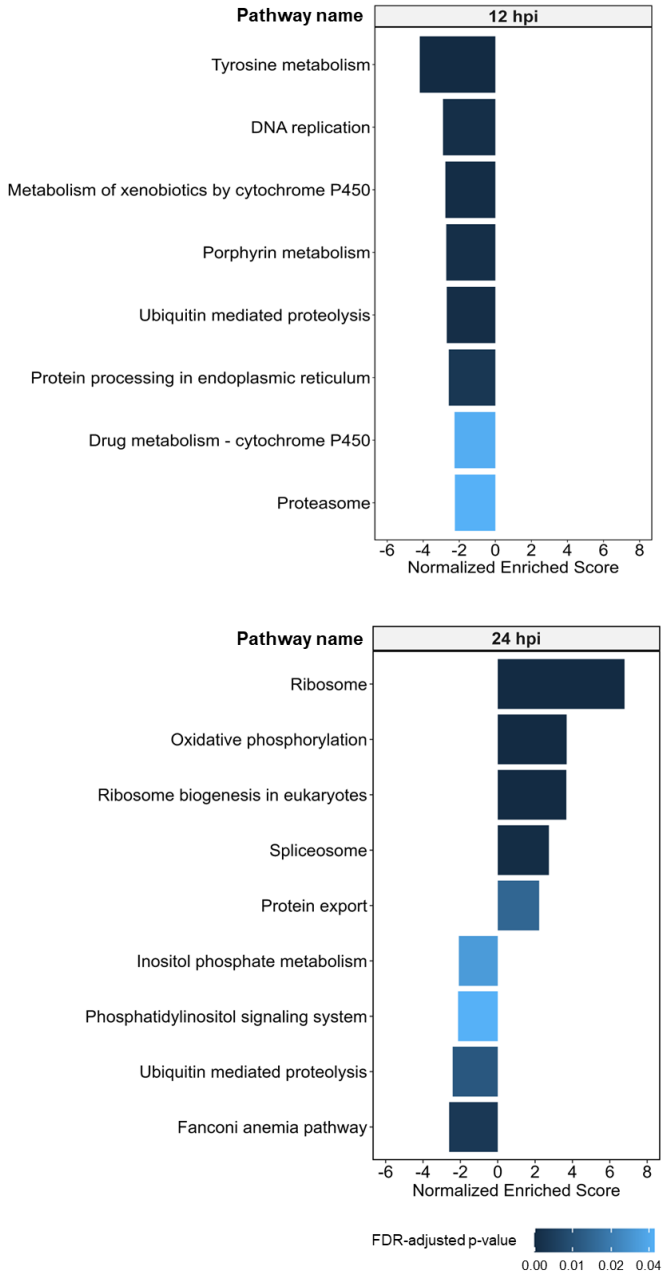


Figure S3. Pathway enrichment analysis using the GSEA method. Barplots show enriched terms at 12 hpi and 24 hours post-infection (hpi). Color gradient of bars indicate the False Discovery Rate (FDR)-adjusted p-value.

8. References

Abaandou, L., Quan, D., & Shiloach, J. (2021). Affecting hek293 cell growth and production performance by modifying the expression of specific genes. *Cells*, *10*(7). https://doi.org/10.3390/cells10071667

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M., & Possee, R. D. (1994). The Complete DNA Sequence of Autographa californica Nuclear Polyhedrosis Virus. *Virology*, *202*(2), 586–605. https://doi.org/10.1006/viro.1994.1380

Babar, M. M., Afzaal, H., Pothineni, V. R., Zaidi, N. U. S. S., Ali, Z., Zahid, M. A., & Gul, A. (2018). Omics approaches in industrial biotechnology and bioprocess engineering. *Omics Technologies and Bio-engineering,* 251–269. https://doi.org/10.1016/B978-0-12-815870-8.00014-0

Benjaminit, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol, 57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Blissard, G. W., & Wenz, J. R. (1992). Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. *Journal of virology*, *66*(11), 6829–6835. https://doi.org/10.1128/JVI.66.11.6829-6835.1992

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Chan, R. B., Tanner, L., & Wenk, M. R. (2010). Implications for lipids during replication of enveloped viruses. *Chemistry and physics of lipids*, *163*(6), 449–459. https://doi.org/10.1016/j.chemphyslip.2010.03.002

Chen, Y.-R., Zhong, S., Fei, Z., Gao, S., Zhang, S., Li, Z., Wang, P., & Blissard, G. W. (2014). Transcriptome Responses of the Host Trichoplusia ni to Infection by the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus. *Journal of Virology*, *88*(23), 13781–13797. https://doi.org/10.1128/jvi.02243-14

Chen, Y.-R., Zhong, S., Fei, Z., Hashimoto, Y., Xiang, J. Z., Zhang, S., & Blissard, G. W. (2013). The Transcriptome of the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus in Trichoplusia ni Cells. *Journal of Virology*, *87*(11), 6391–6405. https://doi.org/10.1128/jvi.00194-13

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Correia, R., Fernandes, B., Alves, P. M., Carrondo, M. J. T., & Roldão, A. (2020). Improving influenza HA-VLPs production in insect high five cells via adaptive laboratory evolution. *Vaccines*, *8*(4), 1–15. https://doi.org/10.3390/vaccines8040589

Cui, G., Sun, R., Veeran, S., Shu, B., Yuan, H., & Zhong, G. (2020). Combined transcriptomic and proteomic analysis of harmine on Spodoptera frugiperda Sf9 cells to reveal the potential resistance mechanism. *Journal of Proteomics*, *211*. https://doi.org/10.1016/j.jprot.2019.103573

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Drugmand, J. C., Schneider, Y. J., & Agathos, S. N. (2012). Insect cells as factories for biomanufacturing. *Biotechnology advances*, *30*(5), 1140–1157. https://doi.org/10.1016/j.biotechadv.2011.09.014

Du, X., & Thiem, S. M. (1997). Responses of insect cells to baculovirus infection: protein synthesis shutdown and apoptosis. *Journal of virology*, *71*(10), 7866–7872. https://doi.org/10.1128/JVI.71.10.7866-7872.1997

Fernandes, B., Vidigal, J., Correia, R., Carrondo, M. J. T., Alves, P. M., Teixeira, A. P., & Roldão, A. (2020). Adaptive laboratory evolution of stable insect cell lines for improved HIV-Gag VLPs production. *Journal of Biotechnology*, *307*, 139–147. https://doi.org/10.1016/j.jbiotec.2019.10.004

Fu, Y., Yang, Y., Zhang, H., Farley, G., Wang, J., Quarles, K. A., Weng, Z., & Zamore, P. D. (2018). The genome of the Hi5 germ cell line from Trichoplusia ni, an agricultural pest and novel model for small RNA biology. *ELife*, *7*. https://doi.org/10.7554/eLife.31628

Goodwin R. H. (1991). Replacement of vertebrate serum with lipids and other factors in the culture of invertebrate cells, tissues, parasites, and pathogens. *In vitro cellular & developmental biology : journal of the Tissue Culture Association*, *27A*(6), 470–478. https://doi.org/10.1007/BF02631147

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. https://doi.org/10.1093/nar/gkn176

Hong, J. K., Lakshmanan, M., Goudar, C., & Lee, D.-Y. (2018). Towards next generation CHO cell line development and engineering by systems approaches. *Current Opinion in Chemical Engineering*, *22*, 1–10. https://doi.org/10.1016/j.coche.2018.08.002

Hu, L., Li, Y., Ning, Y.-J., Deng, F., Vlak, J. M., Hu, Z., Wang, H., & Wang, M. (2019). The Major Hurdle for Effective Baculovirus Transduction into Mammalian Cells Is

Passing Early Endosomes. *Journal of Virology*, *93*(15). https://doi.org/10.1128/jvi.00709-19

J. R. Tennant. (1964). Evaluation of the trypan blue technique for determination of cell viability. *Transplantation*, *2*, 685–694. https://doi.org/10.1097/00007890-196411000-00001

Kalthoff, S., Ehmer, U., Freiberg, N., Manns, M. P., & Strassburg, C. P. (2010). Interaction between oxidative stress sensor Nrf2 and xenobiotic-activated aryl hydrocarbon receptor in the regulation of the human phase II detoxifying UDP-glucuronosyltransferase 1A10. *Journal of Biological Chemistry*, *285*(9), 5993–6002. https://doi.org/10.1074/jbc.M109.075770

Kataoka, C., Kaname, Y., Taguwa, S., Abe, T., Fukuhara, T., Tani, H., Moriishi, K., & Matsuura, Y. (2012). Baculovirus GP64-Mediated Entry into Mammalian Cells. *Journal of Virology*, *86*(5), 2610–2620. https://doi.org/10.1128/jvi.06704-11

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. https://doi.org/10.1016/j.jbiotec.2018.02.001

Kost, T. A., Condreay, J. P., & Jarvis, D. L. (2005). Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nature biotechnology*, *23*(5), 567–575. https://doi.org/10.1038/nbt1095

Kost, T. A., & Kemp, C. W. (2016). Fundamentals of Baculovirus Expression and Applications. *Advances in experimental medicine and biology*, *896*, 187–197. https://doi.org/10.1007/978-3-319-27216-0_12

Lin, C. Y., Huang, Z., Wen, W., Wu, A., Wang, C., & Niu, L. (2015). Enhancing protein expression in HEK-293 cells by lowering culture temperature. *PLoS ONE*, *10*(4). https://doi.org/10.1371/journal.pone.0123562

Marques, J. T., & Imler, J. L. (2016). The diversity of insect antiviral immunity: insights from viruses. *Current opinion in microbiology*, *32*, 71–76. https://doi.org/10.1016/j.mib.2016.05.002

Martínez-Solís, M., Herrero, S., & Targovnik, A. M. (2019). Engineering of the baculovirus expression system for optimized protein production. *Applied microbiology and biotechnology*, *103*(1), 113–123. https://doi.org/10.1007/s00253-018-9474-7

Mena, J. A., Ramírez, O. T., & Palomares, L. A. (2003). Titration of Non-Occluded Baculovirus Using a Cell Viability Assay. *BioTechniques*, *34*(2), 260–264. https://doi.org/10.2144/03342bm05

Monteiro, F., Bernal, V., Chaillet, M., Berger, I., & Alves, P. M. (2016). Targeted supplementation design for improved production and quality of enveloped viral particles in insect cell-baculovirus expression system. *Journal of Biotechnology*, *233*, 34–41. https://doi.org/10.1016/j.jbiotec.2016.06.029

Monteiro, F., Carinhas, N., Carrondo, M. J., Bernal, V., & Alves, P. M. (2012). Toward system-level understanding of baculovirus-host cell interactions: from molecular fundamental studies to large-scale proteomics approaches. *Frontiers in microbiology*, *3*, 391. https://doi.org/10.3389/fmicb.2012.00391

Nayyar, N., Kaur, I., Malhotra, P., & Bhatnagar, R. K. (2017). Quantitative proteomics of Sf21 cells during Baculovirus infection reveals progressive host proteome changes and its regulation by viral miRNA. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-10787-z

Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E., & Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics*, *38*(10), 2943–2945. https://doi.org/10.1093/bioinformatics/btac166

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, *11*(3), R25. https://doi.org/10.1186/gb-2010-11-3-r25

Robinson, M. D., & Smyth, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, *9*(2), 321–332. https://doi.org/10.1093/biostatistics/kxm030

Roldão, A., Oliveira, R., Carrondo, M. J. T., & Alves, P. M. (2009). Error assessment in recombinant baculovirus titration: Evaluation of different methods. *Journal of Virological Methods*, *159*(1), 69–80. https://doi.org/10.1016/j.jviromet.2009.03.007

Rossi, N., Silva, B. G., Astray, R., Swiech, K., Pereira, C. A., & Suazo, C. A. T. (2012). Effect of hypothermic temperatures on production of rabies virus glycoprotein by recombinant Drosophila melanogaster S2 cells cultured in suspension. *Journal of Biotechnology*, *161*(3), 328–335. https://doi.org/10.1016/j.jbiotec.2012.05.016

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Seo, J. S., Kim, Y. J., Cho, J. M., Baek, E., & Lee, G. M. (2013). Effect of culture pH on recombinant antibody production by a new human cell line, F2N78, grown in suspension at 33.0 C and 37.0 C. *Applied Microbiology and Biotechnology*, *97*(12), 5283–5291. https://doi.org/10.1007/s00253-013-4849-2

Shin, H. Y., Choi, H., Kim, N., Park, N., Kim, H., Kim, J., & Kim, Y. B. (2020). Unraveling the genome-wide impact of recombinant baculovirus infection in mammalian cells for gene delivery. *Genes*, *11*(11), 1–12. https://doi.org/10.3390/genes11111306

Reuveny, S., Kim, Y. J., Kemp, C. W., & Shiloach, J. (1993). Effect of temperature and oxygen on cell growth and recombinant protein production in insect cell cultures. *Applied microbiology and biotechnology*, *38*(5), 619–623. https://doi.org/10.1007/BF00182800

van Oers, M. M., Pijlman, G. P., & Vlak, J. M. (2015). Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *The Journal of general virology*, *96*(Pt 1), 6–23. https://doi.org/10.1099/vir.0.067108-0

Vieira, H. L. A., Estêvão, C., Roldão, A., Peixoto, C. C., Sousa, M. F. Q., Cruz, P. E., Carrondo, M. J. T., & Alves, P. M. (2005). Triple layered rotavirus VLP production: Kinetics of vector replication, mRNA stability and recombinant protein production. *Journal of Biotechnology*, *120*(1), 72–82. https://doi.org/10.1016/j.jbiotec.2005.03.026

Wang, X. Y., Yu, H. Z., Geng, L., Xu, J. P., Yu, D., Zhang, S. Z., Ma, Y., & Fei, D. Q. (2016). Comparative transcriptome analysis of Bombyx mori (Lepidoptera) larval midgut response to BmNPV in susceptible and near-isogenic resistant strains. *PLoS ONE*, *11*(5). https://doi.org/10.1371/journal.pone.0155341

Xing, D., Yang, Q., Jiang, L., Li, Q., Xiao, Y., Ye, M., & Xia, Q. (2017). RNA-seq analyses for two silkworm strains reveals insight into their susceptibility and resistance to Beauveria bassiana infection. *International Journal of Molecular Sciences*, *18*(2). https://doi.org/10.3390/ijms18020234

Chapter 3

Evaluation of insect cell heterogeneity during influenza VLP production using single-cell RNA-seq

This chapter was based on the following manuscript:

Silvano, M., Correia, R., Virgolini, N., Clarke, C., Alves P.M., Isidro I.A., Roldão, A. **2023**. Dissecting insect cell heterogeneity during influenza VLP production using single-cell transcriptomics. *Frontiers in Biotechnology and Bioengineering*.

Author contribution

Marco Silvano design and performed the experiments, analysed the data and wrote the chapter

Abstract

The insect cell-baculovirus expression vector system (IC-BEVS) has been widely used to produce recombinant proteins, including complex virus-like particles (VPLs). However, cell-to-cell variability upon infection is yet one of the least understood phenomena in virology, and little is known about its impact on production and quality of therapeutic proteins. This study aimed at providing the first insight on cell population heterogeneity during production of influenza VLPs in IC-BEVS using single-cell RNA-seq (scRNA-seq). High Five cell population was shown to be heterogeneous even before infection, with cell cycle being one of the factors contributing for this variation. In addition, infected insect cells were clustered according to the timing and level of baculovirus genes expression, with each cluster reporting similar influenza VLPs transgenes (i.e. HA and M1) transcript counts. Trajectory analysis enabled to track infection progression throughout pseudotime. Specific pathways such as translation machinery, protein folding, sorting and degradation, endocytosis and energy metabolism were identified as being those which vary the most during insect cell infection and production of Influenza VLPs.

Overall, this study lays the ground for the application of scRNA-seq in IC-BEVS processes to isolate relevant biological mechanisms during recombinant protein expression towards its further optimization.

TABLE OF CONTENTS

1. Introduction

The insect cells (IC) and baculovirus expression vector system (BEVS) constitute an attractive alternative to mammalian cells for manufacturing of heterologous gene products, including recombinant proteins as vaccine candidates and viral vectors for gene therapy (Drugmand et al., 2012b). Recent advances in next-generation sequencing technologies have enabled a considerable improvement of our understanding of the IC-BEVS. For example, RNA-seq has been used to assess the transcriptional changes of alphanodavirus-free High Five cells upon infection by *Autographa californica* multiple nucleopolyhedrovirus (*Ac*MNPV) (Y.-R. Chen et al., 2013) (Y.-R. Chen et al., 2014), providing a global picture of the *Ac*MNPV transcription regulation throughout the infection cycle. As the knowledge of the IC-BEVS grows, potential engineering targets to increase recombinant protein production are being identified (Silvano et al., 2022).

In recent years, technological advances in areas such as cell isolation methods using microfluidics or microwell devices, preparation of next generation sequencing libraries from ultra-low quantities of nucleic acids, and innovative labelling strategies for MS-based proteomics have enabled the characterization of DNA, RNA and proteins at single-cell resolution (Lee et al., 2020). Using different single-cell omics profiling strategies as building blocks, we can now build a multi-omics profile of the same cell. These multi-omics methods will play an important role in many diverse fields, and their applications are rapidly expanding, including delineating cellular diversity (Gulati et al., 2020), lineage tracing (D. E. Wagner & Klein, 2020), identifying new cell types (Ianevski et al., 2022), and deciphering the regulatory mechanisms between omics (Z. J. Cao & Gao, 2022). Single-cell analysis has unambiguously demonstrated that cell populations are often heterogeneous (Goldman et al., 2019). This heterogeneity not only applies to different cell types in a tissue (Karlsson et al., 2021) but also to clonal cell population (Stockholm et al., 2007).

Single-cell RNA sequencing (scRNA-seq) has just recently be applied to virus-based processes. The power of scRNA-seq lies in the simultaneous delivery of snapshots of virus and host transcriptomes, and allows to compare host transcriptome between cells with low and high viral loads (Suomalainen & Greber, 2021). The high-resolution dissection of viral and host cell gene expression patterns reveals that the transcriptional responses of individual infected cells can be divergent, as the interplay between underlying cellular heterogeneity and viral population diversity influences the fate of infection (J. Sun et al., 2020b). For example, cell-to-cell variation in viral transcription has been observed during influenza virus infection in mammalian (A549, MDCK and HEK293) cells (Russell et al., 2018).

To date, the understanding of the IC-BEVS transcriptome has been mostly relying on bulk RNA-seq analysis (Y.-R. Chen et al., 2014; Nguyen et al., 2013; Silvano et al., 2022; Virgolini et al., 2022). For instance, we have previously assessed whole transcriptome changes in High Five insect cells during expression of influenza HA-displaying virus-like particles (HA-VLPs) using IC-BEVS, which enabled to identify key biological processes impacted by virus infection (Silvano et al., 2022). Although these studies uncover transcriptional changes in insect cell response to baculovirus infection, they only provide rough models of the host cell response. Understanding IC-BEVS at the single-cell level could elucidate better the mechanisms of viral infection and potentially enable to identify, within a potentially heterogenous cell population and infection process, the characteristics of cells associated with a more efficient progression of infection and production of heterologous proteins.

In this study, we used scRNA-seq for the characterization of the High Five insect cell line during production of influenza HA-VLPs using IC-BEVS. The transcriptomics pipeline here described allowed to study, at a single-cell level, High Five cell population heterogeneity (prior and during infection), host cell

response to virus infection, and progression of infection (expression of virus genes and transgenes encoding HA-VLPs).

2. Materials and methods

2.1. Cell line and culture media

High Five insect cells (Invitrogen) were routinely sub-cultured to 0.3-0.5×10^6 cell.mL^{-1} every 2-3 days when cell concentration reached 2-3×10^6 cell.mL^{-1} in serum-free Insect-XPRESS™ medium (Sartorius) using 125-500 mL shake flasks with a 10% working volume, and maintained at 27ºC in a Inova 44R shaking incubator (orbital motion diameter of 2.54 cm - Eppendorf) operating at 100 RPM.

2.2. Baculovirus amplification and storage

Recombinant baculoviruses carrying influenza capsid M1 from A/California/06/2009 H1N1 strain and hemagglutinin (HA) from A/Brisbane/59/2007 strain genes were kindly provided by Redbiotec AG (Schlieren, Switzerland). Baculovirus amplification was performed as described elsewhere (Vieira et al., 2005).

2.3. Production of HA-displaying VLPs

HA-VLPs production was carried out in a 0.5 L stirred tank bioreactor (BIOSTAT Qplus – Sartorius) as specified elsewhere (Correia et al., 2020). Cells were expanded in 500-2000 mL shake flasks with a 10% working volume as described above. Infection experiments were performed in bioreactor at cell concentration at the time of infection (CCI) of 2×10^6 cell.mL^{-1} and multiplicity of infection (MOI) of 1 pfu.$cell^{-1}$. Medium exchange was performed at the time of

infection by centrifugation at 200g at room temperature for 10 min. Samples were taken every 24 hours for the assessment of cell concentration and viability, and detection of M1 and HA proteins; for scRNA-seq, samples were taken before infection, and at 8 and 22 hours post-infection (hpi).

2.4. Purification of HA-displaying VLPs

Culture bulk from bioreactor run was harvested 3 days post-infection and centrifuged at (first) 4ºC, 200 g, for 10 min and (second) 4 ºC, 2000 g, for 20 min. The supernatant was filtered using a 0.22 µm Stericup (Millipore), and the HA-VLPs were purified using a SartoBind Q capsule (Sartorius Stedim Biotech) according to manufacturer's instructions. Purified material was formulated in 50 mM HEPES, 300 mM NaCl, 15% (w/v) trehalose, pH 7.4, and stored at -80 or 4°C.

2.5. Analytics

2.5.1. Cell concentration and viability

Cell concentration was determined in a Fuchs-Rosenthal hemocytometer chamber (Brand) and cell viability assessed by trypan blue exclusion method (J R Tennant, 1964).

2.5.2. Baculovirus titration

Baculovirus titers were determined using the MTT assay as described elsewhere (Mena et al., 2003; Roldão et al., 2009).

2.5.3. Western blot

Identification and relative quantification of M1 and HA in culture supernatant were performed as reported elsewhere (Correia et al., 2020).

2.5.4. Transmission electron microscopy

Negative staining transmission electron microscopy was used to assess the conformation and size of HA-VLPs. Briefly, 5 µl of purified VLP sample was fixed for 2 min in a copper grid coated with Formvar-carbon (Electron Microscopy Sciences, Hatfield). Grids were washed with H2O and then stained with 2% (v/v) uranyl acetate for 5 min and left to air dry. Samples were then observed in a Hitachi H-7650 Transmission Electron Microscope (JEOL, USA).

2.6. Single-cell RNA sequencing

For single-cell gene expression profiling, » 6000 cells (at 0 hpi and 8 hpi) or » 8000 cells (at 22 hpi) were loaded into a BD Rhapsody cartridge (BD Biosciences) and libraries were generated according to BD Rhapsody™ System mRNA Whole Transcriptome Analysis (WTA). Upon confirming the quality of the resulting libraries using a Bioanalyser, the quantity of each library was determined using Qubit. scRNA-seq libraries were sequenced using an Illumina NovaSeq (Illumina) configured to yield 150 bp paired end reads.

2.7. Single-cell RNA data analysis

2.7.1. Generation of a UMI count matrix

The cellular barcodes were pre-processed and demultiplexed by the BD Rhapsody WTA bioinformatic workflow (BD Biosciences) on the Seven Bridges Genomics (SBG) cloud platform using default parameters, as reported

elsewhere (Tzani et al., 2021). STAR indexes were generated from the *Trichoplusia ni* (*Tnl*) reference genome (RefSeq assembly accession: GCF_003590095.1) and from *Ac*MNPV (RefSeq assembly accession: GCF_000838485.1, ViralProj14023) (Dobin et al., 2013). Specifically, a hybrid reference genome was used for RNA-seq read mapping using transgenes (M1 and HA) sequences and mtDNA sequence of the *Tnl* reference genome (GenBank accession No. MK714850.1).

2.7.2. Filtering the UMI count matrix

The cell/gene matrices output from the SBG pipeline were imported into the R-4.2.1 Statistical Software Environment and merged to form a single matrix for further analysis. The proportion of unique molecular indexes (UMIs) originating from mtDNA was also determined for each cellular barcode, and cells with > 5% mitochondrial UMIs counts were considered of low-quality and thus removed from further analysis.

2.7.3. UMAP and pseudotime analysis

Seurat v4 was used to apply a graph-based clustering approach (Hao et al., 2021). These methods embed cells in a graph structure with edges drawn between cells with similar feature expression patterns, and then attempt to partition this graph into highly interconnected “communities” (Xu & Su, 2015). To evaluate cell heterogeneity, data sets (0, 8, 22 hpi) were merged prior to global scaling normalization method. Normalized and merged samples were scaled and variations caused by different total UMIs per cell were regressed out. The most variable features were considered for principal component analysis, and 20 principal components were used to perform cluster analysis. The Uniform Manifold Approximation and Projection (UMAP) technique was used to

run non-linear dimensional reduction and to visualize and explore the datasets (McInnes et al., 2018b). Monocle 3 was run to conduct trajectory analysis (Trapnell et al., 2014) and its function graph_test was used to identify genes that change as function of pseudotime. Genes with an average expression change of ≥ 0.5 and *p*-value < 0.05 were considered significant.

2.7.4. Cell cycle correction

Cell cycle scoring function in Seurat v4 was used to determine the likelihood of cells being in either S or G2/M phase, based on reference genes known to play a role in distinct phases of the cell cycle. To conduct this procedure, we mapped the mouse gene list to the Tnl genome to carry out the classification and draw a list of Tnl cell-cycle genes. The resulting scores for S, G2/M and G1 phases were used to regress out the effect of cell cycle in downstream analysis.

2.7.5. Functional annotation and enrichment analysis

For gene annotation, the amino acid sequence of protein-coding genes was used as a query. Blastp search was applied in the NCBI nr protein database using Blast2GO OmicsBox (Götz et al., 2008). No taxonomy filter was applied, and the E-value cutoff was set to 1×10^{-3}. The over-representation of pathways within gene lists found to have a statistically significant association with pseudotime were identified using the Fisher's exact test. Pathway terms with a False Discovery Rate of < 0.05 were considered significant (Benjaminit & Hochberg, 1995).

3. Results

3.1. Production of influenza HA-VLPs

High Five cells were infected at CCI of 2×10^6 cell.mL^{-1} with a MOI of 1 pfu.cell^{-1}, and infection kinetics were assessed throughout (**Figure 1A**). M1 and HA proteins were identified by Western Blot (**Figure 1B**), and particles resembling influenza HA-VLPs, both in size and morphology, detected by TEM (**Figure 1C**). These results are in line with those previously reported (Silvano et al., 2022) and demonstrate that HA-VLPs were successfully produced.

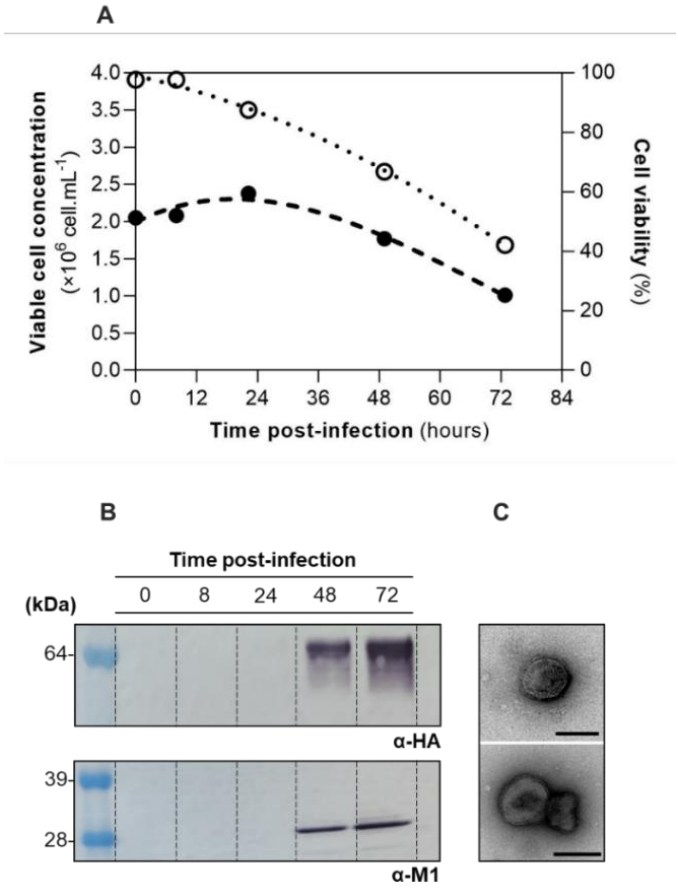
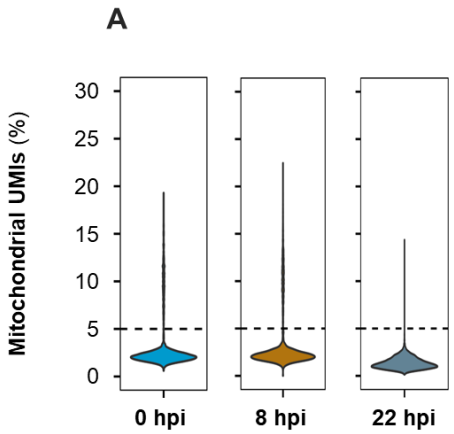


Figure 1. Production of influenza HA-VLPs. A) Cell growth kinetics upon infection. Viable cell concentration and cell viability in full and empty circles, respectively. B) Identification of HA and M1 proteins by Western Blot. C) TEM of purified HA-VLPs; scale bar represents 100 nm.

3.2. Single-cell RNA-seq data processing and quality control

For scRNA-seq, ~ 6,000-8,000 cells were collected at each time point (> 90 % cell viability). After sequencing, an average of 391 million 150 bp-reads were acquired for each sample. Following the completion of this initial pre-processing stage, approx. 4 % of the sequenced RNA-seq reads were removed from further analysis due to insufficient read length, low base quality and/or high single nucleotide frequency; an average of » 377 million reads per time point remained valid for further analysis (**Table S1**). From the reads that passed quality control, 87 % of the reads were successfully assigned to cell barcodes following demultiplexing. Mapping to the reference genome resulted in a unique alignment rate of ~71%. Upon collapsing to UMIs and application of the RSEC algorithm, between 4,496 and 5,408 unique cell barcodes were identified for the three samples taken throughout the process. The mean number of reads and mRNA molecules detected per cell in this experiment were 41,166 and 28,787, respectively, with an average of ~3,372 genes detected in each cell (**Table S1**).

To ensure that only high-quality genes were retained for further analysis, the UMI count matrix was filtered to remove data that might have originated from non-viable cells. An average of 8 % of cells contained > 5 % of detected UMIs originating from mitochondrial genes and thus were eliminated from further analysis (**Figure 2A**). The number of genes per cell and baculovirus UMIs (**Figure 2B**) as well as the amount of total UMIs per cell (**Figure S1**) were assessed; while at 0 and 8 hpi the average number of genes identified per cell is 4000-5000, this number decreases significantly at later stages of infection (i.e. 22 hpi) concomitantly with an increase in the percentage of baculovirus UMIs.



	Number of cells	
hpi	Before filtering	After filtering
0	4,496	3,999
8	5,408	4,755
22	5,162	5,152

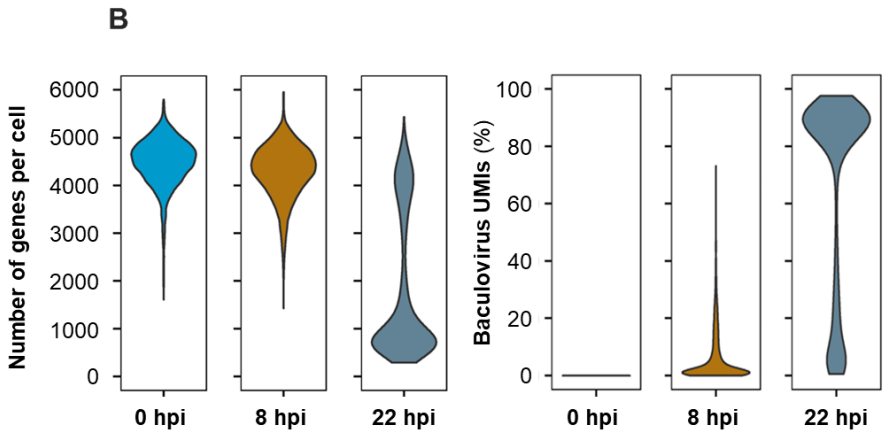


Figure 2. Single-cell RNA-seq quality control. A) Mitochondrial UMIs per cell (on the left) and number of cells before and after filtering per UMIs originating from mitochondrial genes (> 5%) (on the right). B) Number of genes identified per cell and percentage (%) of cells containing UMIs originating from baculovirus genes at each time point after cell filtering.

3.3. High Five insect cells heterogeneity

To identify cell populations existing across the three single-cell datasets (0, 8 and 22 hpi), merged analysis was performed using Seurat v4. A total of 10 clusters were drawn from this analysis (**Figure 3A**), with a noticeable cluster re-

organization being observed throughout infection progression. Importantly, 5 of those clusters were already present before infection (0 hpi) suggesting that High Five cells population was heterogeneous. Cell cycle is known to contribute to heterogeneity in scRNA-seq datasets (McDavid et al., 2016); to ascertain this in our study, cell cycle covariate was estimated using the Cell Cycle Scoring method in the Seurat package (Tirosh et al., 2016). While heterogeneity is clear at 0 hpi, with cells being associated to different cell cycle stages in a proportion of » 1:7:5 (S:G2M:G1), at 22 hpi most cells have been identified as being in G1 phase (74 %) (**Figure S2A-B**). However, at this later timepoint, cell cycle association seems to be misassigned as exemplified by the overall low expression of the G1 cell cycle-related genes *ccnd3*, *ccne1* and *cdk6* in **Figure S2C**. This could be a consequence of overexpression of baculovirus UMIs (hence lower expression of host cell genes), impairing correct cell cycle identification. Thus, the cell cycle regression was not further applied for the merged dataset.

Cluster re-organization throughout infection seems to be correlated with the expression of baculovirus genes as infection progresses (**Figure 3B**); for example, expression of an early baculovirus gene (*ACNVgp135*) was higher in clusters denoting a transitory stage (i.e. cluster #2) whereas a late baculovirus gene (*ACNVgp138*) was more expressed in clusters (e.g. clusters #3, #4, #6) furthest from those identified at 0 hpi. As seen for *ACNVgp138*, expression of transgenes M1 and HA, both under the regulation of the late expression promoter polyhedrin, was mainly identified in clusters #3, #4 and #6 (**Figure 3C**), with similar expression levels of HA and M1 genes being observed regardless of the cell cluster ($r = 0.98$, $q = 0.76$, **Figure 3D**).

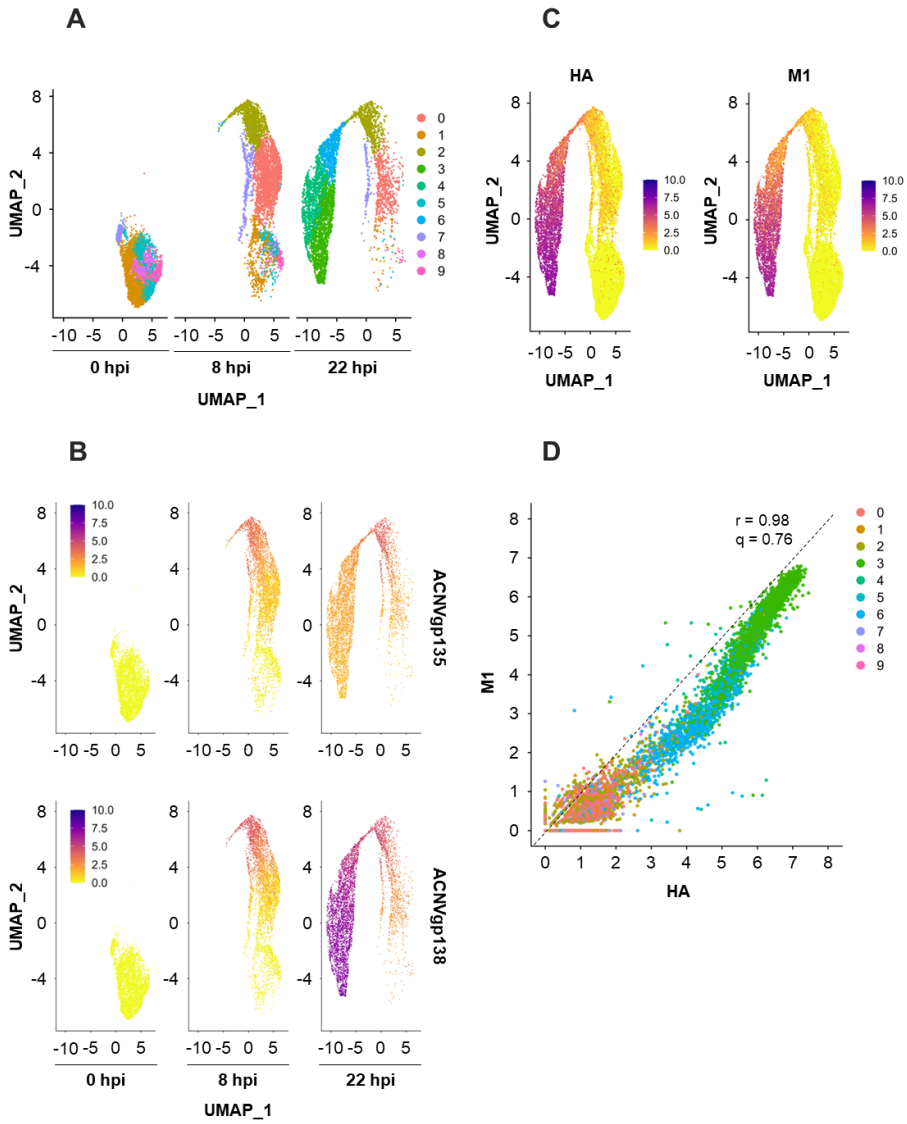


Figure 3. Insect cells clustering and transgenes expression. A) Merged scRNA-seq dataset obtained with UMAP. **B)** Relative expression of early (*ACNVgp135*) and late (*ACNVgp138*) baculovirus genes. **C)** Relative expression of transgenes HA and M1. A color gradient scale was used to show the relative gene expression per cell **D)** Relative HA and M expression per cell, with dashed line representing the best linear fit to the data (r - Pearson's correlation coefficient, q – angular coefficient). A color code was used to identify each cell cluster

3.4. Pseudo-temporal ordering of cells after infection

To assess cell population evolution during infection, pseudo-temporal ordering (i.e. trajectory analysis) was applied. To perform this analysis, 13,906 cells from the merged Seurat analysis were used as input of Monocle 3. The cell cluster with the lowest percentage of HA (cluster #5, see **Figure 3A**) was selected as root state of the trajectory since it was the one most closely resembling the non-infected cell population; the pseudotime variable was then ordered accordingly (**Figure 4A**).

Transcriptomic changes characterizing the progression of cells along HA production were assessed through identification of genes correlated with pseudotime. Overall, 921 host cell genes were found to be significantly (q value < 0.05) associated with pseudotime. Pathway enrichment analysis allowed to identify biological processes varying most along infection, of which those associated to translation machinery, energy metabolism, protein folding and endocytosis (**Figure 4B**) are some examples, in good agreement to what we have previously found in bulk RNA sequencing analysis (Silvano et al., 2022; Virgolini et al., 2022). The relative expression of a selected number of genes involved in these pathways is presented in **Figure S3**, illustrating the significant transcriptional changes in cells upon infection.

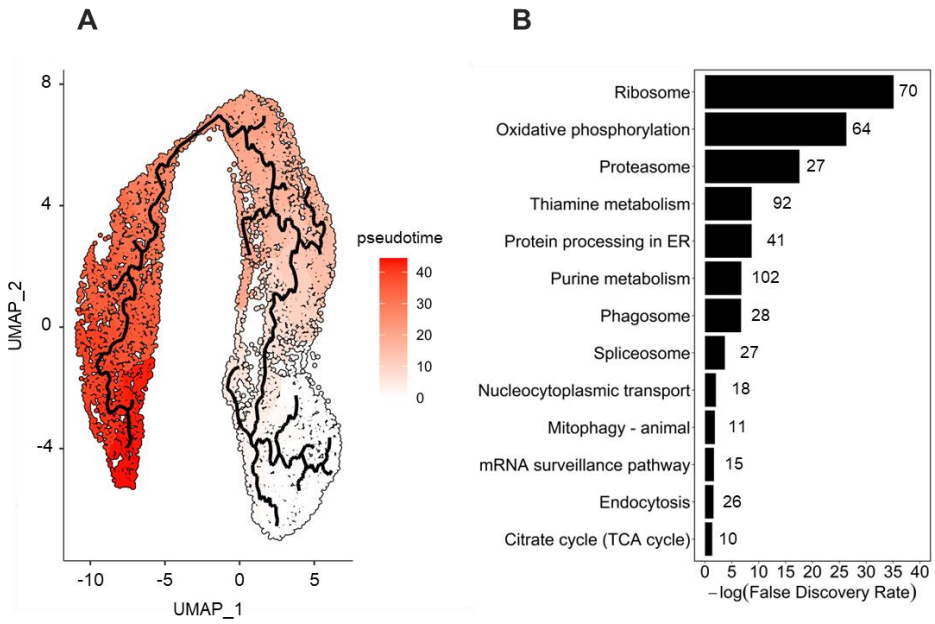


Figure 4. Pseudotime analysis of insect cells along infection. A) Trajectory analysis on merged cell dataset (0, 8 and 22 hpi); a color gradient scale was used to visualize the pseudotime. B) Pathway enrichment analysis performed with genes identified as the most changing in function of pseudotime using the Fisher's exact test; the number of genes involved and found in the enrichment analysis are shown next to the bars.

4. Discussion

In this work, we used scRNA-seq sequencing to analyze, at the single-cell level, alterations in the transcriptome of a High Five insect cell line infected with baculovirus during the production of influenza HA-VLPs.

Stable and high foreign gene expression levels are important criteria during the development of producer cell lines for pharmaceutical applications (Tripathi & Shrivastava, 2019), and understanding the mechanisms behind gene expression variation of genetically identical cells is the first step of this process. Cell population heterogeneity can influence parameters such as cell growth rate, genetic stability and productivity (Schmitz et al., 2019), and thus tracking it is key to avoid process failure and guarantee reproducibility (Olsson et al., 2022).

Most studies on this topic focus on mammalian cells (Lewis et al., 2016; Samoudi et al., 2021), with none to date exploring insect cells. In our study, we could observe heterogeneity in High Five cell population before infection (as demonstrated by the identification of 5 cell clusters), largely resulting from the phase of the cell cycle that the cells are in, which was further amplified upon infection (6 more cell clusters were identified) as consequence of viral DNA replication and gene expression (X. Du & Thiem, 1997) and cell cycle arrest (Braunagel et al., 1998).

Baculovirus genes are known to be transcribed temporally, a process highly regulated by infection-derived mechanisms and mediated by host and viral protein expression (Nguyen et al., 2013). The timing and level of baculovirus gene expression were herein identified as the main factors driving clustering of infected insect cells. Trajectory analysis allowed us to confirm this, in which a clear path along pseudotime is observed although cells separate and order across multiple branches spanning the transcriptomic space.

Biological mechanisms associated with baculovirus infection and transgenes (in our case those coding for influenza HA-VLPs) expression can be identified by correlating changes in gene(s) expression to progression of cells along the infection trajectory. Through pathway enrichment analysis, we found the endocytosis pathway as being one of the most significantly enriched biological processes during infection, which derives from viruses exploiting cellular structures towards endocytosis-mediated viral nucleocapsid transport to the nucleus (Monteiro et al., 2012). In addition, entry of baculovirus is found dependent on clathrin-mediated endocytosis (Long et al., 2006), which was herein corroborated by the up-regulation of clathrin *cltc* and actin-related *arpc5* and *capza1* proteins at early infection stages.

Among the cellular defense responses to environmental and pharmacological stresses, the activation of heat shock response (HSR) is one of the most

important. It leads to rapid and robust expression of members of the chaperone family of heat shock proteins (HSPs) in order to protect the cell from proteotoxic stresses and to maintain protein homeostasis (Fujimoto & Nakai, 2010). Interestingly, viruses can exploit HSR as an infection strategy, making use of HSPs such as HSP70 and HSP90 for regulation of viral gene expression and capsid assembly/disassembly (Mayer & Bukau, 2005; Nagy et al., 2011; Xiao et al., 2010). Our data corroborates this, with the expression of *hsp70* found significantly up-regulated early in infection. In addition, the proteasome pathway was found enriched, in agreement with the reported evidence of close collaboration between HSPs and ubiquitin-proteasome system during the baculovirus replicative cycle (Katsuma et al., 2011b).

Baculovirus infection induces an important metabolic burden on insect cells, enhancing the fluxes through the major catabolic pathways including the tricarboxylic acid cycle (TCA) (Bernal et al., 2009). Within the TCA cycle, the citrate synthase *cs* gene is involved in aerobic energy production and metabolic interconversions in mitochondria (Holloszy et al., 1970b); in our analysis, the expression of *cs* was found significantly increased at the onset of infection, suggesting that this gene plays a key role as a first-line response to infection.

Overall, the enrichment analysis allowed to identify several pathways (e.g., ribosome, spliceosome, oxidative phosphorylation) that were common to those previously identified in our bulk RNA sequencing study (Silvano et al., 2022), demonstrating the robustness and replicability of the data. Importantly, single-cell RNA sequencing allowed to evaluate single cells at different states of infection within the same sample and capture the transcriptional changes associated with the infection process (not possible with the bulk RNA sequencing approach), thus elevating the importance of single-cell omics analysis in the IC-BEVS system.

5. Conclusions

Single-cell transcriptomics enabled us to study host cell and baculovirus gene expression patterns at a resolution previously unobtainable in a bulk approach, allowing to isolate traces of different stages of infection progression. Such understanding can be further applied through genetic engineering approaches for overexpression/knock-out of specific genes, an approach that opens possibilities such as developing cell lines specialized for either virus replication or foreign protein expression, establish inducible systems, and even stimulate infection synchronization across all cells in culture towards a more controlled, homogeneous production process. Notwithstanding, using scRNA-seq to study additional IC-BEVS processes (i.e. comprising different products of interest, infection conditions, among others) is crucial for a more broader understanding of the transcriptome footprints of this expression system.

6. Data availability

The scRNA-seq datasets have been deposited to the Sequence Read Archive (SRA) with accession code PRJNA911494.

7. Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 813453 and by Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Ensino Superior (FCT/MCTES, Portugal) through the following initiatives: "Investigador FCT" Program (IF/01704/2014), Exploratory Research and Development Project (IF/01704/2014/CP1229/CT0001), and PhD fellowships

(SFRH/BD/134107/2017), iNOVA4Health (UIDB/04462/2020 and UIDP/04462/2020) and the Associate Laboratory LS4FUTURE (LA/P/0087/2020). Funding from INTERFACE Programme, through the Innovation, Technology, and Circular Economy Fund (FITEC), is gratefully acknowledged.

8. Supplementary material

Metric	Category	0 hp	8 hpi	22 hpi
Total Reads in FASTQ	Sequencing Quality	367962931	365633934	438009614
Pct_Reads_Too_Short		0	0	0
Pct_Reads_Low_Base_Quality		0.19	0.15	0.22
Pct_Reads_High_SNF		3.27	3.88	2.89
Pct_Reads_Filtered_Out		3.45	4.02	3.1
Total_Reads_After_Quality_Filtering		355254767	350931123	424421367
Total_Filtered_Reads	Library Quality	355254767	350931123	424421367
Pct_Q30_Bases_in_Filtered_R2		87.7	88.24	87.75
Pct_Assigned_to_Cell_Labels		85.18	86.16	86.93
Pct_Cellular_Reads_Aligned_Uniquely_to_Annotated_Transcriptome		59.4	62.0	69.28
Pct_Cellular_Reads_Aligned_Uniquely_to_Other_Genomic_Regions		9.01	8.44	5.96
Pct_Cellular_Reads_Aligned_Not_Unique		9.88	9.66	4.05
Pct_Cellular_Reads_Unaligned		1.22	1.19	3.57
Aligned_Reads_By_Type	Reads and Molecules	211015826	217590582	289780052
Total_Raw_Molecules		152663919	167853173	215652221
Total_RSEC_Molecules		149874976	165304754	192699547
Mean_Raw_Sequencing_Depth		1.38	1.3	1.34
Mean_RSEC_Sequencing_Depth		1.41	1.32	1.5
Sequencing_Saturation		49.22	41.8	51.38
Putative_Cell_Count	Cells RSEC	4496	5408	5162
Pct_Reads_from_Putative_Cells		91.46	89.64	79.28
Mean_Reads_per_Cell		42925.1	36067.14	44505.76
Mean_Molecules_per_Cell		30425.64	27346.1	28591.16
Median_Molecules_per_Cell		31069.5	28145.5	28130
Mean_Targets_per_Cell		4251.6	4053.64	1812.21
Median_Targets_per_Cell		4472	4266	966.5
Total_Targets_Detected		15095	15078	14722

Table S1. Single-cell RNA-seq metrics. The outputs from the Seven Bridges Genomics bioinformatics pipeline are shown including the read-processing steps and deconvolution of cell barcodes for each sample.

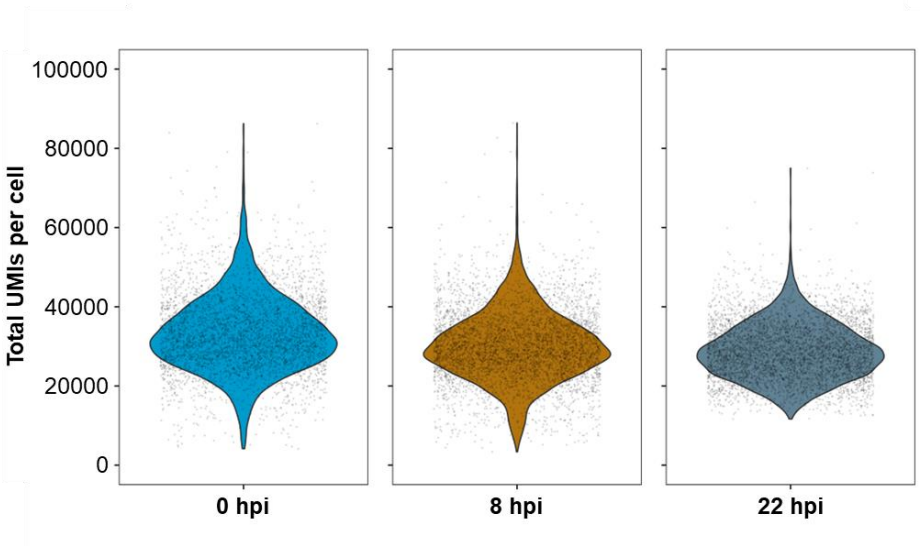


Figure S1. Number of total UMIs per cell at each time point after cell filtering.

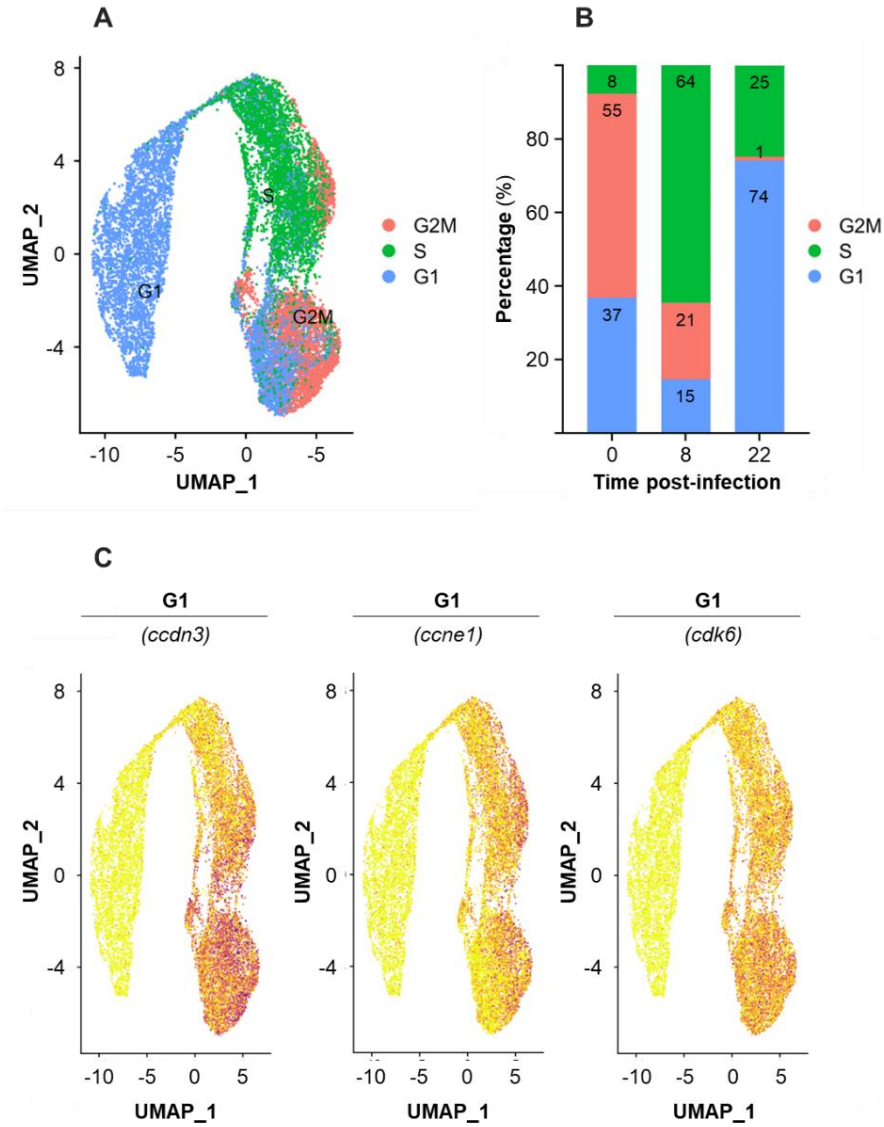


Figure S1. Cell cycle association to non-infected and infected insect cells. A) Cell cycle score assignment for merged scRNA-seq dataset. **B)** Proportion of cells assigned to each cell cycle phase. Pink, green and blue are the colors assigned to cells in G2M, S, G1, respectively. **C)** Relative expression of genes involved in G1 phase: Cyclin D3 (*ccdn3*), Cyclin D1 (*ccne1*) and Cyclin-dependent kinase 6 (*cdk6*).

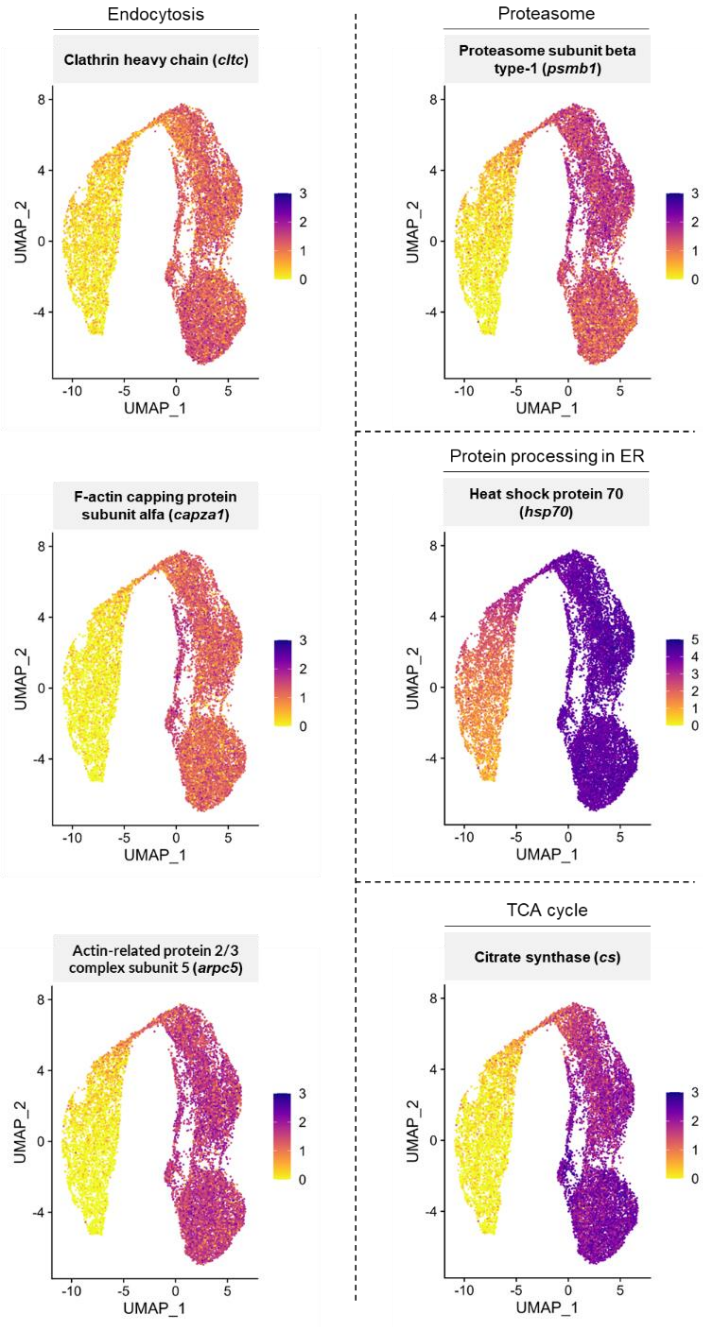


Figure S2. Relative expression of genes associated to pathways found enriched in pseudotime analysis. Cltc, arpc5 and capza1 are genes involved in the endocytosis pathway; hsp70, psmb1 and cs are genes involved in the protein processing in ER, proteasome and citrate cycle (TCA cycle), respectively. A color gradient scale was used to show the relative gene expression per cell.

9. References

Benjaminit, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol, 57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernal, V., Carinhas, N., Yokomizo, A. Y., Carrondo, M. J. T., & Alves, P. M. (2009). Cell density effect in the baculovirus-insect cells system: A quantitative analysis of energetic metabolism. *Biotechnology and Bioengineering*, *104*(1), 162–180. https://doi.org/10.1002/bit.22364

Braunagel, S. C., Parr, R., Belyavskyi, M., & Summers, M. D. (1998). Autographa californica nucleopolyhedrovirus infection results in Sf9 cell cycle arrest at G2/M phase. *Virology*, *244*(1), 195–211. https://doi.org/10.1006/viro.1998.9097

Cao, Z. J., & Gao, G. (2022). Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. *Nature biotechnology*, *40*(10), 1458–1466. https://doi.org/10.1038/s41587-022-01284-4

Chen, Y.-R., Zhong, S., Fei, Z., Gao, S., Zhang, S., Li, Z., Wang, P., & Blissard, G. W. (2014). Transcriptome Responses of the Host Trichoplusia ni to Infection by the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus. *Journal of Virology*, *88*(23), 13781–13797. https://doi.org/10.1128/jvi.02243-14

Chen, Y.-R., Zhong, S., Fei, Z., Hashimoto, Y., Xiang, J. Z., Zhang, S., & Blissard, G. W. (2013). The Transcriptome of the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus in Trichoplusia ni Cells. *Journal of Virology*, *87*(11), 6391–6405. https://doi.org/10.1128/jvi.00194-13

Correia, R., Fernandes, B., Alves, P. M., Carrondo, M. J. T., & Roldão, A. (2020). Improving influenza HA-VLPs production in insect high five cells via adaptive laboratory evolution. *Vaccines*, *8*(4), 1–15. https://doi.org/10.3390/vaccines8040589

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Drugmand, J. C., Schneider, Y. J., & Agathos, S. N. (2012). Insect cells as factories for biomanufacturing. *Biotechnology advances*, *30*(5), 1140–1157. https://doi.org/10.1016/j.biotechadv.2011.09.014

Du, X., & Thiem, S. M. (1997). Responses of insect cells to baculovirus infection: protein synthesis shutdown and apoptosis. *Journal of virology*, *71*(10), 7866–7872. https://doi.org/10.1128/JVI.71.10.7866-7872.1997

Fujimoto, M., & Nakai, A. (2010). The heat shock factor family and adaptation to proteotoxic stress. *The FEBS journal*, *277*(20), 4112–4125. https://doi.org/10.1111/j.1742-4658.2010.07827.x

Goldman, S. L., MacKay, M., Afshinnekoo, E., Melnick, A. M., Wu, S., & Mason, C. E. (2019). The Impact of Heterogeneity on Single-Cell Sequencing. *Frontiers in genetics*, *10*, 8. https://doi.org/10.3389/fgene.2019.00008

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. https://doi.org/10.1093/nar/gkn176

Gulati, G. S., Sikandar, S. S., Wesche, D. J., Manjunath, A., Bharadwaj, A., Berger, M. J., Ilagan, F., Kuo, A. H., Hsieh, R. W., Cai, S., Zabala, M., Scheeren, F. A., Lobo, N. A., Qian, D., Yu, F. B., Dirbas, F. M., Clarke, M. F., & Newman, A. M. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. *Science (New York, N.Y.)*, *367*(6476), 405–411. https://doi.org/10.1126/science.aax0249

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Holloszy, J. O., Oscai, L. B., Don, I. J., & Molé, P. A. (1970). Mitochondrial citric acid cycle and related enzymes: adaptive response to exercise. *Biochemical and biophysical research communications*, *40*(6), 1368–1373. https://doi.org/10.1016/0006-291x(70)90017-3

Ianevski, A., Giri, A. K., & Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-28803-w

J. R. Tennant. (1964). Evaluation of the trypan blue technique for determination of cell viability. *Transplantation*, *2*, 685–694. https://doi.org/10.1097/00007890-196411000-00001

Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., von Feilitzen, K., Zwahlen, M., Arif, M., Altay, O., Li, X., Ozcan, M., Mardinoglu, A., Fagerberg, L., … Lindskog, C. (2021). A single-cell type transcriptomics map of human tissues. *Science advances*, *7*(31), eabh2169. https://doi.org/10.1126/sciadv.abh2169

Katsuma, S., Tsuchida, A., Matsuda-Imai, N., Kang, W. K., & Shimada, T. (2011). Role of the ubiquitin-proteasome system in Bombyx mori nucleopolyhedrovirus infection. *Journal of General Virology*, *92*(3), 699–705. https://doi.org/10.1099/vir.0.027573-0

Lee, J., Hyeon, D. Y., & Hwang, D. (2020). Single-cell multiomics: technologies and data analysis methods. *Experimental & molecular medicine*, *52*(9), 1428–1442. https://doi.org/10.1038/s12276-020-0420-2

Long, G., Pan, X., Kormelink, R., & Vlak, J. M. (2006). Functional Entry of Baculovirus into Insect and Mammalian Cells Is Dependent on Clathrin-Mediated Endocytosis. *Journal of Virology*, *80*(17), 8830–8833. https://doi.org/10.1128/jvi.00880-06

Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and molecular life sciences : CMLS*, *62*(6), 670–684. https://doi.org/10.1007/s00018-004-4464-6

McDavid, A., Finak, G., & Gottardo, R. (2016). The contribution of cell cycle to heterogeneity in single-cell RNA-seq data. *Nature biotechnology*, *34*(6), 591–593. https://doi.org/10.1038/nbt.3498

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. http://arxiv.org/abs/1802.03426

Mena, J. A., Ramírez, O. T., & Palomares, L. A. (2003). Titration of Non-Occluded Baculovirus Using a Cell Viability Assay. *BioTechniques*, *34*(2), 260–264. https://doi.org/10.2144/03342bm05

Monteiro, F., Carinhas, N., Carrondo, M. J., Bernal, V., & Alves, P. M. (2012). Toward system-level understanding of baculovirus-host cell interactions: from molecular fundamental studies to large-scale proteomics approaches. *Frontiers in microbiology*, *3*, 391. https://doi.org/10.3389/fmicb.2012.00391

Nagy, P. D., Wang, R. Y., Pogany, J., Hafren, A., & Makinen, K. (2011). Emerging picture of host chaperone and cyclophilin roles in RNA virus replication. *Virology*, *411*(2), 374–382. https://doi.org/10.1016/j.virol.2010.12.061

Nguyen, Q., Nielsen, L. K., & Reid, S. (2013). Genome scale transcriptomics of baculovirus-insect interactions. *Viruses*, *5*(11), 2721–2747. https://doi.org/10.3390/v5112721

Roldão, A., Oliveira, R., Carrondo, M. J. T., & Alves, P. M. (2009). Error assessment in recombinant baculovirus titration: Evaluation of different methods. *Journal of Virological Methods*, *159*(1), 69–80. https://doi.org/10.1016/j.jviromet.2009.03.007

Russell, A. B., Trapnell, C., & Bloom, J. D. (2018). Extreme heterogeneity of influenza virus infection in single cells. *eLife*, *7*, e32303. https://doi.org/10.7554/eLife.32303

Silvano, M., Correia, R., Virgolini, N., Clarke, C., Alves, P. M., Isidro, I. A., & Roldão, A. (2022). Gene Expression Analysis of Adapted Insect Cells during Influenza VLP

Production Using RNA-Sequencing. *Viruses*, *14*(10), 2238. https://doi.org/10.3390/v14102238

Stockholm, D., Benchaouir, R., Picot, J., Rameau, P., Neildez, T. M. A., Landini, G., Laplace-Builhé, C., & Paldi, A. (2007). The origin of phenotypic heterogeneity in a clonal cell population in vitro. *PLoS ONE*, *2*(4). https://doi.org/10.1371/journal.pone.0000394

Sun, J., Vera, J. C., Drnevich, J., Lin, Y. T., Ke, R., & Brooke, C. B. (2020). Single cell heterogeneity in influenza A virus gene expression shapes the innate antiviral response to infection. *PLoS Pathogens*, *16*(7). https://doi.org/10.1371/journal.ppat.1008671

Suomalainen, M., & Greber, U. F. (2021). Virus Infection Variability by Single-Cell Profiling. *Viruses*, *13*(8), 1568. https://doi.org/10.3390/v13081568

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., 2nd, Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J. R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G., Kazer, S. W., … Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (New York, N.Y.)*, *352*(6282), 189–196. https://doi.org/10.1126/science.aad0501

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, *32*(4), 381–386. https://doi.org/10.1038/nbt.2859

Tzani, I., Herrmann, N., Carillo, S., Spargo, C. A., Hagan, R., Barron, N., Bones, J., Shannon Dillmore, W., & Clarke, C. (2021). Tracing production instability in a clonally derived CHO cell line using single-cell transcriptomics. *Biotechnology and Bioengineering*, *118*(5), 2016–2030. https://doi.org/10.1002/bit.27715

Vieira, H. L. A., Estêvão, C., Roldão, A., Peixoto, C. C., Sousa, M. F. Q., Cruz, P. E., Carrondo, M. J. T., & Alves, P. M. (2005). Triple layered rotavirus VLP production: Kinetics of vector replication, mRNA stability and recombinant protein production. *Journal of Biotechnology*, *120*(1), 72–82. https://doi.org/10.1016/j.jbiotec.2005.03.026

Virgolini, N., Hagan, R., Correia, R., Silvano, M., Fernandes, S., Alves, P. M., Clarke, C., Roldão, A., & Isidro, I. A. (2023). Transcriptome analysis of Sf9 insect cells during production of recombinant Adeno-associated virus. *Biotechnology journal*, *18*(2), e2200466. https://doi.org/10.1002/biot.202200466

Wagner, D. E., & Klein, A. M. (2020). Lineage tracing meets single-cell omics: opportunities and challenges. *Nature reviews. Genetics*, *21*(7), 410–427. https://doi.org/10.1038/s41576-020-0223-2

Xiao, A., Wong, J., & Luo, H. (2010). Viral interaction with molecular chaperones: role in regulating viral infection. *Archives of virology*, *155*(7), 1021–1031. https://doi.org/10.1007/s00705-010-0691-3

Xu, C., & Su, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics*, *31*(12), 1974–1980. https://doi.org/10.1093/bioinformatics/btv088

Chapter 4

Discussion

Author contribution

Marco Silvano wrote the chapter.

TABLE OF CONTENTS

1. Discussion

The IC-BEVS is a robust platform for production of complex biopharmaceuticals. However, little is known about the interactions established between the insect host cell and the baculovirus vector and how they shape the cellular and metabolic rearrangements that occur following infection. Deciphering the baculovirus-host interplay will provide clues on the viral mechanisms subverting host cell metabolism to support biosynthesis, thus aiding in the identification of hidden constraints that govern productivity.

In this PhD thesis, key biological processes of insect cells impacted during infection by baculovirus were revealed using bulk and scRNA-seq, laying the ground for future rational host and/or virus engineering towards improved production of influenza HA-VLPs. Bulk RNA-seq was used to investigate the gene expression patterns associated with high-producer insect cells adapted to neutral pH (in comparison to non-adapted cells) while scRNA-seq was focused on assessing cell population heterogeneity prior and upon infection. The challenges encountered during the implementation of both technologies for insect High Five cells and those laying ahead are discussed in this section.

1.1. The importance of high-quality genome annotations

RNA-seq is nowadays an indispensable approach for comparative transcriptome profiling in model and non-model organisms. Analysing RNA-seq data from non-model organisms poses unique challenges due to scarcity of high-quality genome annotation and/or tools for downstream functional analyses.

Systematic sequence similarity searches against non-redundant sequence sets and BLAST2GO searches (Conesa et al., 2005) are often employed for the functional inference of transcriptomes to be utilized in gene set enrichment analyses. The output of such analyses relies upon the comprehensiveness of the

available datasets and their functional annotation. Up to today, the genomes of around 3000 insect species have been decoded and registered in the genome section of the NCBI Datasets (Bono et al., 2022), but few have the high-quality, accuracy needed for reliable association of genes or gene products to functions. With recent advances in next-generation molecular sequencing technologies, progresses were made on insects belonging to the Lepidoptera order (e.g. *Spodoptera frugiperda* and *Trichoplusia ni*). However, the sparsity of reliable gene, transcript, and protein annotations available to researchers pose a major problem to researchers: only 49% of unique genes represented in a recently published Lepidoptera transcriptome could be successfully annotated (P. A. Tang et al., 2017), belying the great need for further gene function studies on Lepidoptera.

Transcriptome annotation is important to assign functional meaning to the transcripts. In **Chapter 2** and **Chapter 3**, annotation was done at the gene level; by doing this, it is recommended to first create a dataset containing a representative transcript per gene, selected based on length or coverage. A protein-based search is more sensitive. There are many options for selecting a protein database, each with its own advantages and disadvantages. A generic database, such as NCBI nr or UniProt will provide a more comprehensive annotation, whereas searching against a single closely related annotated species may yield more relevant annotations. Using a less redundant and more curated protein database, such as NCBI's RefSeq or the Swiss-Prot section of UniProt, reduces computation time and offers more reliable annotations, at the expense of comprehensiveness.

A major limitation of sequence similarity-based annotation is that the best BLAST hit achieved in a database search is sometimes described as "hypothetical protein," "predicted protein," "uncharacterized protein," or the like, whereas subsequent significant hits do have informative title. In **Chapter 2** and **Chapter**

3, Blast2GO tool was used to extract the best possible protein description from the set of n top BLAST hits passing the specified e-value cutoff (default: up to 20 proteins with e-value <0.001) (Götz et al., 2008).

Lists of differentially expressed genes from treatment comparisons (e.g. infected cells) as well as groups of co-expressed genes from clustering analysis (collectively termed “genes of interest”) may be further analyzed to find enriched biological functions and pathways within them. These analyses require databases which associate the organism genes with functional categories such as ontology terms, conserved domains, metabolic and regulatory pathways, shared regulators or targets etc. For model organisms, there is a vast assortment of analysis tools that include the corresponding databases. With non-model organisms, however, the situation is challenging. Some servers do harbour a large number of organisms (e.g. KEGG, Kanehisa et al., 2017) where it is possible to search for resources that are specific to certain species. One such example is FlyBase, which offers a GO analysis toolkit and database for the insect cell community, currently including around 60 species (Gramates et al., 2017).

Noteworthy, only 2% of RNA is translated to make proteins; other RNA sequences that do not code for any protein are called noncoding RNA (ncRNAs) (Djebali et al., 2012). Given that ncRNAs have potential to interact with other genetic materials like DNA, RNA or protein molecules, it is likely that gene expression studies provide insights about the function of these molecules. In fact, continued efforts in this direction have allowed the discovery of a number of long ncRNAs biological roles in insects (Choudhary et al., 2021), although the significant challenges of filtering out ncRNAs from coding RNA and the lack of proper bioinformatics tools to identify multifunctional RNA still exist.

1.2. Acknowledging cell population heterogeneity

The production of biopharmaceuticals using cell lines has increased over the years. Independently of the product to express, important aspects have to be considered and monitored in every process, including not only cell growth behavior and productivity but also the clonality and stability of the producer cell line (Geisse & Fux, 2009). Conventionally, every cell in an isogenic population is assumed to behave exactly the same, whereby the population average is believed to represent the behaviour of every single cell. Over the last few decades, this theory has changed, and the awareness of heterogeneity inside a genetically identical cell culture has raised interest in studying cell behaviour at single level. This paradigm shift has led to an increasing number of measuring techniques with single-cell resolution (e.g. scRNA-seq), which are helping in understanding the impact of cell population heterogeneity on bioprocess outcome.

Differences in the growth behaviour of each cell within an initially isogenic cell culture can lead to the formation of subpopulations with different properties. Improved culture conditions and extensive empirical screening of large numbers of candidate clones are therefore required, being resource intensive and frequently rate limiting steps in early development. Protein expression stability also tends to be problematic: a variety of mechanisms including transcriptional silencing, gene copy loss, and increased susceptibility to cellular stress have been associated with a sudden or gradual loss of monoclonal antibody production in CHO cell lines (Z. Du et al., 2013; Pilbrough et al., 2009; Tzani et al., 2021). In this context, heterogeneity studies can help to identify such trends early on and thus better understand these processes (Periyannan Rajeswari et al., 2017).

The progression of cells through the cell cycle induces significant metabolic shifts with a direct impact on product titers (Möller et al., 2019). Therefore, it is

important to understand cell cycle regulations, control cell population dynamics and avoid inadvertently induced oscillations of cell cycle distributions in order to improve process efficiency and robustness. In scRNA-seq analysis, cell cycle is one of the major sources of biological noise (Buettner et al., 2015). Cells are at different cell cycle check points and thus they may have different expression profiles even if these are cells of the same type (Singh et al., 2013). This within-type heterogeneity can seriously deteriorate the performance of clustering algorithms for cell type identification: it may blur clusters of cell types or cause cells of similar cell-cycle statuses to stand out as new clusters. In **Chapter 3**, to determine the cell cycle phase of each insect cell, a score indicating the likelihood of cells being in either S or G2/M phase was assigned, based on the supplied reference mouse genes published from Tirosh et al., 2016. The list of mouse cell cycle genes was associated to the *Trichoplusia ni* genome using a protein BLAST search. However, such approach shows limitations since several cell cycle genes were filtered out from the analysis because the e-value was out of the cut-off 0.01. The resulting scores for S and G2/M phase were used to regress out the effect of cell cycle in downstream analyses during data pre-processing. The cell cycle, as the most common covariate, needs to be eliminated in order to get rid of any irrelevant variation in scRNA-seq data. However, it was ultimately concluded that in virus-based processes such as the one herein used (IC-BEVS), cell cycle regression cannot be applied at later stages of cell infection because host cell transcription machinery is taken over by the baculovirus thus leading to global shut-down of host protein synthesis (Monteiro et al., 2012). It is therefore evident the need for an approach/method to successfully remove the effects of the cell cycle and improve the clustering of sub-populations of infected (and non-infected) insect cells.

1.3. The meta-interaction between pathways

To date, most improvements in protein production have been achieved by media and bioprocess optimization (e.g., feeding strategies and process parameter control). However, the availability of high-throughput 'omics data and the emergence of genome editing tools provide novel opportunities for targeted host cell genome engineering. Indeed, they have enabled the overexpression or down-regulation of specific gene candidates to increase yield during culture and control product quality (Dietmair et al., 2011; Fischer et al., 2015; Kim et al., 2012; Lim et al., 2010; Zhu, 2012). Several prominent strategies have targeted cell metabolism, cell cycle regulatory machinery, protein secretion pathway, apoptosis, and protein glycosylation (Klein et al., 2015; Lim et al., 2010).

In **Chapter 2** and **Chapter 3**, several pathways were identified, some already known to increase production yields and product quality (e.g. intracellular trafficking, endocytosis, lipid metabolism, unfolded protein response, cell cycle related functions) (Charaniya et al., 2009; Fischer et al., 2015) and others of unknown impact (e.g. metabolism of xenobiotic and endogenous compounds, metabolism of cofactors and vitamins).

Most successful studies targeted energy metabolism to reduce the accumulation of toxic by-products (i.e., lactate and ammonia) and/or increase metabolic efficiency. Several studies have successfully decreased glucose uptake by up to 50%, leading to reduced lactate production in human and murine derived cells (Inoue et al., 2010; Paredes et al., 1999; Wlaschin & Hu, 2007). An 80% reduction in lactate production was also achieved by knocking down lactate dehydrogenase leading to an improved product titer from 2 to 3-fold (K. Chen et al., 2001; S. H. Kim & Lee, 2007). However, this is a frail strategy to be carried in insect cells, since they were shown to already possess a fully functional TCA cycle and accumulate low levels of lactate (J. Neermann & R.

Wagner, 1996). Other strategies have overexpressed pyruvate carboxylase to improve the connection between glycolysis and the TCA cycle in mammalian and insect cells (Elias et al., 2003; Irani et al., 2002; S. H. Kim & Lee, 2007). In addition to lactate, efforts have been made to control other toxic by-products that impact cell viability and product quality. For example, ammonia production has been reduced by overexpressing the first two steps of the urea cycle by up to 35% and by overexpressing glutamine synthetase (Park et al., 2000). Such efforts have improved the synthesis of glutamine from glutamate (Bell et al., 1995.; Fan et al., 2012; Cockett et al., 1990) and thereby improved cell phenotypes.

Most of the metabolic engineering efforts in the aforementioned studies aimed to reduce toxic by-product accumulation by targeting single genes related to carbohydrate metabolism. However, studies are now tracking the influence of central carbon metabolism on other pathways, including the biosynthesis and metabolism of amino acids, nucleic acids, lipids, and ultimately protein synthesis and protein quality. The knowledge of these pathway connections can be used, for example, to track the cellular switch occurring between exponential growth and protein production, since this event changes the balance of different central metabolic pathways (e.g., pentose phosphate pathway flux and oxidative TCA cycle) (Templeton et al., 2013). For example, the ratio of flux between the pentose phosphate pathway (PPP) and glycolysis flux has been linked to protein production yields (Nicolae et al., 2014). Furthermore, Mulukutla et al., 2016 investigated the regulation of glucose metabolism, including the connections between glycolysis, the pentose phosphate pathway, nucleotide synthesis, glycerol-3 phosphate metabolism and serine/glycine/ threonine biosynthesis. The interconnection between these pathways reiterates the central role of carbon metabolism in regulating the global metabolic state of a cell and other cellular processes (e.g., apoptosis, glycosylation) influencing protein quality attributes. With carbon metabolism connected to other cell processes, it is

possible to control desired cellular attributes by only manipulating genes of central metabolism (Lewis, Croughan, et al., 2016; McAtee et al., 2014).

1.4. Rational bioprocess improvement using omics data – how far along are we?

The ideal insect cell line for protein production would typically include attributes such as high cell viability, cell density, and titer. It also would exhibit robust growth, high stability of expression and control over desired post-translational modifications. To achieve these attributes, multiple modifications are needed in insect cells, and such changes would target diverse cell pathways and physiological functions.

To begin identifying genes whose expression correlate with desirable attributes of protein synthesis and secretion, several studies have compared high vs low producer mammalian cells at multi-'omics level (Samoudi et al., 2021). A similar understanding of the IC-BEVS is critical to devise strategies capable of circumventing production bottlenecks (e.g., lytic nature of baculovirus infection, complete take-over of cell machinery by baculovirus, the cell density effect and the event of asynchronous infection). Limited advancement has been made in understanding the underlying biological mechanisms to enhance recombinant protein expression in IC-BEVS. In **Chapter 2** and **Chapter 3**, progresses were made by profiling the insect High Five cell line during production of Influenza HA-VLPs. Major bottlenecks of such efforts have been the lack of high-quality reference data for gene annotation and reliable omics analysis pipelines. Moreover, we are still missing fundamental knowledge about how these cellular mechanisms are organized and link together; it is often not clear how they are connected to process conditions, and how these factors all impact protein production. Thus, efforts to further enhance production will be facilitated as the molecular basis of these processes are studied and linked to protein production.

In this context, a multi-omics approach could shed further light to improve host cells by enhancing culture longevity and cell viability and/or by boosting specific productivity. Pathway maps and interaction networks are starting points that link the processes, and can help identify new process conditions and cell engineering strategies that control product quantity and quality. Notably, advances in omics data generation and sharing will help augment current modelling techniques, allowing for more context-specific predictions and engineering target discovery.

With the genetic basis established, a holistic understanding of the cellular basis for high productivity could be achieved within the systems biology context. This will be accomplished by the development of detailed metabolic pathway and insect cell-baculovirus interaction maps of the major cell processes, and identifying the genes associated with the pathways. These metabolic networks can be converted into mathematical models that can guide engineering efforts by quantifying the connection of cellular processes to desired phenotypes and protein production using metabolic flux analysis. The analyses of these data in the context of cellular pathways will be particularly informative when investigated along with phenotypic differences of different cell lines, such as variations in growth media, feeding strategies, process conditions, and the type and amount of produced protein.

Finally, with the 'omics basis and pathways mapped out, we can move towards the design and implementation of more complex genetic changes, using genome edits in IC-BEVS (Hong et al., 2022). Similarly, we can harness the multiplex targeting of miRNA (Chavez-Pena & Kamen, 2018). These tools will be invaluable in the future of engineering, wherein multiplex metabolic engineering strategies account for details of cell line, culture environment and product, in the pursuit of a more productive therapeutic production factory.

2. Conclusions and future perspectives

The following major conclusions can be withdrawn from this work:

- Adaptation of insect High Five cells to grow at neutral pH leads to significant gene expression changes upon infection;
- Down-regulation of baculovirus genes together with delayed onset of cell viability drop suggests that infected pH-adapted and non-adapted cells have different susceptibility to baculovirus infection;
- Most enriched pathways in infected pH-adapted cells are down-regulated and associated to baculovirus infection;
- Insect cell population heterogeneity before infection is mostly driven by cell cycle;
- Insect cell population heterogeneity increases along infection;
- Timing and level of baculovirus gene expression drives the clustering of infected insect cells;
- Comparable expression of M1 and HA transgenes levels regardless of the cell cluster.

Overall, this PhD thesis substantially advances our understanding of the impact of baculovirus infection on insect High Five cells during production of influenza HA-VLPs, paving the way for the design of rational approaches towards process optimization. Nonetheless, further studies are still required in order to fully extract the potential of insect cells for the production of this complex biologic, namely (i) analysis of alternative RNA splicing, (ii) lipidomic analysis, (iii) single-cell RNA sequencing of neutral pH adapted cells, (iv) cell cycle synchronization, and (v) cell/virus genetic engineering.

Alternative splicing is a process occurring during gene expression that allows a single gene to code for multiple proteins. In this process, particular exons of a gene may be included within or excluded from the final, processed mRNA

produced from that gene. Consequently, the proteins translated from alternatively spliced mRNAs usually contain differences in their amino acid sequence and, often, in their biological functions. The overall impact of alternative isoforms on insect cells phenotype is still unknown, and the analysis of mRNA variants would potentially be important to reveal differences between non-adapted and adapted cells that could explain the higher productivities achieved by adapted cells.

Lipids are a class of molecules involved in essentially all stages of viral life cycles, and the lipid composition of the host cell membrane can influence the efficiency of viral entry and egress. Comparison of the lipid profiles between non-adapted and adapted High Five cells may reveal important lipids and/or related enzymes in lipid biosynthesis pathway associated with *Ac*MNPV infection. The integration of lipidomic profiles with transcriptomic data could correlate lipid levels to gene expression and bridge the gap between insect cell genotype and phenotype. Currently, there are no lipidomic studies focusing on insect cells; as understanding of insect cell lipid associated signaling and synthetic pathways increases, future lipidomics studies could have significant impact on improving protein titres in IC-BEVS.

pH-adapted cells are the result of a stepwise approach in which the culture pH gradually increased from 6.2 to 7.0 to improve production of influenza HA-VLPs. Single-cell RNA sequencing analysis of neutral pH-adapted cells could help us to understand the insect cell population evolution to atypical culture conditions and therefore reveal imbalanced cell clustering when compared to non-adapted cells. More importantly, it would allow us to differentiate between high- and low-producing subpopulations and thus identify biomarkers, i.e. gene(s) and/or pathway(s), associated to higher protein expression.

Host cell heterogeneity is driven by many factors, of which cell cycle is one of the major contributor. In addition, cells arrested at G1 phase have been shown

to exhibit the highest ribosome biogenesis and protein translation activity of any cell cycle phase, resulting in higher recombinant protein yields. Therefore, a strategy to reduce cell-to-cell variation while enhancing protein expression would be the synchronization of cell cycle to G1 phase by mechanic approach (i.e. isolation of cells at specific phase of the cell cycle) or cell cycle arrest agents (e.g. thymidine).

The transcriptome analysis performed in this thesis assisted the identification of potential engineering targets to increase recombinant protein production. Specifically, the up-regulation of molecular chaperones using CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats – CRISPR-associated protein 9) would act to preserve nascent proteins in a folding-competent conformation and prevent aggregation. Another target pathway is the one associated to cell growth/death, and the inhibition of pro-apoptotic factors by RNA interference could delay the onset of apoptosis, increase the maximum viable cell density and enhance expression HA-VLPs.

3. References

Bell, S. L., Bebbington, C., Scott, M. F., Wardell, J. N., Spier, R. E., Bushell, M. E., & Sanders, P. G. (1995). Genetic engineering of hybridoma glutamine metabolism. *Enzyme and microbial technology*, *17*(2), 98–106. https://doi.org/10.1016/0141-0229(94)00056-w

Bono, H., Sakamoto, T., Kasukawa, T., & Tabunoki, H. (2022). Systematic Functional Annotation Workflow for Insects. *Insects*, *13*(7). https://doi.org/10.3390/insects13070586

Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A., Marioni, J. C., & Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnology*, *33*(2), 155–160. https://doi.org/10.1038/nbt.3102

Carinhas, N., Bernal, V., Monteiro, F., Carrondo, M. J., Oliveira, R., & Alves, P. M. (2010). Improving baculovirus production at high cell density through manipulation of energy metabolism. *Metabolic engineering, 12*(1), 39–52. https://doi.org/10.1016/j.ymben.2009.08.008

Charaniya, S., Karypis, G., & Hu, W. S. (2009). Mining transcriptome data for function-trait relationship of hyper productivity of recombinant antibody. *Biotechnology and Bioengineering*, *102*(6), 1654–1669. https://doi.org/10.1002/bit.22210

Chavez-Pena, C., & Kamen, A. A. (2018). RNA interference technology to improve the baculovirus-insect cell expression system. *Biotechnology advances*, *36*(2), 443–451. https://doi.org/10.1016/j.biotechadv.2018.01.008

Chen, K., Liu, Q., Xie, L., Sharp, P. A., & Wang, D. I. (2001). Engineering of a mammalian cell line for reduction of lactate formation and high monoclonal antibody production. *Biotechnology and bioengineering*, *72*(1), 55–61. https://doi.org/10.1002/1097-0290(20010105)72:1<55::aid-bit8>3.0.co;2-4

Choudhary, C., Sharma, S., Meghwanshi, K. K., Patel, S., Mehta, P., Shukla, N., Do, D. N., Rajpurohit, S., Suravajhala, P., & Shukla, J. N. (2021). Long Non-Coding RNAs in Insects. *Animals : an open access journal from MDPI*, *11*(4), 1118. https://doi.org/10.3390/ani11041118

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Dietmair, S., Nielsen, L.K., Timmins, N.E. (2011). Engineering a mammalian super producer. *J. Chem. Technol. Biotechnol. 86*, 905–914. https://doi.org/10.1002/jctb.2576.

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108. https://doi.org/10.1038/nature11233

Drews, M., Paalme, T., Vilu, R. (1995). The growth and nutrient utilization of the insect cell line Spodoptera frugiperda Sf9 in batch and continuous culture. *Journal of Biotechnology, 40*(3), 187-198. https://doi.org/10.1016/0168-1656(95)00045-R.

Du, Z., Mujacic, M., Le, K., Caspary, G., Nunn, H., Heath, C., & Reddy, P. (2013). Analysis of heterogeneity and instability of stable mAb-expressing CHO cells. *Biotechnology and Bioprocess Engineering*, *18*(2), 419–429. https://doi.org/10.1007/s12257-012-0577-1

Elias, C. B., Carpentier, E., Durocher, Y., Bisson, L., Wagner, R., & Kamen, A. (2003). Improving glucose and glutamine metabolism of human HEK 293 and Trichoplusia ni insect cells engineered to express a cytosolic pyruvate carboxylase enzyme. *Biotechnology Progress*, *19*(1), 90–97. https://doi.org/10.1021/bp025572x

Fan, L., Kadura, I., Krebs, L. E., Hatfield, C. C., Shaw, M. M., & Frye, C. C. (2012). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnology and Bioengineering*, *109*(4), 1007–1015. https://doi.org/10.1002/bit.24365

Fischer, S., Handrick, R., & Otte, K. (2015). The art of CHO cell engineering: A comprehensive retrospect and future perspectives. *Biotechnology advances*, *33*(8), 1878–1896. https://doi.org/10.1016/j.biotechadv.2015.10.015

Geisse, S., & Fux, C. (2009). Recombinant protein production by transient gene transfer into Mammalian cells. *Methods in enzymology*, *463*, 223–238. https://doi.org/10.1016/S0076-6879(09)63015-9

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J., & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*(10), 3420–3435. https://doi.org/10.1093/nar/gkn176

Gramates, L. S., Marygold, S. J., dos Santos, G., Urbano, J. M., Antonazzo, G., Matthews, B. B., Rey, A. J., Tabone, C. J., Crosby, M. A., Emmert, D. B., Falls, K., Goodman, J. L., Hu, Y., Ponting, L., Schroeder, A. J., Strelets, V. B., Thurmond, J., Zhou, P., Perrimon, N., … Baker, P. (2017). FlyBase at 25: Looking to the future. *Nucleic Acids Research*, *45*(D1), D663–D671. https://doi.org/10.1093/nar/gkw1016

Cockett, M. I., Bebbington, C. R., & Yarranton, G. T. (1990). High level expression of tissue inhibitor of metalloproteinases in Chinese hamster ovary cells using glutamine synthetase gene amplification. *Bio/technology (Nature Publishing Company), 8*(7), 662–667. https://doi.org/10.1038/nbt0790-662

Hong, M., Li, T., Xue, W., Zhang, S., Cui, L., Wang, H., Zhang, Y., Zhou, L., Gu, Y., Xia, N., & Li, S. (2022). Genetic engineering of baculovirus-insect cell system to improve protein production. *Frontiers in bioengineering and biotechnology*, *10*, 994743. https://doi.org/10.3389/fbioe.2022.994743

Inoue, Y., Tsukamoto, Y., Yamanaka, M., Nakamura, S., Inoue, A., Nishino, N., & Kawahara, H. (2010). Efficient production of recombinant IgG by metabolic control and co-expression with GLUT5 in a fructose-based medium. *Cytotechnology*, *62*(4), 301–306. https://doi.org/10.1007/s10616-010-9289-6

Irani, N., Beccaria, A. J., & Wagner, R. (2002). Expression of recombinant cytoplasmic yeast pyruvate carboxylase for the improvement of the production of human erythropoietin by recombinant BHK-21 cells. *Journal of biotechnology*, *93*(3), 269–282. https://doi.org/10.1016/s0168-1656(01)00409-6

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. https://doi.org/10.1093/nar/gkw1092

Kim, J. Y., Kim, Y. G., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied microbiology and biotechnology*, *93*(3), 917–930. https://doi.org/10.1007/s00253-011-3758-5

Kim, S. H., & Lee, G. M. (2007a). Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. *Applied Microbiology and Biotechnology*, *74*(1), 152–159. https://doi.org/10.1007/s00253-006-0654-5

Kim, S. H., & Lee, G. M. (2007b). Functional expression of human pyruvate carboxylase for reduced lactic acid formation of Chinese hamster ovary cells (DG44). *Applied Microbiology and Biotechnology*, *76*(3), 659–665. https://doi.org/10.1007/s00253-007-1041-6

Klein, T., Niklas, J., & Heinzle, E. (2015). Engineering the supply chain for protein production/secretion in yeasts and mammalian cells. *Journal of Industrial Microbiology and Biotechnology*, *42*(3), 453–464. https://doi.org/10.1007/s10295-014-1569-2

Lewis, A. M., Croughan, W. D., Aranibar, N., Lee, A. G., Warrack, B., Abu-Absi, N. R., Patel, R., Drew, B., Borys, M. C., Reily, M. D., & Li, Z. J. (2016). Understanding and controlling sialylation in a CHO Fc-Fusion process. *PLoS ONE*, *11*(6). https://doi.org/10.1371/journal.pone.0157111

Lim, Y., Wong, N. S. C., Lee, Y. Y., Ku, S. C. Y., Wong, D. C. F., & Yap, M. G. S. (2010). Engineering mammalian cells in bioprocessing – current achievements and future perspectives. *Biotechnology and Applied Biochemistry*, *55*(4), 175–189. https://doi.org/10.1042/ba20090363

McAtee, A. G., Templeton, N., & Young, J. D. (2014). Role of Chinese hamster ovary central carbon metabolism in controlling the quality of secreted biotherapeutic proteins. *Pharmaceutical Bioprocessing*, *2*(1), 63–74. https://doi.org/10.4155/pbp.13.65

Möller, J., Bhat, K., Riecken, K., Pörtner, R., Zeng, A. P., & Jandt, U. (2019). Process-induced cell cycle oscillations in CHO cultures: Online monitoring and model-based investigation. *Biotechnology and Bioengineering*, *116*(11), 2931–2943. https://doi.org/10.1002/bit.27124

Monteiro, F., Carinhas, N., Carrondo, M. J. T., Bernal, V., & Alves, P. M. (2012). Toward system-level understanding of baculovirus-host cell interactions: From molecular fundamental studies to large-scale proteomics approaches. In *Frontiers in Microbiology* (Vol. 3, Issue NOV). Frontiers Research Foundation. https://doi.org/10.3389/fmicb.2012.00391

Mulukutla, B. C., Yongky, A., Le, T., Mashek, D. G., & Hu, W. S. (2016). Regulation of Glucose Metabolism - A Perspective From Cell Bioprocessing. *Trends in biotechnology*, *34*(8), 638–651. https://doi.org/10.1016/j.tibtech.2016.04.012

Neermann, J., & Wagner, R. (1996). Comparative analysis of glucose and glutamine metabolism in transformed mammalian cell lines, insect and primary liver cells. *Journal of cellular physiology, 166*(1), 152–169. https://doi.org/10.1002/(SICI)1097-4652(199601)166:1<152::AID JCP18>3.0.CO;2-H

Nicolae, A., Wahrheit, J., Bahnemann, J., Zeng, A. P., & Heinzle, E. (2014). Non-stationary 13C metabolic flux analysis of Chinese hamster ovary cells in batch culture using extracellular labeling highlights metabolic reversibility and compartmentation. *BMC Systems Biology*, *8*(1). https://doi.org/10.1186/1752-0509-8-50

Paredes, C., Prats, E., Cairó, J. J., Azorín, F., Cornudella, L., & Gòdia, F. (1999). Modification of glucose and glutamine metabolism in hybridoma cells through metabolic engineering. *Cytotechnology*, *30*(1-3), 85–93. https://doi.org/10.1023/A:1008012518961

Park, H., Kim, I. H., Kim, I. Y., Kim, K. H., & Kim, H. J. (2000). Expression of carbamoyl phosphate synthetase I and ornithine transcarbamoylase genes in Chinese hamster ovary dhfr-cells decreases accumulation of ammonium ion in culture media. *Journal of biotechnology*, *81*(2-3), 129–140. https://doi.org/10.1016/s0168-1656(00)00282-0

Periyannan Rajeswari, P. K., Joensson, H. N., & Andersson-Svahn, H. (2017). Droplet size influences division of mammalian cell factories in droplet microfluidic cultivation. *Electrophoresis*, *38*(2), 305–310. https://doi.org/10.1002/elps.201600316

Pilbrough, W., Munro, T. P., & Gray, P. (2009). Intraclonal protein expression heterogeneity in recombinant CHO cells. *PLoS ONE*, *4*(12). https://doi.org/10.1371/journal.pone.0008432

Samoudi, M., Masson, H. O., Kuo, C. C., Robinson, C. M., & Lewis, N. E. (2021). From omics to cellular mechanisms in mammalian cell factory development. *Current Opinion in Chemical Engineering, 32*. https://doi.org/10.1016/j.coche.2021.100688

Singh, A. M., Chappell, J., Trost, R., Lin, L., Wang, T., Tang, J., Wu, H., Zhao, S., Jin, P., & Dalton, S. (2013). Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem Cell Reports*, *1*(6), 532–544. https://doi.org/10.1016/j.stemcr.2013.10.009

Tang, P. A., Wu, H. J., Xue, H., Ju, X. R., Song, W., Zhang, Q. L., & Yuan, M. L. (2017). Characterization of transcriptome in the Indian meal moth Plodia interpunctella (Lepidoptera: Pyralidae) and gene expression analysis during developmental stages. *Gene*, *622*, 29–41. https://doi.org/10.1016/j.gene.2017.04.018

Templeton, N., Dean, J., Reddy, P., & Young, J. D. (2013). ARTICLE Peak Antibody Production is Associated With Increased Oxidative Metabolism in an Industrially Relevant Fed-Batch CHO Cell Culture. *Biotechnol. Bioeng*, *110*, 2013–2024. https://doi.org/10.1002/bit.24858

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J. R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., … Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, *352*(6282), 189–196. https://doi.org/10.1126/science.aad0501

Tzani, I., Herrmann, N., Carillo, S., Spargo, C. A., Hagan, R., Barron, N., Bones, J., Shannon Dillmore, W., & Clarke, C. (2021). Tracing production instability in a clonally derived CHO cell line using single-cell transcriptomics. *Biotechnology and Bioengineering*, *118*(5), 2016–2030. https://doi.org/10.1002/bit.27715

Wlaschin, K. F., & Hu, W. S. (2007). Engineering cell metabolism for high-density cell culture via manipulation of sugar transport. *Journal of Biotechnology*, *131*(2), 168–176. https://doi.org/10.1016/j.jbiotec.2007.06.006

Zhu, J. (2012). Mammalian cell protein expression for biopharmaceutical production. *Biotechnology Advances*, *30*(5), 1158–1170. https://doi.org/10.1016/j.biotechadv.2011.08.022

ITQB NOVA, Avenida da República, 2780-157 Oeiras, Portugal | Tel +351 214 469 100 | itqb.unl.pt