universität wien

# DISSERTATION | DOCTORAL THESIS

Titel | Title

## Modelling of Adeno-Associated Virus-Production Processes using HEK293-specific Genome-scale Metabolic Models


verfasst von | submitted by

Dipl.-Ing. Leopold Zehetner BSc MSc


angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2025

| | |
|---|---|
| Studienkennzahl lt. Studienblatt \| Degree programme code as it appears on the student record sheet: | UA 796 605 419 |
| Dissertationsgebiet lt. Studienblatt \| Field of study as it appears on the student record sheet: | Chemie |
| Betreut von \| Supervisor: | Univ.-Prof. Dipl.-Ing. Dr. Jürgen Zanghellini |

# Acknowledgments

Above all, I want to express my deep gratitude to **Univ.-Prof. Dipl.-Ing. Dr. Jürgen Zanghellini** for his inspirational way of teaching, supervision, and mentoring. His enthusiastic manner motivated me throughout my PhD project and will have a major impact on my future career. Besides his invaluable support, which was a cornerstone of this project, I am exceptionally grateful for his advice and feedback throughout the journey.

Secondly, I would like to thank my manager at Takeda Austria, **DI Dr. Juan A. Hernandez Bort**, who supported and supervised me whenever possible and necessary. His exceptional leadership skills allowed him to guide me through my PhD project. Furthermore, he was always open to discussing any problems or ideas and encouraged me to become an independent researcher. Without his insights, mentoring, and supervision, it would not have been possible for me to complete my PhD project on time.

Next, I would like to sincerely thank my co-supervisor, **Mgr. Diana Szeliova, PhD**, for the inspiring meetings, through which I gained new knowledge, and for always being available whenever I needed support. Her patience with me was incredible, especially during the final stages of my projects. I admire her commitment to maintaining high standards, which makes her an indispensable researcher and colleague at the University of Vienna.

I also want to express my deep gratitude to **Dr. Barbara Kraus, MBA**, for the trust she placed in me from the very beginning of my project. Her strong leadership and dedication to her employees made her an ideal role model for me.

Additionally, I would like to thank my colleagues at the University of Vienna for creating a great atmosphere and for their insightful presentations. I am also grateful to my former colleagues at Takeda Austria for their support, especially Manfred Ostermann, Katharina Portner, Bernd Innthaler, Peter Eisenhut, Silvain Moser, Dmytro Iurashev, Lucia Micutkova, Felix Fuchsberger, Peter Andorfer, Ronald Leitner, and all other colleagues from Orth/Danube.

Next, I want to thank my girlfriend, **Victoria Kafol, LLM**, for her unwavering support in daily life and her patience—especially during the final months of my PhD project. I love you!

From my family, I want to particularly acknowledge the support and guidance of my first-ever mentor and older brother, **Michael Zehetner, MBA**, who helped shape the person I am today and laid the foundation for my development and progress in life. Furthermore, I want to thank my parents, **Michaela** and **Prof. Helmut Zehetner**, for giving me the opportunity to pursue an academic career, as well as my siblings, **Nikolaus Zehetner, MSc**, and **Antonia Zehetner**, for their endless support.

Last but not least, I want to express my gratitude to my best friend, **DI Arch. Christoph Bonholzer**, for our numerous meetings over the years, always being available for discussions and personal conversations. Additionally, I would like to thank all my friends for their support.

# Abstract

Recombinant adeno-associated virus (rAAV)-based gene therapy (GT) holds significant promise for treating genetic disorders, yet the high costs associated with its production remain a major bottleneck. Improving rAAV manufacturing efficiency is crucial for expanding its clinical applicability. This dissertation aims to address this challenge by implementing HEK293 strain-specific genome-scale metabolic models (GSMMs) combined with multi-omics datasets to optimize rAAV production.

As a first step, we reconstructed over 80 GSMMs from two HEK293 strains across five time points. These models provided insight into metabolic differences between high- and low-producing strains, identifying HIF-1$\alpha$ as a key regulator in the low-producing cells. By inhibiting HIF-1$\alpha$, we successfully stalled cellular growth and increased specific rAAV capsid production by 2.5-fold. However, this intervention also led to a simultaneous reduction in viral genome numbers, as HIF-1$\alpha$ plays a critical role in the pentose phosphate pathway (PPP), which is essential for nucleotide biosynthesis. These findings highlight the complex trade-offs in metabolic engineering for rAAV production.

To systematically compare and analyze the reconstructed GSMMs, we employed logistic principal component analysis (LPCA), a method specifically designed to handle large sets of GSMMs. LPCA allows for efficient clustering and identification of metabolic pathways driving observed differences, which is particularly useful in the context of automated GSMM reconstructions. This method enhances the interpretability of GSMMs, facilitating targeted interventions for improving rAAV production.

Beyond HEK293 cells, virus-like particle (VLP) production in various host cell lines presents additional challenges in GT applications. The need for robust and scalable production platforms is exacerbated by the variability in host cell metabolism, as well as the impact of viral vector replication on cellular pathways. The review outlines the state-of-the-art methodologies for GSMM integration into VLP-producing cell lines, addressing current limitations and highlighting opportunities for process optimization.

Overall, this dissertation contributes to the advancement of rAAV and VLP production by integrating computational modeling with multi-omics approaches, paving the way for more efficient and cost-effective GT manufacturing strategies.

# Abstract (deutsch)

Recombinant adeno-associated virus (rAAV)-basierte gene therapy (GT) bietet vielversprechende Möglichkeiten zur Behandlung genetischer Erkrankungen, jedoch bleiben die hohen Produktionskosten ein wesentliches Hindernis. Die Verbesserung der rAAV-Herstellungseffizienz ist entscheidend für die Erweiterung ihrer klinischen Anwendbarkeit. Diese Dissertation zielt darauf ab, diese Herausforderung durch die Implementierung HEK293-stammspezifischer genomweiter genome-scale metabolic models (GSMMs) in Kombination mit Multi-Omics-Datensätzen zur Optimierung der rAAV-Produktion zu bewältigen.

Als erster Schritt rekonstruierten wir 80 GSMMs aus zwei HEK293-Stämmen über fünf Zeitpunkte hinweg. Diese Modelle ermöglichten Einblicke in die metabolischen Unterschiede zwischen hoch- und niedrigproduzierenden Stämmen und identifizierten HIF-1$\alpha$ als einen Schlüsselregulator in den niedrigproduzierenden Zellen. Durch die Inhibition von HIF-1$\alpha$ konnten wir das Zellwachstum erfolgreich stoppen und die spezifische rAAV-Kapsidproduktion um das 2,5-fache steigern. Diese Intervention führte jedoch gleichzeitig zu einer Verringerung der Anzahl viraler Genome, da HIF-1$\alpha$ eine entscheidende Rolle im pentose phosphate pathway (PPP) spielt, der für die Nukleotidsynthese essenziell ist. Diese Ergebnisse verdeutlichen die komplexen Kompromisse im metabolischen Engineering für die rAAV-Produktion.

Um die rekonstruierten GSMMs systematisch zu vergleichen und zu analysieren, verwendeten wir die logistic principal component analysis (LPCA), eine Methode, die speziell für die Verarbeitung großer GSMM-Datensätze entwickelt wurde. LPCA ermöglicht eine effiziente Clusterbildung und Identifikation der Stoffwechselwege, die die beobachteten Unterschiede bestimmen, was insbesondere im Kontext automatisierter GSMM-Rekonstruktionen von Bedeutung ist. Diese Methode verbessert die Interpretierbarkeit metabolischer Modelle und erleichtert gezielte Eingriffe zur Steigerung der rAAV-Produktion.

Über HEK293-Zellen hinaus stellt die Produktion von virusähnlichen Partikeln virus-like particles (VLPs) in verschiedenen Wirtszelllinien zusätzliche Herausforderungen für GT-Anwendungen dar. Der Bedarf an robusten und skalierbaren Produktionsplattformen wird durch die Variabilität des Wirtszellstoffwechsels sowie die Auswirkungen der viralen Vektorreplikation auf zelluläre Signalwege verstärkt. Diese Übersicht beschreibt die modernsten Methoden zur Integration von GSMMs in VLP-produzierende Zelllinien, geht auf bestehende Einschränkungen ein und hebt Optimierungsmöglichkeiten hervor.

Insgesamt trägt diese Dissertation zur Weiterentwicklung der rAAV- und VLP-Produktion bei, indem sie computergestütztes Modellieren mit Multi-Omics-Ansätzen kombiniert und so den Weg für effizientere und kostengünstigere Produktionsstrategien in der GT ebnet.

# Contents

# 1 Introduction

## 1.1 Gene therapy

Gene therapy (GT) is an advanced form of treatment that changes the genetic makeup of a patient's cells to cure genetic diseases. This approach is categorized into two principal modalities: *in vivo* and *ex vivo* GT [1]. *Ex vivo* GT is a method where cells are extracted from a patient, genetically modified outside the body, and then reintroduced to treat a disease [2]. In *in vivo* GT, the therapeutic gene is directly administered to the patient, targeting specific cell populations to rectify underlying genetic dysfunctions that contribute to disease pathogenesis, such as spinal muscular atrophy, caused by a mutation in the *SMN1* gene [2,3]. Virus-like particles (VLPs), lipid nanoparticless (LNPs), and extracellular vesicles (EVs) are some of the vector systems that are needed to deliver genes effectively. VLPs mimic viral architecture to enhance cellular uptake; LNPs provide synthetic and tunable gene delivery frameworks; and EVs serve as endogenous carriers, facilitating the intercellular transport of genetic material. VLP-based vectors are the only vectors that have been approved by the authorities for use in patients, which shows that they are safe and effective in clinical settings [1–3].

The global market for cell and gene therapy is experiencing significant expansion and is projected to increase substantially in the coming years. In 2023, the combined market cap for cell and GT was estimated at approximately 18.9 billion USD, with an increase to 21 billion USD in 2024 and a projected valuation of 60 billion USD by 2030 (Fig 1 a) [4]. Effective viral vectors are important for GT to work in the clinic, with recombinant adeno-associated virus (rAAV) vectors currently being the most popular choice. The market for rAAV-based GT manufacturing is expected to grow from 1.02 billion USD in 2024 to 2.78 billion USD by 2030 (Fig 1 a) [5]. This trajectory is also reflected in regulatory approvals, with seven FDA-approved (Food and Drug Administration) GT drugs utilizing VLP-based rAAV vectors (Fig 1 c). Four of the approved rAAV-based drugs received FDA approval in 2022 and 2023, showing an accelerated trend in rAAV-based GT innovations [3]. This trend aligns with academic research output, as evidenced by an increase in publications focusing on rAAV biology (CAGR (Compound Annual Growth Rate): 12.28%) and manufacturing (CAGR: 10.02%), indicating sustained efforts to optimize rAAV platforms in response to the expanding GT market (Fig 1 b).

## 1.2 Adeno-associated virus

### 1.2.a Biology of AAV

rAAV was first discovered in the early 1960s as a contaminant in adenovirus preparations, leading to its classification as a member of the *Parvoviridae* family [6]. Initially considered a defective virus, rAAV requires a helper virus, such as adenovirus, for replication, which has shaped its unique life cycle and biological characteristics [7].

rAAV is a small, non-enveloped virus with a single-stranded DNA genome [3,7]. The genome size of rAAV typically ranges from 4.7 to 5.0 kilobases (kbs), which is smaller than many other viral vectors, such as adenoviruses and lentiviruses, making it suitable for applications where compactness is advantageous [8]. rAAVs are classified into multiple serotypes, with rAAV 1 to rAAV 9 being most studied [9]. Each serotype exhibits distinct tissue tropisms, which allows for targeted gene delivery to specific cell types [9,10]. The advantages of rAAV as a gene delivery vector are numerous: rAAVs are known for their low immunogenicity, which is especially important for patients with immune deficiency [8,11]. rAAV can transduce both dividing [12] and non-dividing cells [13], allowing for versatile applications in GT. Moreover, rAAV vectors can achieve stable long-term expression of transgenes, particularly in postmitotic tissues, without the need for integration into the host genome [14,15]. This characteristic is particularly advantageous compared to lentiviral vectors, which integrate into the host genome and can pose the risks of insertional mutagenesis [8].

However, rAAVs also have notable disadvantages. One significant limitation is their relatively small packaging capacity, which restricts the size of transgenes that can be delivered [1,3,8], and the

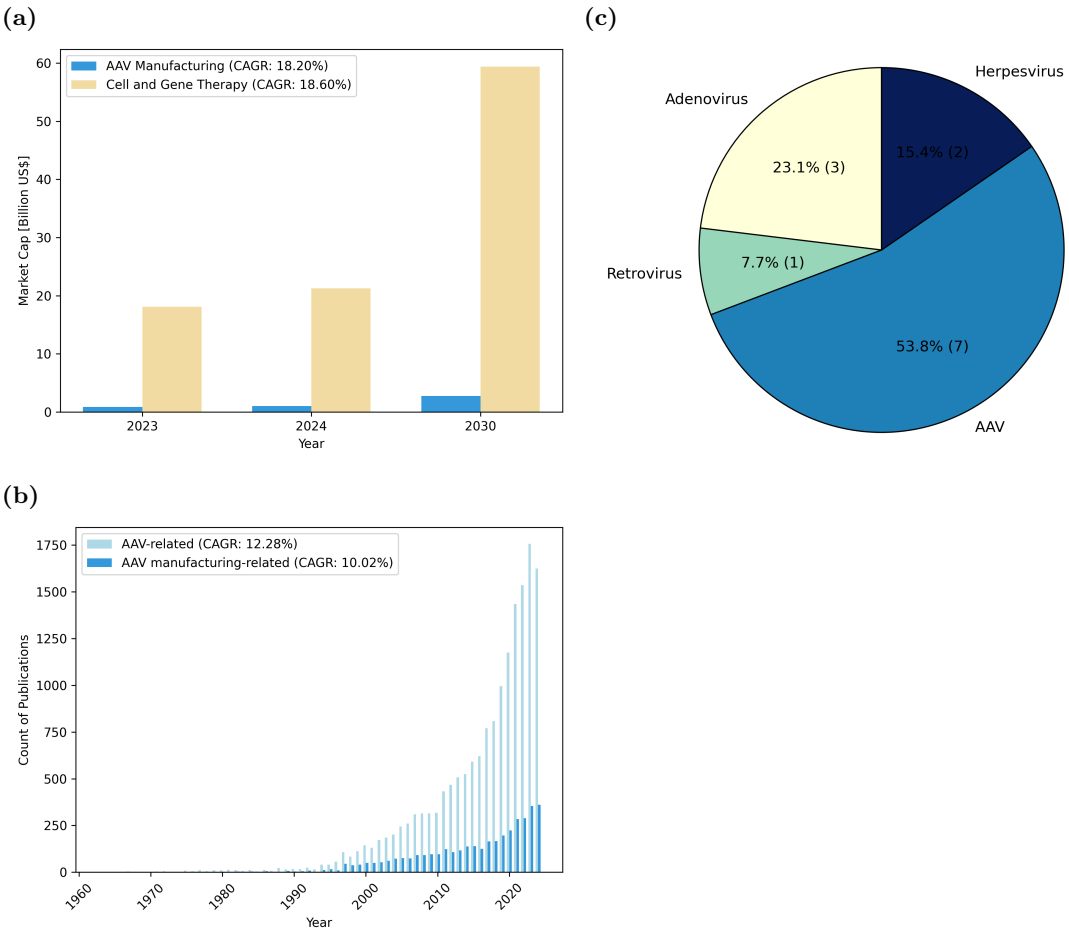


**Fig 1 . Industrial and academic relevance of rAAV manufacturing and cell and GT market.** (a) Predicted growth of rAAV manufacturing market (blue), and cell and GT market (yellow), both expecting an annual average growth of 18%. (b) FDA-approved *in vivo* GT drugs grouped by viral vector. (c) Number of publications related to rAAV (light blue), and rAAV manufacturing (dark blue), both with an average growth of over 10% per year. Data were obtained as of January 10th, 2025 from PubMed (https://pubmed.ncbi.nlm.nih.gov/).

scale-up of rAAV production remains a bottleneck in meeting the increasing demand for sufficient rAAV quantities for widespread clinical applications [16, 17]. Additionally, the presence of pre-existing neutralizing antibodies in the human population can hinder the efficacy of rAAV-based therapies, as these antibodies can bind to the viral capsid and prevent successful transduction [18]. This issue is particularly pronounced in rAAV 2, which is the most widely used serotype [19]. Recent advancements, such as the encapsulation of rAAVs in EVs, have shown promise in overcoming these limitations by enhancing stability and evading neutralizing antibodies, thus improving gene delivery efficacy [20].

In comparison to other gene delivery vectors, such as adenoviruses, lentiviruses, LNPs, and EVs, rAAVs offer a unique set of benefits and challenges: Adenoviruses can deliver larger transgenes [1] and do not require helper viruses for replication; however, they are often associated with stronger immune responses and potential toxicity [21]. Lentiviruses, while capable of stable integration and long-term expression, carry risks of insertional mutagenesis [22]. LNPs provide a non-viral alternative but may lack the efficacy and tissue-specificity of rAAVs in certain contexts [23]. Additionally, rAAV vectors have already received regulatory approval for clinical use, with several rAAV-based drugs

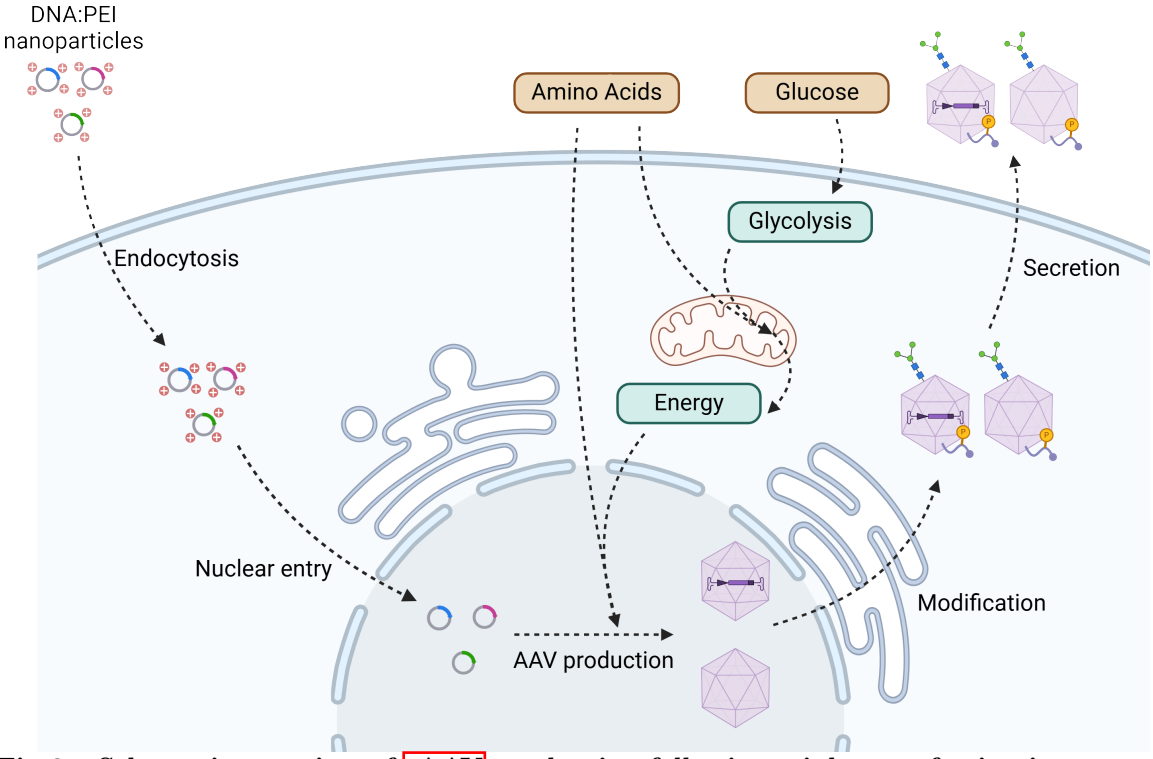


**Fig 2 . Schematic overview of rAAV production following triple transfection in mammalian cells.** DNA:PEI nanoparticles are taken up via endocytosis. After entering the nucleus, rAAV are produced using resources, and energy from the host cell. Following post-translational modification, rAAV are released via secretion or cell lysis.

approved by authorities for the treatment of genetic disorders, while EVs and LNPs are still under investigation and have yet to achieve similar regulatory status in *in vivo* GT [1,3].

### 1.2.b Production of rAAV based on the triple transfection process

Currently, the most common platform for the production of rAAV are HEK293 (Human embryonic kidney) cells using transient triple transfection, a process in which three plasmids are introduced into cells to allow viral replication, genome packaging, and capsid formation (see Fig 2) [1,3,24,25]. The triple transfection system consists of a transfer plasmid, a packaging plasmid, and a helper plasmid. The transfer plasmid contains the gene of interest (GOI) flanked by inverted terminal repeatss (ITRs), which are required for rAAV genome replication and packaging. The packaging plasmid encodes the rAAV replication (Rep) and capsid (Cap) proteins, which are essential for viral genome replication and capsid formation. The helper plasmid supplies the necessary adenoviral (Fig 3) or herpes simplex virus type 1 (HSV-1) genes that provide helper functions required for efficient rAAV replication [1,3,25].

Specifically, the Rep proteins play a critical role in rAAV genome replication and packaging. The large Rep proteins, Rep78 and Rep68, bind to the ITR sequences of the viral genome and initiate replication through their helicase and endonuclease activities [26]. The small Rep proteins, Rep52 and Rep40, are primarily involved in genome packaging and strand displacement during the final stages of rAAV assembly [26,27]. The Cap proteins, consisting of viral protein (VP) 1, VP 2, and VP 3, self-assemble to form a 60-subunit icosahedral capsid in the ratio 1:1:10 (VP 1:VP 2:VP 3). The VP 1 protein possesses phospholipase A2 (PLA2) activity, which is necessary for endosomal escape during viral entry, while VP 2 and VP 3 contribute to receptor binding and transduction

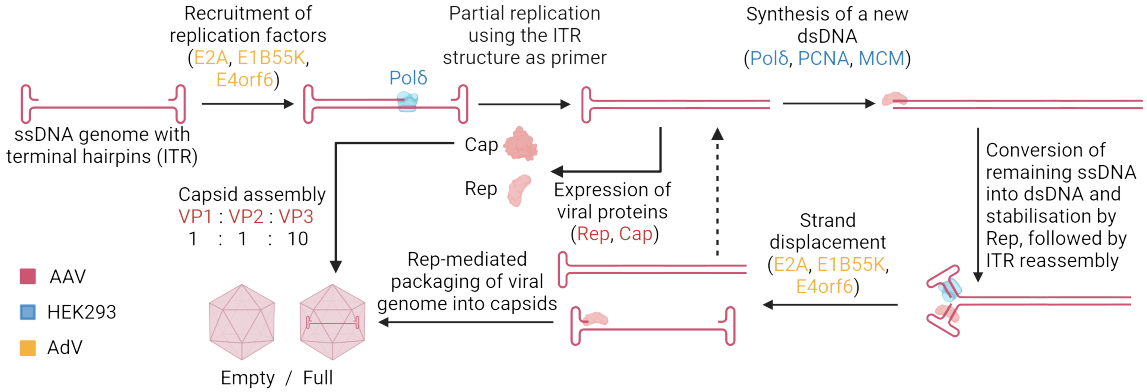


**Fig 3 . Schematic overview of rAAV replication in mammalian cells.** Starting from a ssDNA AAV genome, the replication process requires AAV-specific proteins (red), host proteins (blue), and helper virus proteins (yellow), such as Adenoviruses.

efficiency [27]. The assembly-activating protein (AAP) facilitates the proper folding and assembly of the rAAV capsid [28].

rAAV requires helper virus functions for efficient replication and transcription. These helper functions are typically provided by adenoviral genes or, alternatively, by HSV-1 genes. In the adenovirus-based system, the *E1A* gene promotes cell cycle progression and enhances Rep/Cap gene expression, while *E4orf6*, and *E1B55K* facilitate the degradation of *p53*, thereby allowing viral replication to proceed efficiently, and inhibiting cellular apoptosis. The adenoviral *E2A* gene further contributes by supporting viral DNA replication and inhibiting the antiviral PKR response (see Fig 3) [29]. Alternatively, HSV-1 helper genes can be used to support rAAV replication. The *UL5*, *UL8*, and *UL52* genes encode the helicase-primase complex, which unwinds DNA and synthesizes primers for replication. *ICP0* acts as an *E3* ubiquitin ligase, targeting cellular restriction factors for degradation, thereby enhancing viral replication. *ICP4* serves as a transcriptional activator, ensuring high-level expression of viral genes by recruiting host polymerases. Additionally, the HSV-1 DNA polymerase, encoded by *UL30*, and its processivity factor, *UL42*, facilitate the replication of rAAV genomes in the absence of adenoviral helper functions [29,30].

In addition to viral factors, several host cell enzymes are required for efficient rAAV replication. The minichromosomal maintenance complex (MCM) functions as a DNA helicase, unwinding the rAAV genome to facilitate replication. Proliferating cell nuclear antigen (*PCNA*) serves as a sliding clamp that enhances the processivity of cellular DNA polymerases. Furthermore, rAAV relies on host DNA polymerase $\delta$ to synthesize new genome copies, using the self-priming mechanism provided by the ITRs (see Fig 3) [31].

### 1.2.c Post-translational modification (PTM) of rAAV

Across the rAAV 1-rh10 serotypes, 52 PTMs were identified, with glycosylation being the most prevalent (36%), followed by phosphorylation (21%), ubiquitination (17%), SUMOylation (13%), and acetylation (11%) [32]. These modifications play important roles in regulating capsid function, stability, and transduction efficiency [33]. Genetic engineering showed the necessity of PTMs for rAAV. Mutation of the glycosylation site in rAAV 2 (N253Q) resulted in a 100-fold reduction in vector yield, indicating that glycosylation is crucial for viral capsid assembly and stability. Mutation of the SUMOylation site in rAAV 2 (K258Q) led to a 28% reduction in transduction *in vitro* and completely abolished *in vivo* transduction in the mouse retina, suggesting that SUMOylation plays a role in nuclear transport and transcriptional activation. Furthermore, mutation of the ubiquitination site in rAAV 2 (K544E) increased viral transduction by preventing proteasomal degradation, indicating that ubiquitination functions as a degradation signal [32]. Regarding glycosylation, nine

rAAV serotypes (rAAV 1–rAAV 9) exhibited N-glycosylation with comparable overall profiles. The dominant feature was the high abundance of mannosylated N-glycans, comprising approximately 74–83% of total N-glycans. Additionally, fucosylated N-glycans were present at 23–40%, while sialylation accounted for 10–17%. No significant glycan differences were observed between empty and full rAAV 2 capsids, suggesting that glycosylation is an intrinsic property of the capsid rather than being influenced by the presence of vector genomes [34]. The glycan structures on the rAAV capsid surface play a crucial role in tissue tropism by mediating interactions with cell surface receptors, as well as in cellular uptake, intracellular trafficking, immune recognition, and vector stability. A deeper understanding of N-glycosylation differences among rAAV serotypes may aid in the rational design of rAAV vectors for GT by enabling the selection of serotypes with favorable glycan structures for specific applications [34]. rAAV 6 is primarily O-glycosylated rather than N-glycosylated. O-glycosylation affects the capsid composition by altering the VP 1 to VP 2/VP 3 ratio. Since VP 1 contains a phospholipase A2 domain that is essential for endosomal escape, glycosylated rAAV 6 exhibited lower transduction efficiency. *In vitro* studies showed that non-glycosylated rAAV 6 had a transduction efficiency of 94.4%, whereas glycosylated rAAV 6 only achieved 10.0%. Similarly, *in vivo* experiments demonstrated that glycosylated rAAV 6 led to lower levels of human factor IX (*hFIX*) expression, further confirming its reduced functional efficiency [35]. rAAV 8 is glycosylated specifically on the VP 3 protein. The presence of terminally galactosylated N-glycans suggests potential interactions between rAAV 8 and host-cell receptors. This glycosylation may influence its tissue tropism, transduction efficiency, and immunogenicity, highlighting its potential role in vector performance and host interactions [36]. Early studies initially classified glycans as "primary receptors" for rAAV; however, they are now recognized as "attachment factors" rather than true receptors. Different rAAV serotypes exhibit distinct glycan specificities, influencing their interactions with host cells. For example, rAAV 2, rAAV 3B, rAAV 6, and rAAV 13 preferentially bind to heparan sulfate, while rAAV 1, rAAV 5, rAAV 6, and rAAV 4 recognize sialic acid. In contrast, rAAV 9 primarily interacts with galactose. These glycan interactions play a crucial role in increasing rAAV concentration at the plasma membrane, thereby facilitating engagement with cellular receptors and enhancing viral entry [37].

### 1.2.d Full-empty ratio

One of the key factors during the production of rAAV particles is the full-empty-ratio (F/E ratio). This ratio refers to the proportion of rAAV particles that contain a full-length genome (full capsids) compared to those that are devoid of genetic material (empty capsids) or carry truncated DNA sequences. A high F/E ratio is desirable because empty capsids do not contribute to therapeutic efficacy and can dilute the effectiveness of the treatment by competing with full capsids for cellular uptake and receptor binding [38–40]. The existence of a F/E ratio arises from the inherent complexities of the rAAV packaging process (Fig 3). During the assembly of rAAV particles, not all capsids successfully encapsulate the viral genome. Several factors can influence the F/E ratio. Firstly, the genetic payload size is a significant determinant; larger genomes tend to have lower packaging efficiency, resulting in a higher proportion of empty capsids [27]. Secondly, the production system used (e.g., mammalian cells vs. insect cells) can affect the assembly and packaging efficiency of rAAV particles. For example, insect cell systems may yield higher F/E ratios due to differences in PTMs, and capsid assembly dynamics [41–43]. Furthermore, the optimization of capsid protein ratios (VP 1:VP 2:VP 3) during the production process can enhance the packaging efficiency of full genomes [42]. Recently, a mechanistic model has been used to identify bottlenecks during triple transfection for rAAV 5 production. Only about 5% of the initial plasmid input is taken up by cells, with only 0.6% reaching the nucleus, limiting rAAV production due to inefficient nuclear entry and plasmid degradation [25].

Capsid production peaks at 24 hours post-transfection, while viral genome replication continues beyond this point, resulting in newly synthesized DNA not being efficiently packaged into capsids. This mismatch between capsid production and DNA replication contributes to a low F/E ratio, limiting the overall yield of functional rAAV particles. Rep protein plays a dual role by catalyzing

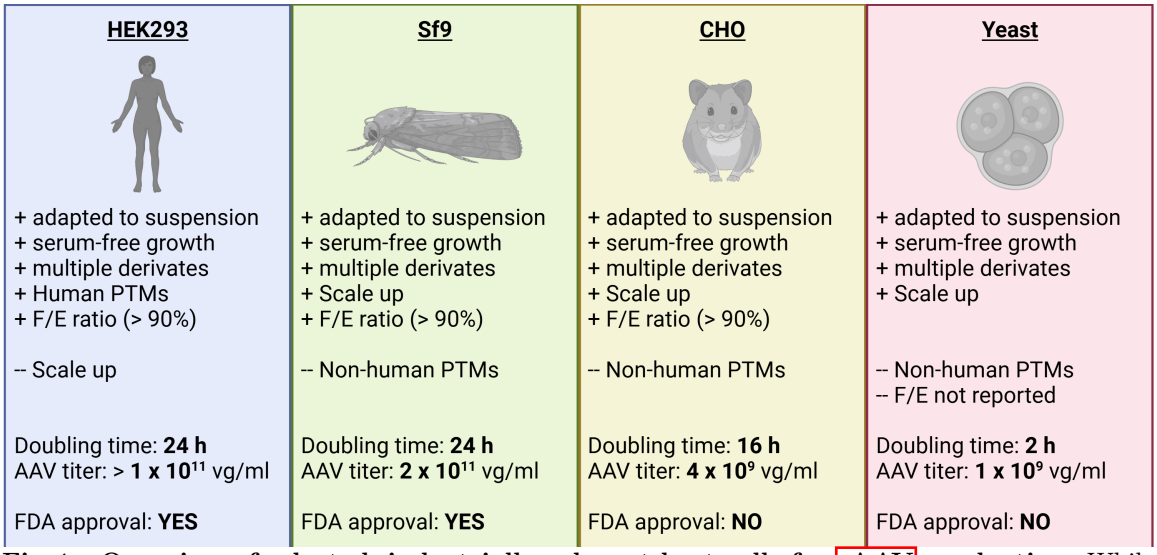


**Fig 4 . Overview of selected, industrially relevant host cells for rAAV production.** While HEK293 cells are the most important production hosts for rAAV production, especially due to the human PTM signature, other host cells bear significant advantages, such as higher growth rates (Chinese hamster ovary (CHO), *S.cerevisiae*) or rAAV titers (Sf9).

viral DNA replication and packaging while also repressing cap gene expression, which can limit capsid production and reduce overall rAAV yield [25]. A potential reason for the earlier expression of capsid genes might be the earlier uptake of rep/cap, and helper plasmids compared to the transgene plasmid [25], which is the reason why the authors suggested time-staggering, with transgene plasmids added earlier to the cells, while other plasmids could be added later to synchronize expression [25].

## 1.3 rAAV production in different host cells

For the production of rAAV vectors various host cells can be utilized for this purpose, including HEK293 [1], Sf9 [44], HeLa [45], A549 [46], CHO [47], *S.cerevisiae* [48], and *Brassicaceae* [49]. Each of these hosts offers distinct advantages and challenges in the context of rAAV production. However, A549 and HeLa cells are derived from tumors and should not be used for the manufacturing of recombinant drugs [46,50] (Fig 4).

### 1.3.a rAAV production in HEK293 cells

rAAV production has been significantly improved through optimized transient transfection methods in suspension HEK293 cells [51–53]. One approach adapted HEK293 cells to serum-free suspension culture, achieving $> 1 \cdot 10^{10}$ vg/mL post-purification [53]. Implementing a perfusion system further increased total yields to $1.1 \cdot 10^{11}$ vg/mL from a 10 L WAVE bioreactor by continuously harvesting vectors from the media [24]. A design-of-experiment strategy to optimize plasmid ratios and transfection conditions lead to a production of $3.64 \cdot 10^{11}$ vg/mL for AAV9 and $4.46 \cdot 10^{9}$ vg/mL for AAV2 in small-scale cultures. Scaling up to 125 mL cultures resulted in $2.23 \cdot 10^{10}$ vg/mL for AAV9 [52]. A different study using orbitally shaken HEK293 cells reported titers of $> 1 \cdot 10^{8}$ vg/mL, allowing scale up to 1000 L [51].

### 1.3.b rAAV production in Sf9 cells

Sf9 cells, derived from the fall armyworm *Spodoptera frugiperda*, are a widely utilized insect cell line in biotechnology, particularly for the production of recombinant proteins (RPs) and VLPs, including

rAAV. The Sf9 cell line is part of the baculovirus expression vector system (BEVS), which exploits the natural ability of baculoviruses to infect insect cells and facilitate high-level expression of RPs [54, 55]. Sf9 cells are favored for their rapid growth, the ability to grow in suspension cultures even in high density [56], and capacity for PTMs similar to those in mammalian cells, making them suitable for producing complex biopharmaceuticals [44, 57]. Additionally, Sf9 cells have a robust metabolic profile that supports the efficient production of RPs as they preferentially utilize glutamine over asparagine, which is generally benefitial for growth and energy production, particularly in the tricarboxylic acid cycle (TCA) cycle [58]. Furthermore, Sf9 cells are less prone to contamination with human pathogens, providing a safer alternative for producing therapeutic proteins [54].

BEVS utilizing insect cells, particularly the Sf9 cell line have been adapted to modified systems within BEVS framework to enhance rAAV production, including OneBac [59], Monobac [60], ExpiSf9 [61], and stable Sf9 cell pools [62]. For example, the ExpiSf9 system is a commercially available platform that utilizes a chemically defined medium for the growth of Sf9 cells [61]. The OneBac system has been reported to achieve significantly higher rAAV titers compared to traditional Sf9 systems giving the currently highest reported titers with $> 2 \cdot 10^{11}$ vg/mL in a 50 L bioreactor [59].

### 1.3.c rAAV production in CHO cells

CHO cells are the most important mammalian cell line in biopharmaceutical production [63, 64], and are primarily used for the production of monoclonal antibodies and RPs. Recently, CHO strains were engineered to be infectible by HSV-1 viruses to allow rAAV production. The top-performing clone, CHO-HV-C1, demonstrated high production yields, achieving titers of $1.62 \cdot 10^9$ vg/mL for rAAV6.2-GFP, $2.51 \cdot 10^9$ vg/mL for rAAV 8-GFP, and $4.07 \cdot 10^9$ vg/mL for rAAV 9-GFP. The best results were obtained with an MOI of 4:1 (rHSV-1 to CHO cells), while further increasing the MOI did not lead to significant improvements in yields. Additionally, a temperature shift from 37°C to 33°C significantly improved cell viability and rAAV yield, a finding consistent with previous studies that reported greater HSV-1 stability and replication efficiency at lower temperatures. CHO-derived rAAV 6.2-GFP and rAAV 9-GFP contained 91% and 79.5% full capsids [47], respectively, a similar percentage compared to the 90% recently seen in HEK293-produced rAAVs [52]. Further characterization of the capsid proteins confirmed proper assembly, with VP 1:VP 2:VP 3 ratios matching those of naturally occurring rAAV capsids, unlike the BEV-Sf9 system, which often requires genetic modifications to achieve proper capsid formation. *In vitro* infectivity tests demonstrated that CHO-derived rAAVs exhibited transduction efficiencies comparable to those produced in HEK293 cells using triple transfection. The rAAV were tested for biodistribution in mice, where they showed effective transduction in major organs, including the liver, heart, and lungs [47].

### 1.3.d rAAV production in yeast (*S. cerevisiae*)

rAAV production in yeast has been explored as an alternative to mammalian and insect cell-based systems. Several studies have demonstrated that yeast can express rAAV capsid proteins, assemble rAAVs, and generate single-stranded DNA (ssDNA) genomes, though with significantly lower efficiency than in human cells. In a first attempt, it has been demonstrated that yeast can replicate an rAAV genome and generate ssDNA from a circular plasmid containing ITRs and the Rep68 protein. This process, however, did not follow the conventional rAAV replication pathway observed in mammalian cells. Yeast could serve as a cost-effective, genetically tractable system for understanding the molecular biology of rAAV, though further optimization is required to achieve efficient replication and packaging [65]. In a following study the ability of yeast to assemble rAAV capsids by expressing the VP 1 and VP 3 structural proteins was investigated. The correct stoichiometry was achieved using a combination of the natural p40 promoter and the inducible yeast GAL1 promoter, with optimal induction conditions identified as 4.5 hours in a medium containing 0.5% glucose and 5% galactose. The rAAVs were compared to confirm that their morphology was

similar to rAAV capsids produced in mammalian cells [66]. Finally, the expression of rAAV capsid proteins and ssDNA was accomplished. Various promoters were tested, and the GAL1/10 bidirectional promoter provided the best results for VP 1, VP 2, and VP 3 expression. Additionally, attempts were taken to enhance rAAV genome replication by introducing adenoviral helper proteins, *E4orf6* and *E1b55k*, which are known to facilitate rAAV replication in human cells. However, their expression in yeast did not improve ssDNA formation. While yeast could generate rAAV capsids, both capsid assembly and ssDNA formation were significantly less efficient, approximately 1000-fold and 100-fold ($1 \cdot 10^9$ vg/ml) lower, respectively, than in human cells [48]. Although yeast offers advantages such as ease of genetic manipulation, cost-effectiveness, and rapid growth, it is not yet an efficient platform for large-scale rAAV vector production. The primary limitations include low capsid assembly efficiency, reduced ssDNA formation, and the lack of a clear replication mechanism comparable to that in mammalian cells. Nevertheless, these studies provide a foundation for further research aimed at improving rAAV production in yeast, potentially making it a viable alternative to current mammalian and insect cell-based systems.

In summary, HEK293 cells remain the predominant host for rAAV production in the biopharmaceutical industry due to their high yields, well-established protocols, and biopotency [24, 61], while alternative hosts such as Sf9 insect cells, CHO cells, and yeast offer distinct advantages. Sf9-based BEVS have demonstrated promising scalability, achieving high titers, but often require modifications to ensure proper capsid formation. CHO cells show potential as a mammalian alternative, particularly with engineered strains optimized for HSV-1-based production, however, more research on biopotency needs to be done. Although yeast presents a cost-effective and genetically tractable platform, its efficiency remains significantly lower than mammalian and insect cell-based systems, highlighting the need for further optimization before it can be considered a viable alternative.

## 1.4 HEK293 cells in biopharmaceutical production

### 1.4.a Relevance of HEK293 cells

The HEK293 cell line was established in 1973 from the kidney of an aborted human embryo via transformation with sheared Adenovirus 5 DNA [67]. As of today, thousands of HEK293 strains are registered in the cellosaurus database [68], with the majority being knockout strains. The HEK293 cell line and its derivates are essential in biotechnological research and production, valued for their ability to produce therapeutic proteins with human-like PTMs, such as glycosylation and proper folding [69, 70]. HEK293 is the most used cell line in the production of VLPs like rAAV, adenoviruses, and lentivirus for GT and vaccine development [1], as well as in manufacturing complex biologics [70]. Certain HEK293 derivatives were optimized for high-density suspension cultures, enabling large-scale production in suspension and under serum-free conditions [69, 71]. Additionally, their high transfectability and reliable expression of exogenous genes make them valuable for studying signal transduction, protein interactions, and gene expression, cementing their role as a cornerstone in both academic research and industrial bioprocessing [72].

From an economic perspective, HEK293 cell line market was valued at $1.2 billion in 2022 and is projected to reach $2.5 billion by 2030. Even more impressive is, that the HEK293 media market was valued at $6.3 billion in 2023 and is expected to reach $13.1 billion by 2030. These markets are driven by advancements in biotechnology, increasing demand for biopharmaceuticals, and cell-based therapies [73].

### 1.4.b Biology of HEK293 cells

During establishing the HEK293 cell line, a 4-kbp adenoviral genome fragment was integrated into chromosome 19, encoding proteins *E1A*, and *E1B*, which interfere with cell cycle control and apoptosis, making HEK293 cells immortal [67]. Cytogenetic analysis identified pseudotriploidity in

HEK293 cells, indicating abnormal chromosomal numbers, potentially due to extensive passaging [74], which was confirmed during a resequencing, and karyotyping effort of six HEK293 strains [75]. Over the past decade, eight HEK293 derivates have been resequenced and analyzed [75,76], including HEK293T, HEK293S, HEK293SG, HEK293SGGD, HEK293E, HEK293-H, HEK293-F, Freestyle 293-F. Among these strains, some were adapted to grow under suspensional conditions for increased growth rates, and facilitate scale-up in manufacturing. Despite several studies, describing adaption from adherent to suspensional growth of HEK293 strains [71,75,76], several questions remain to be solved how cells change during this adaption. When comparing five progeny strains to the parental HEK293 strain, downregulation of cell adhesion, extracellular matrix organization, and cell motility was observed, while amino acid metabolism and cholesterol biosynthesis were upregulated in progeny cells, independent of the growth conditions. Strains grown in suspension had a decreased retinol metabolism, oxidative phosphorylation (OXPHOS), and nucleotide metabolism [76], indicating that the adaption process is more prevalent for HEK293 cells than the production state [77].

In terms of production of RPs and VLPs, HEK293 cells provide several advantages over other mammalian cells:

- One of the primary benefits of using HEK293 cells is their high transfection efficiency, which allows for rapid and robust production of RPs and VLPs. For instance, engineered HEK293 cell lines have been shown to support high-throughput expression of RPs with titers reaching up to 650 mg/L within seven days [78].
- The use of the cytomegalovirus promoter in HEK293 cells has been widely adopted to facilitate high-yield production of RPs, as well as lentiviral and rAAV vectors [79]. This efficiency is further enhanced by the adaptation of HEK293 cells to suspension culture, which allows for easier scaling and automation of the production process [51].
- HEK293 cells have very specific morphological features. The large nucleus of HEK293 cells, comprising approximately two-thirds of the cell volume [80], is beneficial for the production of RPs and VLPs, since it provides more space for transcription, potentially allowing for higher levels of mRNA production [72].
- HEK293 cells are capable of performing complex PTMs that are critical for the functionality of many RPs and VLPs, such as human-like glycosylation pattern. The ability to produce RPs and VLPs with native PTMs reduces the risk of immunogenicity and enhances the therapeutic potential [69,81].
- Numerous therapeutic products, including RPs, VLPs, and vaccines produced in HEK293 cells, have already been approved by the FDA, making the HEK293 platform a trustworthy host for production [3,69].

## 1.5 Challenges in culturing HEK293 cells

Despite HEK293 strains have multiple advantages in the production of complex biologics, there are still a lot of challenges in culturing HEK293 cells (or mammalian cells in general) related to metabolism, endoplasmic reticulum (ER) stress, and transfection efficiency.

### 1.5.a Metabolic challenges and solutions

**Pseudohypoxia:** In bioprocesses, the activation of HIF-1$\alpha$ has been shown to negatively impair productivity in HEK293 cells [82], which is in accordance with results from our experiments [83], and makes a detailed discussion of HIF-1$\alpha$ associated metabolism indispensible.

Pseudohypoxia refers to a cellular state that mimics the physiological responses of hypoxia, despite the presence of normal oxygen levels. This phenomenon is primarily characterized by the activation of HIF-1$\alpha$, which leads to the expression of genes typically activated under low oxygen

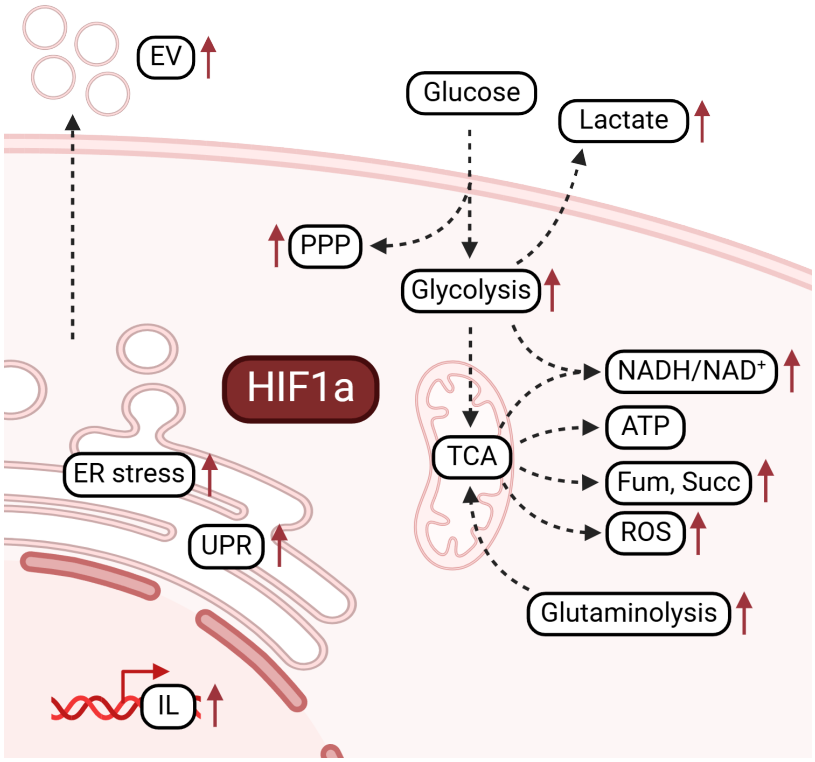


**Fig 5 . Schematic overview of cellular pathways, upregulated by HIF-1$\alpha$ activation in hypoxia or normoxia.** HIF-1$\alpha$ activation upregulates glycolysis, while downregulating overall TCA activity, HIF-1$\alpha$ enhances parts of the TCA by enhancing glutaminolysis, fuelling the TCA. Metabolic changes lead to enhanced reactive oxygen species (ROS) and $NADH$ levels. as well as to ER stress, unfolded protein response (UPR), increased EV secretion, and expression of immune response related genes (see main text for detailed description).

conditions [84]. Pseudohypoxia is often associated with various pathological conditions, including cancer, where it can promote tumor progression and therapeutic resistance by altering metabolic pathways, such as TCA, and pentose phosphate pathway (PPP), and enhancing glycolysis, even in the presence of sufficient oxygen [85,86]. HIF-1$\alpha$ is tightly controlled by prolyl hydroxylases ($PHDs$), which hydroxylate HIF-1$\alpha$, and lead to its proteasomal depletion. By inhibiting $PHDs$ using pharmacological agents ($CoCl_2$), HIF-1$\alpha$ is activated even under normoxic conditions [87]. In cell culture, various causes for HIF-1$\alpha$ stabilization, and its consequences were described, inducing pseudohypoxia without the addition of pharmacological agents (see Fig 5):

- The TCA cycle plays a central role in cellular metabolism and can significantly impact HIF-1$\alpha$ stability, with fumarate and succinate being $PHD$ inhibitors [88]. Succinate accumulation has been observed in HEK293 cells following adaptation to suspension growth, potentially indicating active HIF-1$\alpha$ signaling [71]. ROS generated from enhanced TCA cycle activity further contribute to HIF-1$\alpha$ stabilization by promoting oxidative stress, a known activator of HIF-1$\alpha$ [89,90], despite ROS not being necessary for HIF-1$\alpha$ stabilization [91]. This creates a feedback loop where HIF-1$\alpha$ activation suppresses oxidative metabolism while enhancing glycolytic pathways, thereby perpetuating pseudohypoxia and supporting tumor growth [88]. Despite TCA activity and OXPHOS are reduced during HIF-1$\alpha$ activation, HIF-1$\alpha$ enhances glutaminase ($GLS$) expression, promoting glutaminolysis as an adaptive mechanism in rapidly proliferating cells [92]. Forming a positive feedback loop, glutaminolysis provides alternative

carbon sources for the TCA cycle, and its upregulation can contribute to HIF-1$\alpha$ stabilization by producing TCA intermediates [93].

- Enhanced glycolysis leads to the accumulation of intermediates such as pyruvate and lactate, and contributes to HIF-1$\alpha$ stabilization [94], reinforcing a feedback loop that further upregulates glycolysis [95]. In turn, HIF-1$\alpha$ suppresses *PDH* activity through *PDK* upregulation, limiting TCA cycle flux and reinforcing glycolytic dependence [96]. Enhanced glycolysis leads to reductive stress by shifting the $NADH/NAD^+$ ratio towards $NADH$ [97].

- Inflammatory cytokines such as TNF-$\alpha$ and IL-1$\gamma$ can directly induce HIF-1$\alpha$ stabilization, contributing to metabolic reprogramming associated with chronic inflammation and tumor progression [98, 99]. HIF-1$\alpha$, in turn, enhances the expression of pro-inflammatory cytokines, promoting a self-sustaining inflammatory microenvironment [100]. This interplay is crucial in cancer and immune cell function, where inflammation-driven metabolic shifts favor glycolysis over OXPHOS.

- The activation of the PPP in pseudohypoxia is primarily driven by the increased demand for ribose-5-phosphate, a product of the PPP that is necessary for nucleotide synthesis due to high growth rates. The stabilization of HIF-1$\alpha$ upregulates genes involved in the PPP, such as *G6PD*, which is the rate-limiting enzyme in the PPP [101, 102]. The enhanced glycolytic flux provides the necessary substrates for the PPP, thereby facilitating nucleotide synthesis in cancer cells [103].

- HIF-1$\alpha$ regulates the expression of genes involved in the UPR. For example, it has been demonstrated that HIF-1$\alpha$ activation can lead to increased expression of C/EBP homologous protein (*CHOP*), a critical transcription factor in the UPR that promotes apoptosis when ER stress is severe [104–106] (see chapter 1.4.2).

- Hypoxia-induced HIF-1$\alpha$ activation in HEK293 cells significantly increased EV release, while HIF-1$\alpha$ knockdown reduced this effect, indicating its crucial role in EV secretion under low oxygen conditions [107]. Another study demonstrated that pharmacological stabilization of HIF-1$\alpha$ enhanced EV production in renal proximal tubular cells, further confirming that HIF-1$\alpha$ activation promotes EV release, although additional hypoxia-responsive mechanisms may contribute to this process [108].

Differences in culturing techniques may contribute to variations in the productivity of certain HEK293 strains, particularly when comparing suspension to adherent culture methods. A recent study highlighted significant metabolic changes when an adherent HEK293 strain was adapted to suspension culture. This adaptation led to increased glucose consumption, elevated lactate secretion, and higher succinate production, all of which suggest the occurrence of pseudohypoxia in these cells [71]. Similar metabolic shifts have been observed in CHO cells, where reducing pseudohypoxia through the inhibition of pyruvate dehydrogenase kinase (using dichloroacetate) resulted in improved culture performance and enhanced antibody production [109]. Furthermore, in a separate study, knocking down HIF-1$\alpha$ in HEK293 cells led to a remarkable 35-fold increase in HIV particle production, emphasizing the critical role of HIF-1$\alpha$ in optimizing bioreactor-based production processes [82].

**Warburg effect:** first described by Otto Warburg in the 1920s, the Warburg effect is a metabolic phenomenon observed in highly proliferative mammalian cells, including cancer cells, immune cells, and stem cells, where glucose is preferentially fermented into lactate even in the presence of oxygen [110–112]. This process diverges from conventional cellular metabolism, in which glucose-derived pyruvate is oxidized via the TCA cycle and OXPHOS in mitochondria for efficient ATP production. Instead, Warburg-positive cells convert the majority of pyruvate into lactate through the action of lactate dehydrogenase (*LDH*), a process regulated by pyruvate dehydrogenase

(*PDH*) and pyruvate dehydrogenase kinases (*PDKs*). *PDKs* phosphorylate and inhibit *PDH*, preventing pyruvate from entering the mitochondria and redirecting it toward lactate production [64].

In HEK293 cells, 77% of glucose was found to be converted into lactate [113], indicating a strong Warburg effect, instead of fuelling pyruvate into the TCA, which makes up only 22%. TCA cycle carbon mostly originates from amino acid metabolism, with 16% of the TCA cycle carbon originating from glutamine [113], primarily entering as AKG [93]. When pH drops below 6.75 and lactate concentration reaches 13 mM, HEK293 cells shift to consuming both glucose and lactate, preventing lactate buildup and improving metabolic balance under pH-uncontrolled conditions. This metabolic shift enables higher growth rates and reduces acidification, making it beneficial for bioprocess optimization. In glucose-only metabolism, HEK293 cells rely heavily on glycolysis, with only 24% of pyruvate entering mitochondria. However, in glucose-lactate co-metabolism (pH-uncontrolled), glucose uptake decreases by 89%, allowing both glucose and lactate to fuel the TCA cycle, stabilizing energy metabolism, sustaining cell growth, and reducing toxic by-products like lactate [114].

Identifying the inefficient pyruvate entry into the TCA as bottelneck in HEK293 strains, studies were conducted to improve central carbon metabolism. The overexpression of pyruvate carboxylase (*PC*) leads to a more efficient conversion of pyruvate to oxaloacetate, resulting in increased glucose utilization, decreased lactate and ammonia production, as well as higher viability [115–117]. The knockdown of *PDK*, which downregulates the activity of *PC*, resulted in lower amounts of lactate secreted, and VLP production was enhanced by 30-fold [82]. In a recent approach, CHO and HEK293 cells were adapted to Warburg-zero cells, meaning that they do not produce lactate anymore, while maintaining wild-type growth rates, disproving the hypothesis that lactate production is essential for rapid proliferation. Cells exhibited a drastic reduction in glucose uptake but maintained energy balance via increased oxidative metabolism, mainly enhanced TCA activity. Oxygen consumption increased, indicating reliance on mitochondrial ATP production, and cells were more sensitive to mitochondrial inhibitors. $NADH$ levels increased significantly, disrupting the $NADH/NAD^+$ balance. Cells compensated by upregulating alternative $NAD^+$ regenerating pathways, including glycerol-3-phosphate dehydrogenase (*GPD1*) [64]. With these insights, it could be shown that cell growth is independent from the Warburg effect.

### 1.5.b ER stress and unfolded protein response (UPR)

ER stress is a cellular condition that arises when there is an accumulation of unfolded or misfolded proteins within the ER lumen. This accumulation can overwhelm the ER capacity to properly fold proteins, leading to a state of stress that triggers a protective signaling pathway known as the UPR. The UPR serves to restore normal function by enhancing the folding capacity of the ER, reducing the synthesis of new proteins, and promoting the degradation of misfolded proteins through a process called ER-associated degradation (ERAD). If the stress is unresolved, the UPR can lead to apoptosis, highlighting its dual role in cell survival and death [118]. The occurrence of ER stress and the activation of the UPR are particularly relevant in the context of producing RPs or VLPs in mammalian cells. When cells are engineered to produce large amounts of RPs, the influx of nascent polypeptides can exceed the folding capacity of the ER, resulting in an accumulation of misfolded proteins [118]. This situation is exacerbated by the complex folding requirements of many RPs, which may include PTMs that are critical for their functionality [119]. Similarly, viral infections can induce ER stress due to the unscheduled accumulation of VLPs and alterations to ER membranes, which further disrupt the normal protein folding environment [120, 121].

The UPR is mediated by three primary signaling pathways involving the transmembrane proteins *IRE1*, *ATF6*, and *PERK*, each of which plays a distinct role in the cellular response to ER stress [122, 123]. For instance, *IRE1* activates *XBP1*, a transcription factor that enhances the expression of chaperones and components of the ERAD machinery, and therefore inducing the UPR [124]. Overexpression of *XBP1* led to an improved volumetric and specific production of $\gamma$-retroviruses in HEK293 cells [125]. In terms of rAAV production, the down-regulation of nucleophosmin (*NPM1*) resulted in up to a 35-fold increase in rAAV 2 and rAAV 8 vector production, suggesting that manipulating cellular stress responses can significantly enhance rAAV yield [126].

*NPM1* has been identified as a chaperone-like protein that can bind to misfolded proteins, which is particularly relevant during ER stress when the accumulation of such proteins occurs [127].

### 1.5.c Challenges and solutions during transfection

In a transient transfection process for rAAV production, plasmids are complexed with transfection reagents such as PEI [16,25,77] to allow uptake into cells. However, numerous factors impair transfection efficiency including the correct plasmid ratio (in case of a triple transfection process) [24,52], cell density [128,129], size of pDNA-particles [130], the interaction with host macromolecules after entering the host cells [131], or the number of plasmids (one vs. three) [132].

**Cell density effect:** The cell density effect (CDE) in mammalian cell culture describes how cell growth and productivity are influenced by cell concentration [128,129,133]. As cell density increases, regulatory mechanisms such as contact inhibition, nutrient depletion, and the accumulation of waste products can limit cell proliferation and productivity [129,133]. This phenomenon is particularly important in bioprocessing and RP production, where it affects product yield and quality [128,133]. For instance, hybridoma cells producing monoclonal antibodies can be optimized at low growth rates in high cell density (HCD) bioreactors, while other recombinant mammalian cells may experience diminished production under similar conditions [134]. The CDE is a significant barrier in transient gene expression processes, where transfection efficiency and productivity decrease at HCD. Transient gene expression is most commonly used in bioprocesses for producing VLPs, and RPs, but its large-scale application is hindered by the CDE. At $> 8 \cdot 10^6$ cells/mL, no detectable VLP production was observed, indicating a severe impact of CDE [128]. One reason for the low transfection efficiency at HCD might be the increased production of EVs by cells [128]. Higher EV concentrations correlated with lower transfection efficiency. When fresh medium was replaced with spent medium from HCD cultures, transfection efficiency dropped from 80% to 5%. Interestingly, EVs seem to interact with PEI:pDNA polyplexes, causing their breakdown and aggregation.

Depleting EVs from conditioned medium restored transfection efficiency to near-optimal levels. At HCD, polyplexes were trapped in intracellular vesicles, preventing transfection. Ceramide glucosyltransferase (*UGCG*) overexpression disrupted these intracellular structures, allowing polyplexes to reach the nucleus. The transfection efficiency was improved by 45% at $12 \cdot 10^6$ cells/mL, when combining EV depletion and *UGCG* overexpression. VLP production increased by 60%, and polyplex trafficking was normalized [128]. The interaction between EVs and polyplexes during transient transfection processes may be influenced by recently published alternatives to PEI as a transfection reagent. One such alternative is FectoVir, a novel transfection reagent that forms smaller DNA-conjugates, reducing potential interactions with EVs and enhancing transfection efficiency. Additionally, Polyplus has developed a smaller helper plasmid, further improving transfection efficiency [130]. However, increasing the transfection efficiency of all plasmids during transient transfection could lead to a lower F/E ratio of VLPs. This is because the uptake of the transgene plasmid is the rate-limiting step in rAAV production in HEK293 cells [25]. Faster uptake of *rep/cap* and helper plasmid would lead to even earlier expression of the corresponding genes, and therefore production of more empty capsids, due to the lower transfection efficiency of the transgene plasmid.

Another primary challenge associated with HCD is the accumulation of metabolic by-products, such as lactate and ammonia, which can become toxic to the cells. These by-products can inhibit cell growth and viability, leading to a decrease in overall productivity. For instance, it has been shown that elevated lactate levels can negatively impact the metabolic activity of HEK293 cells, thereby reducing their capacity to produce VLPs effectively [133]. Additionally, HCD can lead to nutrient depletion, as the demand for essential nutrients increases with cell proliferation, potentially limiting the metabolic resources available for VLP assembly and secretion [135].

**AAVone system:** the **AAVone system** [132] has been developed to overcome the limitations of the standard process, where three plasmids are taken up at different rates [25]. The AAVone system

is a single-plasmid approach designed to improve rAAV vector production by consolidating all necessary genetic components into a compact 13-kb plasmid. Compared to the conventional triple plasmid method, AAVone increases rAAV yields by 2 to 4 times across multiple serotypes while maintaining high production consistency and reducing batch variability. Beyond yield improvements, AAVone enhances vector quality by reducing DNA impurities, particularly unwanted plasmid backbone sequences, and lowering the presence of truncated or non-functional genomes. It also increases the F/E capsid ratio, producing a higher proportion of fully packaged, functional vectors. With reduced plasmid DNA input requirements, the system minimizes manufacturing costs while maintaining equivalent or superior transduction efficiency *in vitro* and *in vivo*. These advantages make AAVone a scalable, cost-effective, and good manufacturing practice (GMP)-compatible solution for rAAV-based gene therapies [132].

**Stable cell lines** for rAAV production provide significant advantages over transient transfection, particularly in scalability, cost efficiency, and product consistency [62, 136–139]. Unlike transient transfection, which requires expensive GMP-grade plasmids and reagents, stable cell lines streamline manufacturing and lower costs [62]. They also enhance reproducibility and quality control, ensuring consistent batch-to-batch production with an improved F/E capsid ratio, reducing empty capsids that could trigger immune responses [139]. Additionally, stable cell lines eliminate the need for helper virus infection, simplifying downstream purification and minimizing contamination risks [139]. Inducible stable expression systems further optimize production by allowing controlled activation of rAAV replication and capsid gene expression, preventing cytotoxic effects and improving overall yields. Stable HEK293 cell lines achieve vector titers comparable to or exceeding those of transient transfection and produce higher proportions of full capsids, with up to 73% full particles, compared to the 20–50% typically seen in transient transfection [62]. These advantages make stable cell lines a promising approach for large-scale, high-quality rAAV vector manufacturing; however, the establishment of stable cell lines remains laborious and time-consuming.

## 1.6 Genome-scale metabolic modeling of biopharmaceutical processes

### 1.6.a Fundamentals of GSMMs

Genome-scale metabolic models (GSMMs) are mathematical representations that describe the biochemical reactions occurring within a cell or organism, allowing for the simulation and analysis of metabolic processes. These models are constructed using a stoichiometric matrix, which organizes the stoichiometric coefficients of metabolites involved in various biochemical reactions. The stoichiometric matrix is crucial for understanding the relationships between different metabolites and reactions, as it encapsulates the conservation of mass across the metabolic network [140, 141]. Within this framework, the null space of the stoichiometric matrix plays a significant role; it represents the set of steady-state flux distributions that satisfy mass balance constraints, indicating how metabolites can be interconverted without accumulating or depleting [142, 143].

### 1.6.b Constraint-based modeling

Constraint-based modelling (CBM) is a computational approach used to study biological systems, particularly metabolic networks. It involves the application of mathematical constraints to represent the biochemical, thermodynamic, and physiological limitations that govern metabolic reactions. Instead of explicitly modeling the kinetics of each reaction, CBM uses optimization techniques to predict cellular behavior under different environmental and genetic conditions. In CBM, different constraints are applied to define the solution space of feasible metabolic flux distributions [144, 145].

Flux balance analysis (FBA) is the most prominent computational method used to analyze the flow of metabolites through metabolic networks, particularly in the context of GSMMs. FBA operates under the assumption that the system is at a steady state, meaning that the rate of input and output of metabolites is balanced. The most critical parameters in FBA include the

stoichiometric matrix, which represents the relationships between metabolites and reactions, the objective function, which defines the goal of the analysis (e.g., maximizing biomass production), and the constraints that reflect biological limitations such as nutrient availability and reaction capacities [146,147]. Mathematically, FBA is grounded in linear programming. The core of the method involves formulating the metabolic network as a system of linear equations, where the stoichiometric matrix (S) relates the rates of reactions (v) to the concentrations of metabolites (x) through the equation

$$S \cdot v = 0. \tag{1}$$

The objective function is then maximized (or minimized) subject to constraints that define the feasible region of the solution space [146,148]. FBA has been applied in various research contexts, particularly in metabolic engineering and bioprocess optimization. For instance, studies have utilized FBA to improve the production of metabolites such as succinate and lactate in engineered strains of *E.coli*, demonstrating significant enhancements in yield through targeted gene knockouts and pathway optimizations. A 1.5-fold increase in succinate production was observed, achieving a yield of 0.78 g/L/h [149]. A 2.3-fold increase in mevalonate production through strategic gene knockouts was observed, resulting in a final concentration of 1.5 g/L [148]. In the context of CHO cells, FBA was employed to optimize media formulations and feeding strategies, leading to a 30% increase in monoclonal antibody productivity, with titers reaching up to 4 g/L [150]. Similarly, enhanced feeding strategies resulted in a 25% increase in antibody titers, achieving a final concentration of 3.5 g/L [151]. These studies highlight the effectiveness of FBA in guiding metabolic engineering efforts to enhance productivity and optimize bioprocesses.

**Flux capacity constraints** define the upper and lower bounds of reaction fluxes, ensuring that reactions operate within biologically feasible limits. These constraints are represented mathematically as

$$v_{min} < v < v_{max}, \tag{2}$$

where fluxes may be restricted based on experimental data or enzymatic capacities. For instance, irreversible reactions are constrained to have only non-negative fluxes (0), while reversible reactions can have both positive and negative values within a given range [152].

An important set of flux capacity constraints are environmental constraints. Environmental constraints define the availability and uptake/secretion rates of nutrients, such as glucose or oxygen. These constraints impose upper limits on nutrient influx ensuring that simulated metabolic behavior reflects realistic growth conditions [63,153,154].

Objective function constraints define the metabolic goal that the model aims to optimize. Common objective functions include maximizing biomass production to simulate growth or production rate [63,148,150,151,155], maximizing ATP yield to reflect energy efficiency [156], or minimizing enzyme usage to optimize resource allocation [157]. These constraints guide the computational optimization process in predicting cellular behavior under different scenarios. As recently shown, the biomass composition does not seem to not have a large impact on correct predictions, whereas accurately measured exchange rates were shown to be massively important for the correct prediction of growth rates [63,158].

**Thermodynamic constraints** Thermodynamic constraints ensure that metabolic reactions obey the principles of thermodynamics, particularly reaction directionality and feasibility. These constraints often involve the Gibbs free energy change ($\Delta G$), requiring that spontaneous reactions have $\Delta G < 0$. Thermodynamically constrained models integrate these conditions to refine flux predictions and prevent biologically implausible solutions [159,160]. Introducing thermodynamic constraints in GSMMs poses challenges due to data limitations, computational complexity, and feasibility issues. The values for $\Delta G$ might be incomplete or could be biased due to uncertain experimental data making an accurate estimation difficult. Since reaction feasibility depends on intracellular metabolite concentrations, which are difficult to measure directly, assumptions often

introduce inaccuracies [161–163]. Additionally, $\Delta G$ values depend on environmental factors like pH, which vary across cellular compartments, making general constraints unreliable if not carefully adjusted [164].

**Enzyme constraints** Enzyme constraints limit the metabolic flux based on the catalytic efficiency and availability of enzymes [165,166]. The reaction rate is constrained by the enzyme's kinetic parameters ($k_{cat}$) and its concentration ($E_i$), expressed as $v_i \leq k_{cat} \cdot E_i$ [166]. These constraints are crucial for accurately modeling metabolic reactions that depend on enzyme kinetics. Introducing enzyme constraints in GSMMs presents several challenges, primarily due to the limited availability and variability of experimental kinetic parameters ($k_{cat}$) [167–169]. Additionally, kinetic parameters are not universal; they can change significantly depending on environmental conditions such as pH, temperature, nutrient availability, PTMs, and cellular stress, making their direct application in GSMMs difficult [170,171]. To address the scarcity of kinetic parameters, artificial intelligence-based approaches have been developed to predict $k_{cat}$ values based on protein sequences, structure, and reaction context [167–169]. These models significantly improve the coverage of kinetic data and enable better-informed constraints in enzyme-constrained GSMMs. However, they do not solve the fundamental issue of environmental variability. Predicted $k_{cat}$ values are typically derived from datasets measured under specific conditions, and their applicability under different physiological or industrial settings remains uncertain [167–169]. Since enzymes can exhibit different catalytic efficiencies in response to temperature, pH, or intracellular crowding, these models still require careful validation and adaptation for specific applications.

### 1.6.c Contextualization of global GSMMs

Global reconstructions, which cover the whole set of reactions of an organism have been used extensively for modeling [63,172–174]. However, context-specific GSMMs allow for more accurate predictions [172] as their construction typically relies on the integration of context-specific transcriptomic, proteomic or metabolomic data. To generate context-specific GSMMs, various model extraction methods (MEMs) have been designed [175,176]. However, certain challenges remain, including the selection of transcriptomic thresholds [177,178] and the order of constraint introduction [178,179]. Thus, consensus pipelines remain to be established.

It has been reported that the choice of MEM had a greater influence on GSMM variability than the differences between biological conditions [179–182] with ranges of up to 100% in size of reactions in context-specific GSMMs [183]. Furthermore, thresholds set on transcriptomic datasets have been shown to play a pivotal role in reconstruction, altering biological function of GSMMs [177,180], having an even larger impact than the choice of MEMs [178,179]. Adding metabolic constraints had the least impact on model composition [179]. Aside from gene expression thresholds, it was found that including differentially expressed genes significantly increased predictability of GSMMs, independent of the chosen MEM [177].

In the following section, an overview of various MEMs (see Tab 1), sorted by family, including MBA-like, GIMME-like, iMAT-like, and MADE-like MEMs will be given.

**MBA-like methods.** The MBA-like family of GSMMs reconstruction algorithms is based on the principle of defining a core set of essential metabolic reactions and iteratively removing other reactions while maintaining network consistency. Unlike the GIMME-like family, which relies on a RMF, and the iMAT-like family, which maximizes agreement between reaction activity and gene expression levels, MBA-like methods take a different approach by focusing on network topology and consistency. This makes them particularly useful in cases where defining an RMF is challenging or where multiple types of experimental data need to be integrated [175].

**Mathematical formulation:**

$$\min \sum_{i \in \mathcal{R} \setminus \mathcal{C}} z_i \tag{3}$$

**Table 1 . Summary of MEM families.** MEM families can be seprated based on required metabolic function (RMF) requirement, type of input data, and the algorithm.

| Feature | MBA-like | GIMME-like | iMAT-like | MADE-like |
|---|---|---|---|---|
| RMF required | NO | YES | NO | NO |
| Main input data | Transcriptomics, Proteomics, Metabolomics | Transcriptomics, Proteomics, Metabolomics | Transcriptomics, Proteomics Metabolomics | Differential Gene Expression |
| Discretization | YES | NO | YES | YES |
| Optimization Problem | LP | LP | MILP | MILP |
| Examples | MBA, mCADRE, FASTCORE, SWIFTCORE, CORDA | GIMME, GIMMEp RIPTiDe GIM3E | iMAT, INIT tINIT, RegEx | MADE RMetD2 $\delta$ FBA |

**Subject to:**

- **Flux balance constraints (steady-state):**

$$Sv = 0 \tag{4}$$

- **Core reactions must remain active:**

$$v_i \neq 0, \quad \forall i \in \mathcal{C} \tag{5}$$

- **Non-core reactions can be removed:**

$$z_i = \begin{cases} 1, & \text{if } v_i = 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

**MBA.** The Model Building Algorithm (MBA) was the first method in this family and set the foundation for many later refinements. MBA works by categorizing metabolic reactions into two groups: core reactions, which are expected to be active in a specific context, and non-core reactions, which may or may not be necessary. Core reactions are typically identified using a combination of transcriptomic, proteomic, metabolomic data, or biochemical knowledge. The algorithm then removes non-core reactions while ensuring that the core reactions remain functional and that the network retains metabolic consistency. The process is performed iteratively, where each removed reaction is checked to ensure that it does not disrupt any essential pathways. If removing a reaction leads to an inconsistency, the reaction is reintroduced into the GSMM. Since the final GSMM depends on the order in which reactions are removed, MBA runs multiple iterations to generate a consensus reconstruction. This consensus approach ensures that the resulting model remains as compact as possible while still being functionally viable [184]. A significant limitation of MBA is its computational intensity, as the iterative pruning and consistency-checking process can be slow, especially for large-scale metabolic models.

**mCADRE.** To address these issues, mCADRE (Metabolic Context-specificity Assessed by Deterministic Reaction Evaluation) was developed as an improvement over MBA. Like MBA, mCADRE relies on defining a core set of reactions, but it introduces reaction scoring methods to improve efficiency. These scores are based on multiple factors, including gene expression levels, network connectivity, and experimental confidence scores. Reactions with high scores are more likely

to be retained, while those with low scores are pruned more aggressively. Additionally, mCADRE allows for the incorporation of key metabolites that are required in the given biological context, ensuring that the final model includes all pathways necessary to maintain these metabolites. The use of Fast Flux Variability Analysis (FastFVA) in mCADRE significantly improves the efficiency of consistency checking, making it faster and more scalable than MBA [185].

**FASTCORE.** To further improve computational efficiency and scalability, FASTCORE (Fast Context-Specific Reconstruction of Metabolic Networks) was developed as another alternative to MBA. FASTCORE eliminates the need for multiple iterations by formulating the pruning process as a linear problem (lp) problem. The goal of FASTCORE is to identify a minimal yet consistent set of metabolic reactions that can support all core reactions while minimizing the inclusion of unnecessary reactions. Instead of iteratively pruning and reintroducing reactions like MBA, FASTCORE solves a single optimization problem that identifies the sparsest possible network while keeping all core reactions active. This results in a more parsimonious model that is computationally efficient to construct. FASTCORE is particularly well suited for large-scale applications, as it significantly reduces computational time compared to both MBA and mCADRE [186].

**SWIFTCORE.** While FASTCORE improved efficiency, its approach to handling reversible reactions was somewhat suboptimal, as it required additional steps to determine reaction directionality. To overcome this limitation, SWIFTCORE was introduced. Instead of determining reaction directions through mixed-integer linear problem (MILP), which can be computationally expensive, SWIFTCORE uses a soft constraint approach that encourages reactions to be active in one direction without enforcing strict constraints. This results in faster execution times and a more compact final GSMM, making SWIFTCORE an attractive alternative for large-scale metabolic reconstructions [187].

**CORDA.** Another notable extension of MBA-like methods is CORDA (Cost-Optimization Reaction Dependency Assessment), which introduces a more flexible approach to reaction pruning. While most MBA-like methods aim to remove as many non-core reactions as possible, CORDA takes a less aggressive approach, allowing for the retention of some non-core reactions if they contribute to overall network feasibility. This is particularly useful in cases where experimental data is incomplete, and removing all non-core reactions might lead to an overly restrictive model. Instead of forcing a minimal set of reactions, CORDA evaluates the dependency of core reactions on non-core reactions and only removes non-core reactions if they are not needed to maintain core function. This makes CORDA particularly useful for large-scale tissue-specific reconstructions, as it produces more biologically realistic models compared to parsimonious approaches like FASTCORE [188].

**GIMME-like methods.** The GIMME-like family of MEMs is designed to integrate high-throughput transcriptomics data into global GSMMs while ensuring that the reconstructed models retain essential biological functions. These methods primarily focus on adjusting reaction flux constraints based on gene expression levels while simultaneously maintaining a predefined RMF. This RMF typically represents an essential physiological objective, such as biomass production, ATP maintenance, or a known metabolic demand specific to a given biological context [175].

**Mathematical Formulation:**

$$\min \sum_{i \in \mathcal{R}} p_i \cdot v_i \tag{7}$$

Subject to:

- **Flux balance constraints (steady-state assumption):**

$$S \cdot v = 0 \tag{8}$$

where $S$ is the stoichiometric matrix, and $v$ is the vector of reaction fluxes.

- RMF constraint: Maintain at least a fraction $\alpha$ of optimal RMF flux:

$$v_{\text{rmf}} \geq \alpha \cdot v^*_{\text{rmf}} \tag{9}$$

- Penalty function: Penalizes reactions inconsistent with experimental data:

$$p_i = \begin{cases} c_i \cdot (e_i - t), & \text{if } e_i < t \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

where $e_i$ is the gene expression value of gene $i$, $t$ is the threshold, and $c_i$ is a weighting coefficient for gene $i$.

**GIMME.** The foundational algorithm in this family is GIMME (Gene Inactivation Moderated by Metabolism and Expression), which was first introduced as a method to tailor generic metabolic models to specific cellular conditions. GIMME operates by modifying the flux distribution of a reference metabolic network based on gene expression data, ensuring that reactions associated with low-expression genes receive minimal flux while still maintaining the predefined RMF. The algorithm proceeds in two main steps: first, it solves a standard FBA problem to determine the maximal possible flux through the RMF under normal conditions. In the second step, it penalizes reactions that are linked to genes with expression levels below a given threshold while ensuring that the RMF remains above a fraction of the original maximal flux. This means that GIMME deactivates reactions with low expression evidence only if doing so does not disrupt essential metabolic functions. An important feature of GIMME is that it prioritizes metabolic feasibility over strict adherence to transcriptomics data, meaning that some lowly expressed reactions may remain active if they are necessary for achieving the RMF. This characteristic ensures that the reconstructed model remains biologically relevant and functional, avoiding the complete loss of critical pathways due to potential inaccuracies or incompleteness in gene expression data. However, one limitation of GIMME is its reliance on an arbitrary threshold for gene expression levels, which can impact the robustness and accuracy of the reconstructed model [189].

**GIM3E.** Another advancement in the GIMME-like family is GIM3E (GIMME with Metabolomics Evidence Integration), which further extends the original GIMME approach by incorporating metabolomics data. GIM3E introduces a mechanism that ensures the network uses experimentally detected metabolites. This is achieved by introducing turnover reactions for each metabolite present in the metabolomics dataset and forcing their fluxes to remain above a minimum threshold. By constraining metabolite usage within the metabolic network, GIM3E provides a more biologically realistic depiction of metabolic activity, allowing it to better capture condition-specific metabolic behaviors. However, the inclusion of metabolomics data adds additional computational complexity, as the model needs to accommodate constraints from multiple data sources simultaneously. Moreover, the method requires additional processing to ensure that metabolite measurements are correctly mapped to the corresponding reactions in the GSMM [190].

**RIPTiDe.** A more recent addition to the GIMME-like family is RIPTiDe (Reaction Inclusion by Parsimony and Transcript Distribution), which refines the process of integrating transcriptomics data into metabolic models by removing the need for an explicit gene expression threshold. Unlike traditional GIMME-based approaches that define a binary threshold to classify reactions as "on" or "off", RIPTiDe uses a continuous weighting system to determine reaction flux penalties based on transcript levels. The algorithm applies a parsimonious approach to enzyme usage, minimizing overall flux values while ensuring that metabolic functions remain intact. Specifically, RIPTiDe first identifies the optimal flux distribution for a given RMF, similar to GIMME, but then applies an iterative refinement process where it assigns a weight to each reaction based on its corresponding

gene expression level. Reactions with higher gene expression are favored, while those with lower expression are progressively penalized. This ensures that the final model reflects metabolic activities that align with transcriptomic data without relying on arbitrary cutoffs. A key advantage of RIPTiDe is that it avoids the sharp cutoff effects seen in GIMME and other threshold-dependent algorithms. Instead of forcing a strict distinction between expressed and non-expressed reactions, RIPTiDe allows for a more gradient-based selection of metabolic activities, making it more robust to fluctuations in transcriptomics data [191].

**iMAT-like methods.** The iMAT-like family of MEMs represents a class of methods that integrate high-throughput transcriptomic or proteomic data into metabolic networks without requiring a predefined RMF. Instead of optimizing for a specific metabolic function, iMAT-like methods focus on maximizing the agreement between reaction activity states (active or inactive) and the expression levels of their associated genes or proteins. This makes these algorithms particularly useful in cases where the RMF is unknown or difficult to define, such as in the reconstruction of models for specific tissues, disease states, or poorly understood biological contexts [175].

**Mathematical Formulation:**

$$\max \sum_{i \in \mathcal{R}} \left( w_i^+ \cdot y_i - w_i^- \cdot (1 - y_i) \right) \tag{11}$$

**Subject to:**

- **Flux balance constraints (steady-state):**

$$S \cdot v = 0 \tag{12}$$

- **Reaction state constraints (binary activation):**

$$M \cdot y_i \geq v_i \geq -M \cdot y_i \tag{13}$$

  where $M$ is a large constant and $y_i$ is a binary variable (1 if reaction is active, 0 otherwise).

- **Weights for consistency with expression data:**

$$w_i^+ = \text{high for } e_i \text{ in active range}, \quad w_i^- = \text{high for } e_i \text{ in inactive range} \tag{14}$$

  with $e_i$ being the gene expression of gene $i$

**iMAT.** The foundational algorithm in this family is iMAT (Integrative Metabolic Analysis Tool), which was developed to systematically integrate transcriptomic and proteomic data into metabolic networks by classifying reactions as either active or inactive based on expression data. The method first discretizes gene expression levels into categories such as high, moderate, or low expression. It then assigns corresponding states to metabolic reactions, aiming to maximize the number of reactions in the model that align with the expression data while ensuring overall metabolic feasibility. This is achieved by solving a MILP problem, where the objective is to find a flux distribution that best matches the expected activity of reactions based on gene expression levels. iMAT does not enforce a RMF but rather seeks to maintain metabolic network consistency while ensuring that the final model reflects experimentally observed gene or protein activity. One of the limitations of iMAT is its reliance on the discretization of expression data, which can lead to information loss. Additionally, because iMAT assigns binary states to reactions (active or inactive), it can sometimes produce overly simplified models that do not fully capture intermediate activity levels of metabolic reactions [192].

**INIT.** An important refinement of iMAT is INIT (Integrative Network Inference for Tissues), which was designed to incorporate qualitative metabolomic data alongside transcriptomic and proteomic data. Unlike iMAT, INIT assigns weights to metabolic reactions based on the level of experimental evidence supporting their presence in a given tissue or condition. These weights can be derived from transcriptomic or proteomic datasets, where high-confidence reactions receive positive weights, while reactions with strong evidence of absence receive negative weights. The goal of INIT is to find a metabolic flux distribution that maximizes the sum of these reaction weights, ensuring that reactions with strong experimental evidence remain active while those with weak or contradictory evidence are removed. A key innovation in INIT is that it does not assume a steady-state metabolic network, allowing for some degree of metabolite accumulation or depletion. This makes INIT particularly well suited for integrating metabolomics data, where the presence or absence of specific metabolites can be used to constrain metabolic fluxes. The method also allows the user to define a set of required metabolic tasks, ensuring that key metabolic functions remain intact during the reconstruction process. However, the increased complexity of INIT comes at a computational cost, as the MILP optimization problem it solves is more demanding than standard iMAT [193].

**RegrEx.** Another important extension of iMAT is RegrEx (Regularized Metabolic Model Extraction), which introduces a continuous rather than discrete approach to integrating gene expression data into GSMMs. Unlike iMAT, and INIT which rely on binary classification of reactions as active or inactive, RegrEx seeks to directly correlate reaction fluxes with gene expression levels using a regularized regression approach. This allows RegrEx to maintain a more nuanced representation of metabolic activity, avoiding the sharp cutoffs imposed by discretization in traditional iMAT-based methods. The key innovation in RegrEx is its use of regularization techniques to prevent overfitting and ensure that only biologically relevant reactions are retained in the final model. Specifically, RegrEx employs L1-norm and L2-norm regularization to impose sparsity on the GSMM, ensuring that only the most essential reactions are included. A major advantage of RegrEx over traditional iMAT-based methods is that it does not require arbitrary thresholds to classify gene expression levels, making it more robust to variability in transcriptomic data. However, the method is computationally intensive due to the need to solve large-scale mixed-integer quadratic problems, which are more complex than the MILP problems used in iMAT and INIT [194].

**MADE-like methods.** The MADE-like family of MEMs is fundamentally different from other reconstruction approaches, such as GIMME-like, iMAT-like, and MBA-like methods, in that it does not focus on extracting a single context-specific GSMM. Instead, MADE-like methods aim to capture metabolic changes between different conditions by incorporating differential gene expression data into the GSMM. These algorithms are particularly valuable in studying dynamic metabolic reprogramming, such as the transition from a healthy to a disease state, the response to drug treatments, or metabolic shifts occurring during cell differentiation or infection. Instead of simply identifying an active or inactive metabolic state for a given condition, MADE-like algorithms define a sequence of metabolic states and adjust reaction fluxes in accordance with observed gene expression changes [175].

**Mathematical Formulation:**

$$\max \sum_{i \in \mathcal{R}} w_i \cdot \left( v_i^{\text{condition 1}} - v_i^{\text{condition 2}} \right) \tag{15}$$

**Subject to:**

- **Flux balance constraints for both conditions:**

$$S \cdot v^{\text{condition 1}} = 0, \quad S \cdot v^{\text{condition 2}} = 0 \tag{16}$$

- **Flux change consistency with differential expression data:**

$$w_i = \begin{cases} +1, & \text{if } e_i^{\text{condition 1}} > e_i^{\text{condition 2}} \\ -1, & \text{if } e_i^{\text{condition 1}} < e_i^{\text{condition 2}} \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

- **Minimize total flux change penalty:**

$$\sum_{i \in \mathcal{R}} |v_i^{\text{condition 1}} - v_i^{\text{condition 2}}| \tag{18}$$

**MADE.** The foundational method in this family is MADE (Metabolic Adjustment by Differential Expression), which was developed to use time-series or comparative transcriptomic data to adjust reaction fluxes in a way that reflects changes in gene expression levels across different conditions. MADE operates under the assumption that metabolic fluxes should change in accordance with gene expression fold changes when comparing two or more states. The algorithm first determines which genes exhibit significant upregulation or downregulation between conditions. It then applies these gene expression changes to the metabolic model using gene-protein-reaction rule (GPR) associations, ensuring that upregulated genes correspond to increased reaction fluxes and downregulated genes lead to decreased fluxes. One limitation of MADE is its reliance on a discretization of gene expression changes, which can lead to information loss. Because the algorithm primarily focuses on identifying binary changes in reaction activity (i.e., active or inactive), it does not capture gradual shifts in metabolic fluxes that may occur in response to gene expression changes. Additionally, MADE assumes that changes in gene expression always result in corresponding metabolic flux changes, which may not always be the case due to post-transcriptional regulation, enzyme activity modulation, or compensatory metabolic mechanisms [195].

**RMetD2.** To address some of these limitations, RMetD2 (Relative Metabolic Differences version 2) was developed as an extension of MADE that introduces a more nuanced approach to integrating differential gene expression data into GSMMs. Instead of simply binarizing expression changes, RMetD2 gradually adjusts flux constraints based on expression differences between two conditions. The key idea behind RMetD2 is that metabolic changes do not always occur instantaneously but rather progressively shift over time. To account for this, the algorithm pushes flux constraints in incremental steps that reflect the magnitude of gene expression changes, rather than enforcing abrupt on/off states. RMetD2 operates through a series of iterative steps that progressively tighten or relax reaction constraints based on expression evidence while ensuring that the overall network remains metabolically feasible. This is done by evaluating the Spearman correlation between metabolic flux changes and gene expression changes, ensuring that the reconstructed model is consistent with observed transcriptomic alterations. Unlike MADE, which requires an RMF (such as biomass production) to guide the reconstruction, RMetD2 can operate without an explicit objective function, making it more adaptable to contexts where a metabolic function is not well defined [196].

**Δ FBA.** Another important contribution to the MADE-like family is Δ FBA, which builds upon the same principles as MADE and RMetD2 but takes a more systematic and constrained approach to flux adjustments. Unlike previous methods that simply push reaction constraints based on gene expression, Δ FBA formulates the problem as an optimization task, where the goal is to maximize the consistency between gene expression changes and reaction flux differences while minimizing unnecessary perturbations to the network. Δ FBA works by first identifying reactions that are significantly upregulated or downregulated between two conditions based on Min/Max GPR mapping, which assigns flux changes to reactions based on gene expression fold changes. The algorithm then formulates a MILP problem that seeks to maximize the agreement between gene

expression changes and flux adjustments while ensuring that the overall metabolic network remains in a steady state. Additionally, Δ FBA introduces an L2 norm minimization step, which ensures that the overall flux differences between conditions are as small as possible while still capturing biologically relevant changes. A major advantage of Δ FBA over MADE and RMetD2 is its robustness to noisy gene expression data. Because the method seeks a minimal yet consistent solution, it avoids overfitting to small fluctuations in gene expression that may not have functional significance. Additionally, Δ FBA does not require prior knowledge of metabolic tasks or an RMF, making it more flexible for studying dynamic metabolic changes in poorly characterized systems [197].

### 1.6.d Evaluation of large sets of reconstructed GSMMs

Since the generation of large amounts of GSMMs is possible through automized or semiautomized techniques, such as MEMs, it is necessary to compare GSMMs at scale, and identify the underlying differences efficiently. t-distributed stochastic neighborhood embedding (t-SNE) and Jaccard similarity index are commonly used techniques for analyzing binary reaction matrices of GSMMs [198–200]. Each method offers unique advantages and limitations in terms of clustering and similarity measurement. t-SNE is particularly useful for clustering GSMMs in a lower-dimensional space based on the presence or absence of reactions. This technique helps visualize relationships between different GSMMs by mapping similar GSMMs closer together. To achieve this, t-SNE calculates pairwise similarity using the Hamming distance, which measures the number of differing elements between two binary vectors. The non-linear transformation performed by t-SNE effectively separates groups of GSMMs while maintaining local similarity. However, t-SNE has several drawbacks: it is non-deterministic, meaning that different runs may produce slightly different clusters, making reproducibility a challenge. Furthermore, the results are sensitive to hyperparameters such as perplexity and learning rate, requiring careful tuning to achieve meaningful clustering [201]. Jaccard similarity index provides a simple and interpretable metric for comparing GSMMs by quantifying the proportion of shared reactions between two models. The Jaccard similarity index is computed as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{19}$$

where A and B are the sets of reactions in two GSMMs, and the numerator represents the number of reactions present in both models while the denominator represents the total number of unique reactions across the two GSMMs. This method is particularly useful for hierarchical clustering, allowing GSMMs to be grouped based on their overall metabolic similarity. However, Jaccard similarity index has limitations in that it does not provide reactiob-level insights. While it effectively quantifies overall similarity, it does not highlight specific reactions or metabolic subsystems that drive differences between GSMMs [202].

### 1.6.e Assessment of the solution space

**Flux variability analysis** is a computational technique used to assess the flexibility and robustness of metabolic networks. It builds upon the principles of FBA by exploring the range of possible flux distributions that can occur within a metabolic network while still satisfying the constraints. The results from flux variability analysis (FVA) include the maximum and minimum flux values for each reaction in the network, which are derived from the optimization of an objective function, such as biomass production or metabolite yield [203, 204]. This is achieved by solving a series of linear programming problems: first, to maximize the objective function and then to minimize it, thereby establishing the upper and lower bounds of each reaction's flux [203, 204].

Currently, several tools are available for performing FVA. The fastFVA algorithm is noted for its computational efficiency, allowing for rapid calculations of flux variability across large-scale GSMMs [205]. However, while fastFVA excels in speed, it may sacrifice some accuracy in the

representation of complex metabolic interactions. On the other hand, traditional FBA tools provide a more comprehensive analysis but can be computationally intensive, especially for large GSMMs [203]. FVA revealed that growth variability is the primary driver of metabolic flexibility in *E. coli*, while internal and external variability had minimal impact. In anaerobic conditions, flux variability dropped to 20% of its aerobic value, and nitrogen limitation increased external/internal contributions to 50%. Most importantly, reducing growth to suboptimal levels increased metabolic flexibility by up to 5-fold, highlighting a fundamental trade-off between growth rate and adaptability [206]. In another approach an adapted version of FVA significantly improved biomass prediction accuracy by reducing the overestimation of growth rate from 0.052 (FBA) to 0.017, aligning more closely with the typical experimental value of HEK293 cells (0.03) [207].

**Flux sampling** is a computational approach designed to explore the range of feasible metabolic flux distributions within GSMMs. Unlike FBA, which focuses on identifying a single optimal flux distribution based on predefined objective functions, flux sampling provides a probabilistic representation of all feasible steady-state flux solutions without assuming a cellular objective [208]. This unbiased approach is particularly advantageous when biological objectives are unknown or when studying metabolism under varied or non-optimal environmental conditions. Important parameters for flux sampling include the thinning factor, which controls the interval at which samples are collected to reduce autocorrelation, and the choice of the sampling algorithm, such as Coordinate Hit-and-Run with Rounding (CHRR) or OptGpSampler. Additional parameters include the number of warmup steps for achieving convergence and constraints derived from experimental data, which influence the solution space [209–212].

Various flux sampling tools are currently available, each with distinct advantages and limitations. CHRR is renowned for its efficiency and uniformity in sampling convex spaces but struggles with high computational demands for large models [212]. OptGpSampler introduces improvements in convergence speed and computational efficiency, outperforming older methods such as gpSampler [209]. GAPSPLIT excels in handling non-convex spaces, offering better coverage but at the cost of increased computational complexity [211]. `LooplessFluxSampler` addresses the issue of thermodynamically infeasible loops, ensuring realistic biological predictions [213].

Particularly, two tools have been published to analyze flux sampling datasets: CSFA [214], and FluxGAT [215]. Comparative Flux Sampling Analysis (CSFA) is a methodological framework designed to compare the metabolic flux distributions of different phenotypes. It uses flux sampling to evaluate the entire solution space of GSMMs under various constraints. CSFA identifies reactions with altered flux distributions across conditions or phenotypes through statistical analyses, making it a powerful tool for guiding metabolic engineering strategies. For instance, CSFA has been applied to identify genetic targets for strain optimization, such as improving lipid production in *Cutaneotrichosporon oleaginosus* and naringenin synthesis in *Saccharomyces cerevisiae*. By pinpointing reactions for upregulation, downregulation, or deletion, CSFA enables a systematic and stepwise approach to strain design [214]. FluxGAT integrates flux sampling with graph neural networks (GNNs) to predict gene essentiality in GSMMs without the need for predefined objective functions. This integration addresses a key limitation of traditional constraint-based modeling methods, which require assumptions about cellular objectives and introduce observer bias. FluxGAT constructs a graphical representation of flux sampling data, where nodes represent reactions and edges capture their interactions within the network. The method achieves higher sensitivity compared to traditional approaches, such as FBA, nearly doubling the predictive accuracy for gene essentiality. A key strength of FluxGAT lies in its ability to generalize across biological systems and conditions, making it suitable for applications where cellular objectives are poorly understood [215].

# 2 Aims of the Thesis

The aim of this thesis is to establish a genome-scale metabolic model (GSMM) for HEK293 strains, used for recombinant adeno-associated virus (rAAV) production under good manufacturing practice (GMP) conditions. To do so we investigate differences in cell composition and exchange rates among HEK293 strains during rAAV production through experimental measurements and develop time-resolved GSMMs based on these data. This work is composed of two parts:

**Optimizing rAAV production using HEK293-specific GSMMs**

1. What are the metabolic differences between HEK293 strains during rAAV production and mock states?

2. How can a multi-omics approach, including lipidomic, transcriptomic, and exometabolomic data, integrated in GSMMs be used to capture strain-specific metabolic shifts during rAAV production?

3. Can the identified target be experimentally validated to enhance rAAV production in the low-producing (LP) HEK293 strain?

**Implementation of logistic principal component analysis (LPCA) as a method to compare and analyze GSMMs**

The development of novel tools for large-scale comparison of GSMMs is essential to simultaneously distinguish GSMMs and elucidate the underlying metabolic drivers responsible for their separation. Current methodologies predominantly emphasize GSMM differentiation without systematically identifying key contributing factors or necessitate prior assumptions and simulations to infer them.

1. Can GSMMs be clustered more efficiently while identifying reaction-specific differences that drive categorization, without prior simulation using LPCA?

2. How does LPCA perform compared to state-of-the-art algorithms like t-distributed stochastic neighborhood embedding (t-SNE) or Jaccard similarity index, when applied to pan-GSMMs from bacteria, fungi, and human tissue-specific models? Is it possible to preserve phylogenetic relationships and differentiating tissue-specific metabolic profiles.

In addition, we authored a review article that examines the current state-of-the-art viral vector platforms used in gene therapy (GT) production, highlighting their unique advantages and limitations. We suggest the integration of GSMMs across these platforms, advocating for the development of GSMMs to fill the current gap, as only small-scale models are presently available.

# 3 Summary of Results

My contributions to our published papers are summarized in tableTab 2. Though the main focus was on experiments, analysis, and writing, I got involved in other aspects as well.

**Table 2 .** Contributor roles across different papers.
**Paper 1:** Optimizing VLP production in gene therapy: opportunities and challenges for in silico modeling.
**Paper 2:** Logistic PCA explains differences between genome-scale metabolic models in terms of metabolic pathways.
**Paper 3:** Improving HEK293-based rAAV production using GSMMs, and a multi omics approach.

| Contributor roles | Paper 1 | Paper 2 | Paper 3 |
|---|---|---|---|
| Conceptualization | ■ | ■ | ■ |
| Data curation | | ■ | ■ |
| Investigation | ■ | ■ | ■ |
| Methodology | ■ | ■ | ■ |
| Software | | ■ | ■ |
| Validation | | ■ | ■ |
| Visualization | ■ | ■ | ■ |
| Writing – original draft | ■ | ■ | ■ |
| Writing – review & editing | ■ | ■ | ■ |

## 3.1 Optimizing virus-like particle (VLP) production in gene therapy (GT): Opportunities and challenges for *in silico* modeling

In this review, we explore the current challenges and opportunities in optimizing VLP production for GT through the application of genome-scale metabolic models (GSMMs). GT has emerged as a promising field for treating inherited genetic disorders and cancer, with VLPs serving as vectors for the safe and effective delivery of therapeutic genes. However, the production of these vectors remains hindered by low yields, high costs, and complex bioprocessing requirements. Our objective is to highlight the potential of GSMMs in overcoming these challenges by providing a systematic, computational approach to bioprocess optimization. Despite significant advancements in cell line engineering and culture media optimization, traditional approaches to improving VLP production are largely reliant on empirical, time-consuming trial-and-error methodologies. In contrast, GSMMs offer a data-driven, predictive framework that enables a deeper understanding of cellular metabolism, facilitating the rational design of more efficient production strategies.

A major barrier to applying GSMMs in VLP production is the dynamic metabolic reprogramming that occurs in host cells during viral vector synthesis. Viral infections and vector production significantly alter cellular metabolism, often in ways that are not yet fully understood. While GSMMs have been extensively used to enhance cell growth and recombinant protein (RP) yields, their application to VLP production remains unexplored. To address this gap, we advocate for the development of cell line-specific GSMMs, particularly for HEK293, Per.C6, MDCK, and Vero cells, which are among the most commonly used platforms for GT vector production. Constructing such models would allow for the identification of key metabolic bottlenecks and the implementation of targeted interventions to enhance productivity. We further emphasize the necessity of integrating multi-omics data—genomics, transcriptomics, proteomics, and metabolomics—into GSMMs to improve predictive accuracy and model robustness. While extensive omics datasets exist for Chinese hamster ovary (CHO) cells, such data for GT-relevant cell lines remain limited. We propose that systematic omics studies, in combination with GSMM-based simulations, could provide novel insights into cellular metabolism, enabling the rational design of optimized culture conditions and genetic engineering strategies. Another key aspect of our review is the potential for computational

strain design (CSD) tools to refine VLP production. These tools have been successfully applied in microbial systems to enhance metabolic flux toward desired products, yet their use in mammalian cell lines remains largely unexplored. We argue that implementing CSD techniques in mammalian production hosts could lead to substantial improvements in yield and cost-effectiveness.

In conclusion, our review underscores the untapped potential of GSMMs in optimizing VLP production for GT applications. While significant challenges remain–such as the complexity of host cell metabolism, limited omics data availability, and the need for GSMM refinement–we believe that the strategic application of GSMMs, in conjunction with emerging computational and biotechnological tools, can drive substantial improvements in production efficiency. Moving forward, we advocate for the development of cell line-specific GSMMs and a greater emphasis on integrating omics-driven modeling approaches to bridge the gap between current production limitations and the future demands of GT manufacturing.

## 3.2 Logistic principal component analysis (LPCA) explains differences between genome-scale metabolic models in terms of metabolic pathways

Benchmarking LPCA is crucial for effectively comparing GSMMs, as it provides an orthogonal method to validate metabolic model reconstructions. LPCA allows improved and interpretable clustering of GSMMs while identifying distinct metabolic pathways that differentiate GSMMs, thus enhancing confidence in GSMM analyses prior to simulation.

We employed LPCA to analyze binary matrices that indicate the presence or absence of reactions across multiple GSMMs, grouping reactions by subsystems to observe metabolic differences. LPCA was applied to diverse datasets, including strains of *Escherichia*, human tissues, fungi, and *Firmicutes*. For comparative analysis, t-distributed stochastic neighborhood embedding (t-SNE) and Jaccard similarity index were also calculated. Phylogenetic relationships, derived from genomic data, were verified by comparing LPCA, thus validating LPCA's accuracy in reflecting biological relationships.

For *Escherichia* strains, LPCA identified three primary subsystems–"Murein Biosynthesis", "Murein Recycling", and "Anaplerotic Reactions"–as central to strain clustering. The LPCA method outperformed t-SNE by providing stable clustering across multiple permutations and offering insights into the reactions that drive metabolic differences, such as the reaction D-alanyl-D-alanine dipeptidase in the "Murein Recycling" pathway, absent in specific clades like *E. albertii* and *E. fergusonii*.

For fungi, LPCA successfully grouped phylogenetically similar strains, with the most prominent pathways influencing clusters being "Glycerolipid Metabolism" and "Fructose and Mannose Metabolism". Compared to the *Escherichia* pan-GSMMs, fungal models exhibited less variation across the first two principal components, possibly due to a higher proportion of unassigned reactions in fungal GSMMs.

When applied to human tissue samples, LPCA distinguished between healthy and cancerous tissues based on reaction-specific differences, with "Linoleate Metabolism" being a particularly strong marker of cancerous tissues. This subsystem, which includes reactions involved in prostaglandin synthesis (notably *PTGS2*), has a known role in cancer progression and was detected in cancerous models, whereas healthy tissue samples had higher loading values for reactions within the "Ascorbate and Aldarate Metabolism" subsystem.

For Firmicutes, LPCA clustered species into subgroups based on distinct metabolic characteristics, aligning closely with the main phylogenetic clades. However, incomplete subsystem annotations limited the utility of LPCA for detailed pathway analysis within this group. LPCA showed improved clade discrimination and allowed more precise identification of metabolic subsystems driving these separations, while computation time was significantly higher compared to t-SNE and Jaccard similarity index.

### 3.3 Improving HEK293-based recombinant adeno-associated virus (rAAV)-production using GSMMs, and a multi-omics approach

The HEK293 strains were initially cultured in chemically defined, serum-free FreeStyle F17 medium. Cell expansion was conducted in spinner flasks, followed by transfer to a modular fermentation system for transient rAAV production. To standardize comparisons, cells were divided into two sets for either mock (control) or plasmid transfection, involving a triple-plasmid transfection using polyethylenimine. Samples were collected at five time points post-transfection (0, 4, 24, 48, and 72 hours post transfection (HPT)) for analysis, covering a total of 80 samples. Experimental analyses encompassed multiple cell and media assessments. Biomass was measured using centrifugation and PBS washing, followed by drying to ensure consistent mass measurements. Biomass composition, including DNA, RNA, glucose, and lipids, was quantified using already established methods [63]. Metabolite concentrations in the spent media were analyzed using LC-MS, with further quantification of amino acids and metabolic components like lactate and ammonia. rAAV capsid production and genome quantification were conducted using ELISA and droplet digital PCR, respectively. For computational analysis, lipidomic data were normalized and analyzed with principal component analysis (PCA) to identify patterns and drivers of metabolic differences. Growth rates were calculated based on biomass changes and fitted using a logistic function. Metabolic exchange rates were derived from concentration profiles, enabling the development of context-specific GSMMs using the CORDA [188] algorithm and transcriptomic data [77]. Simulated growth rates were compared with experimental growth rates, confirming that the models accurately captured growth trends across conditions. As an orthogonal method, LPCA was used to capture a similar clustering pattern based on reconstructed GSMMs compared to transcriptome based PCA.

The comparative analysis between high-producing (HP) and low-producing (LP) HEK293 strains revealed substantial differences in their biomass composition, energy storage strategies, and metabolic efficiencies, all of which impacted rAAV production. The HP strain maintained a more stable dry cell mass over time with a higher protein content, while the LP strain displayed a higher but declining dry mass over the fermentation period. This discrepancy pointed to underlying metabolic inefficiencies in the LP strain. Lipid analysis provided further insights into these metabolic variations: Phosphatidylcholine (PC) and phosphatidylethanolamine (PE) represented the dominant lipids across both strains, consistent with their roles in cell membrane structure. However, the HP strain accumulated higher levels of energy-rich triacylglycerides (TG) and ceramides (Cer) as fermentation progressed, while the LP strain stored glucose as glycogen rather than TG. This shift in the LP strain suggested a metabolic bottleneck possibly associated with oxygen limitation, as glycogen is less efficient than TG for long-term energy storage. PCA of lipid profiles clearly distinguished between HP and LP strains, primarily driven by higher TG and Cer levels in HP, underscoring its superior energy storage capability. The GSMMs developed for each strain illustrated a significant allocation of cellular resources toward growth rather than rAAV production, accounting for over 99.8% of metabolic activity in both strains. This observation guided the strategic decision to suppress growth and redirect resources to enhance rAAV productivity. Simulation results showed that these GSMMs accurately captured growth trends and metabolic fluxes across conditions, with growth rate predictions within a reasonable margin of experimental data. The theoretical production potential calculated for both strains was 1000-fold higher than actual rates. This gap suggested that additional metabolic and environmental optimizations could be implemented to maximize rAAV yield.

A crucial finding of this study was the identification of HIF-1$\alpha$ as a metabolic bottleneck in the LP strain. HIF-1$\alpha$ was identified based on time-resolved flux sampling, and linked to a pseudohypoxic state in the LP cells, shifting resources toward growth rather than rAAV production. Recognizing HIF-1$\alpha$ as a limiting factor, we tested the effects of inhibiting this transcription factor on rAAV productivity. The inhibition of HIF-1$\alpha$ led to a remarkable outcome: the specific productivity of rAAV capsids in the LP strain increased by 250%. Additionally, this intervention completely halted cellular growth in the bioreactor, reallocating metabolic resources toward viral capsid production. While this inhibition improved capsid output, the full-to-empty ratio of rAAV

decreased. This decrease stemmed from inhibited nucleotide synthesis within the pentose phosphate pathway (PPP), which impaired genome replication capabilities. This trade-off highlighted the importance of balancing capsid production with genome replication efficiency when optimizing rAAV production in HEK293 strains.

# 4 Conclusion and Outlook

## 4.1 Improving HEK293-based AAV-production using GSMMs, and a multi-omics approach

In the main part of this thesis, we explored the metabolic and transcriptomic differences in HEK293 strains used for recombinant adeno-associated virus (rAAV) production, applying genome-scale metabolic models (GSMMs) and multi-omics data integration to optimize production efficiency. One of the most significant findings of this study is that the primary metabolic distinction in HEK293 cells lies between the high-producing (HP) and low-producing (LP) HEK293 strains, rather than between the production states (mock vs. transfected). Lipidome, transcriptome, reactome, and fluxome analyses all reinforce this observation, indicating that strain-specific metabolic characteristics dictate production outcomes more than the presence of viral vector transfection. Furthermore, GSMM simulations revealed that the theoretical maximum production potential of both HEK293 strains is approximately 1000-fold higher than the observed experimental values. This discrepancy suggests that metabolic constraints prevent cells from achieving their full capacity, highlighting opportunities for strain improvement through metabolic engineering.

One major limiting factor identified in bioreactor conditions is the induction of pseudohypoxia, predominantly due to the activation of HIF-1$\alpha$. This metabolic rewiring significantly contributes to the reduced production rates in the LP strain. By inhibiting HIF-1$\alpha$, cellular growth was immediately halted, while rAAV-specific productivity increased by 2.5-fold, demonstrating the potential for targeted metabolic interventions to enhance production efficiency. Interestingly, while HIF-1$\alpha$ inhibition improved specific capsid productivity, the number of viral genomes was decreased, potentially due to downregulation of the pentose phosphate pathway (PPP). This effect is likely due to the regulatory role of HIF-1$\alpha$ in PPP activity, which in turn impacts nucleotide biosynthesis. Understanding these interactions is crucial for further optimization of HEK293-based rAAV production. Despite these insights, further studies are necessary to refine the metabolic engineering strategies required for strain optimization. Future work should focus on balancing growth and productivity, overcoming metabolic bottlenecks, and integrating computational models with experimental validation to enhance bioprocess performance.

### 4.1.a Logistic PCA explains differences between genome-scale metabolic models in terms of metabolic pathways

We introduced logistic principal component analysis (LPCA) as a novel approach for clustering GSMMs, enabling efficient identification of reaction-specific differences without relying on prior simulations or environmental conditions. LPCA demonstrated its effectiveness across diverse datasets, including *Escherichia* species, *Saccharomycotina* (yeast), and human tissue-specific GSMMs, by preserving phylogenetic relationships and revealing tissue-specific metabolic profiles. Unlike traditional clustering methods, LPCA provides mechanistic insights into the drivers of clustering by pinpointing the metabolic reactions and subsystems that contribute to the differentiation of GSMMs. This ability to directly analyze binary matrices of reactions offers a more objective and interpretable method for studying metabolic networks. Despite its advantages, LPCA required a significantly higher run time than alternative algorithms, such as t-distributed stochastic neighborhood embedding (t-SNE) or the Jaccard similarity index, highlighting an area for future work to improve its computational efficiency.

# References


1. Bulcha JT, Wang Y, Ma H, Tai PW, Gao G. Viral vector platforms within the gene therapy landscape. Signal transduction and targeted therapy. 2021;6(1):53.

2. Volodina O, Smirnikhina S. The future of gene therapy: a review of in vivo and ex vivo delivery methods for genome editing-based therapies. Molecular Biotechnology. 2025;67(2):425–437.

3. Wang JH, Gessler DJ, Zhan W, Gallagher TL, Gao G. Adeno-associated virus as a delivery vector for gene therapy of human diseases. Signal Transduction and Targeted Therapy. 2024;9(1):78.

4. CMI2475 RC. Cell & Gene Therapy Market. Market Insights. 2023;.

5. GVR-4-68040-037-5 RI. Adeno Associated Virus Vector Manufacturing Market Size, Share & Trends Analysis Report. Market Analysis Report. 2023;.

6. Atchison RW, Casto BC, Hammon WM. Adenovirus-associated defective virus particles. Science. 1965;149(3685):754–756.

7. Hoggan MD, Blacklow NR, Rowe W. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. Proceedings of the National Academy of Sciences. 1966;55(6):1467–1474.

8. Samulski RJ, Muzyczka N. AAV-mediated gene therapy for research and therapeutic purposes. Annual review of virology. 2014;1(1):427–451.

9. Issa SS, Shaimardanova AA, Solovyeva VV, Rizvanov AA. Various AAV serotypes and their applications in gene therapy: an overview. Cells. 2023;12(5):785.

10. Zincarelli C, Soltys S, Rengo G, Rabinowitz JE. Analysis of AAV serotypes 1–9 mediated gene expression and tropism in mice after systemic injection. Molecular therapy. 2008;16(6):1073–1080.

11. Arjomandnejad M, Dasgupta I, Flotte TR, Keeler AM. Immunogenicity of recombinant adeno-associated virus (AAV) vectors for gene transfer. BioDrugs. 2023;37(3):311–329.

12. Chatterjee S, Li W, Wong CA, Fisher-Adams G, Lu D, Guha M, et al. Transduction of primitive human marrow and cord blood-derived hematopoietic progenitor cells with adeno-associated virus vectors. Blood, The Journal of the American Society of Hematology. 1999;93(6):1882–1894.

13. Wu P, Phillips MI, Bui J, Terwilliger EF. Adeno-associated virus vector-mediated transgene integration into neurons and other nondividing cell targets. Journal of virology. 1998;72(7):5919–5926.

14. Greig JA, Martins KM, Breton C, Lamontagne RJ, Zhu Y, He Z, et al. Integrated vector genomes may contribute to long-term expression in primate liver after AAV administration. Nature Biotechnology. 2024;42(8):1232–1242.

15. Nguyen GN, Everett JK, Kafle S, Roche AM, Raymond HE, Leiby J, et al. A long-term study of AAV gene therapy in dogs with hemophilia A identifies clonal expansions of transduced liver cells. Nature biotechnology. 2021;39(1):47–55.

16. Grieger JC, Samulski RJ. Adeno-associated virus vectorology, manufacturing, and clinical applications. Methods in enzymology. 2012;507:229–254.

17. Qu G, McClelland A, Wright J. Scaling-up production of recombinant AAV vectors for clinical applications. Current Opinion in Drug Discovery & Development. 2000;3(6):750–755.

18. Cao L, Ledeboer A, Pan Y, Lu Y, Meyer K. Clinical enrollment assay to detect preexisting neutralizing antibodies to AAV6 with demonstrated transgene expression in gene therapy trials. Gene Therapy. 2023;30(1):150–159.

19. Louis Jeune V, Joergensen JA, Hajjar RJ, Weber T. Pre-existing anti–adeno-associated virus antibodies as a challenge in AAV gene therapy. Human gene therapy methods. 2013;24(2):59–67.

20. Li X, La Salvia S, Liang Y, Adamiak M, Kohlbrenner E, Jeong D, et al. Extracellular Vesicle–Encapsulated Adeno-Associated Viruses for Therapeutic Gene Delivery to the Heart. Circulation. 2023;148(5):405–425.

21. Kreppel F, Hagedorn C. Capsid and genome modification strategies to reduce the immunogenicity of adenoviral vectors. International Journal of Molecular Sciences. 2021;22(5):2417.

22. Modlich U, Baum C, et al. Preventing and exploiting the oncogenic potential of integrating gene vectors. The Journal of clinical investigation. 2009;119(4):755–758.

23. Herrera-Barrera M, Ryals RC, Gautam M, Jozic A, Landry M, Korzun T, et al. Peptide-guided lipid nanoparticles deliver mRNA to the neural retina of rodents and nonhuman primates. Science advances. 2023;9(2):eadd4623.

24. Grieger JC, Soltys SM, Samulski RJ. Production of recombinant adeno-associated virus vectors using suspension HEK293 cells and continuous harvest of vector from the culture media for GMP FIX and FLT1 clinical vector. Molecular Therapy. 2016;24(2):287–297.

25. Nguyen TN, Sha S, Hong MS, Maloney AJ, Barone PW, Neufeld C, et al. Mechanistic model for production of recombinant adeno-associated virus via triple transfection of HEK293 cells. Molecular Therapy Methods & Clinical Development. 2021;21:642–655.

26. Weger S, Wendland M, Kleinschmidt JA, Heilbronn R. The adeno-associated virus type 2 regulatory proteins rep78 and rep68 interact with the transcriptional coactivator PC4. Journal of virology. 1999;73(1):260–269.

27. Bennett A, Mietzsch M, Agbandje-McKenna M. Understanding capsid assembly and genome packaging for adeno-associated viruses. Future virology. 2017;12(6):283–297.

28. Earley LF, Powers JM, Adachi K, Baumgart JT, Meyer NL, Xie Q, et al. Adeno-associated virus (AAV) assembly-activating protein is not an essential requirement for capsid assembly of AAV serotypes 4, 5, and 11. Journal of virology. 2017;91(3):10–1128.

29. Meier AF, Fraefel C, Seyffert M. The interplay between adeno-associated virus and its helper viruses. Viruses. 2020;12(6):662.

30. Alazard-Dany N, Nicolas A, Ploquin A, Strasser R, Greco A, Epstein AL, et al. Definition of herpes simplex virus type 1 helper activities for adeno-associated virus early replication events. PLoS Pathogens. 2009;5(3):e1000340.

31. Nash K, Chen W, Muzyczka N. Complete in vitro reconstitution of adeno-associated virus DNA replication requires the minichromosome maintenance complex proteins. Journal of virology. 2008;82(3):1458–1464.

32. Mary B, Maurya S, Arumugam S, Kumar V, Jayandharan GR. Post-translational modifications in capsid proteins of recombinant adeno-associated virus (AAV) 1-rh10 serotypes. The FEBS journal. 2019;286(24):4964–4981.

33. Kontogiannis T, Braybrook J, McElroy C, Foy C, Whale AS, Quaglia M, et al. Characterization of AAV vectors: A review of analytical techniques and critical quality attributes. Molecular Therapy Methods & Clinical Development. 2024;32(3).

34. Xie Y, Butler M. N-glycomic profiling of capsid proteins from Adeno-Associated Virus serotypes. Glycobiology. 2024;34(1):cwad074.

35. Yamaguchi Y, Ishii K, Koizumi S, Sakaue H, Maruno T, Fukuhara M, et al. Glycosylation of recombinant adeno-associated virus serotype 6. Molecular Therapy Methods & Clinical Development. 2024;32(2).

36. Aloor A, Zhang J, Gashash EA, Parameswaran A, Chrzanowski M, Ma C, et al. Site-specific N-glycosylation on the AAV8 capsid protein. Viruses. 2018;10(11):644.

37. Meyer NL, Chapman MS. Adeno-associated virus (AAV) cell entry: structural insights. Trends in microbiology. 2022;30(5):432–451.

38. Mingozzi F, Anguela XM, Pavani G, Chen Y, Davidson RJ, Hui DJ, et al. Overcoming preexisting humoral immunity to AAV using capsid decoys. Science translational medicine. 2013;5(194):194ra92–194ra92.

39. McColl-Carboni A, Dollive S, Laughlin S, Lushi R, MacArthur M, Zhou S, et al. Analytical characterization of full, intermediate, and empty AAV capsids. Gene Therapy. 2024; p. 1–10.

40. Heldt CL, Skinner MA, Anand GS. Structural Changes Likely Cause Chemical Differences between Empty and Full AAV Capsids. Biomedicines. 2024;12(9):2128.

41. Grosse S, Penaud-Budloo M, Herrmann AK, Börner K, Fakhiri J, Laketa V, et al. Relevance of assembly-activating protein for adeno-associated virus vector production and capsid protein stability in mammalian and insect cells. Journal of virology. 2017;91(20):10–1128.

42. Kondratov O, Zolotukhin S. Exploring the Comprehensive Kozak Sequence Landscape for AAV Production in Sf9 System. Viruses. 2023;15(10):1983.

43. Rumachik NG, Malaker SA, Poweleit N, Maynard LH, Adams CM, Leib RD, et al. Methods matter: standard production platforms for recombinant AAV produce chemically and functionally distinct vectors. Molecular therapy Methods & clinical development. 2020;18:98–118.

44. Kurasawa JH, Park A, Sowers CR, Halpin RA, Tovchigrechko A, Dobson CL, et al. Chemically defined, high-density insect cell-based expression system for scalable AAV vector production. Molecular Therapy-Methods & Clinical Development. 2020;19:330–340.

45. Reed SE, Staley EM, Mayginnes JP, Pintel DJ, Tullis GE. Transfection of mammalian cells using linear polyethylenimine is a simple and effective means of producing recombinant adeno-associated virus vectors. Journal of virological methods. 2006;138(1-2):85–98.

46. Gao Gp, Lu F, Sanmiguel JC, Tran PT, Abbas Z, Lynd KS, et al. Rep/Cap gene amplification and high-yield production of AAV in an A549 cell line expressing Rep/Cap. Molecular Therapy. 2002;5(5):644–649.

47. Nagy A, Chakrabarti L, Kurasawa J, Mulagapati SHR, Devine P, Therres J, et al. Engineered CHO cells as a novel AAV production platform for gene therapy delivery. Scientific Reports. 2023;13(1):19210.

48. Galli A, Della Latta V, Bologna C, Pucciarelli D, Cipriani F, Backovic A, et al. Strategies to optimize capsid protein expression and single-stranded DNA formation of adeno-associated virus in Saccharomyces cerevisiae. Journal of applied microbiology. 2017;123(2):414–428.

49. Braun S, Cardon F, Dickx Y, Guillet M, Martin S, Randoux M. Production of recombinant viral vectors from plant hairy roots; 2023.

50. Fu Q, Polanco A, Lee YS, Yoon S. Critical challenges and advances in recombinant adeno-associated virus (rAAV) biomanufacturing. Biotechnology and Bioengineering. 2023;120(9):2601–2621.

51. Blessing D, Vachey G, Pythoud C, Rey M, Padrun V, Wurm FM, et al. Scalable production of AAV vectors in orbitally shaken HEK293 cells. Molecular Therapy Methods & Clinical Development. 2019;13:14–26.

52. Park S, Shin S, Lee H, Jang JH, Lee GM. Enhancing the production of adeno-associated virus (AAV) 2 and AAV9 with high full capsid ratio in HEK293 cells through design-of-experiment optimization of triple plasmid ratio. Biotechnology Journal. 2024;19(3):2300667.

53. Zhao H, Lee KJ, Daris M, Lin Y, Wolfe T, Sheng J, et al. Creation of a high-yield AAV vector production platform in suspension cells using a design-of-experiment approach. Molecular Therapy Methods & Clinical Development. 2020;18:312–320.

54. Sokolenko S, George S, Wagner A, Tuladhar A, Andrich JM, Aucoin MG. Co-expression vs. co-infection using baculovirus expression vectors in insect cell culture: Benefits and drawbacks. Biotechnology advances. 2012;30(3):766–781.

55. Wu Y, Jiang L, Geng H, Yang T, Han Z, He X, et al. A recombinant baculovirus efficiently generates recombinant adeno-associated virus vectors in cultured insect cells and larvae. Molecular therapy Methods & clinical development. 2018;10:38–47.

56. Kaiser SC, Decaria PN, Seidel S, Eibl D. Scaling-up of an Insect Cell-based Virus Production Process in a Novel Single-use Bioreactor with Flexible Agitation. Chemie Ingenieur Technik. 2022;94(12):1950–1961.

57. Joshi PR, Venereo-Sanchez A, Chahal PS, Kamen AA. Advancements in molecular design and bioprocessing of recombinant adeno-associated virus gene delivery vectors using the insect-cell baculovirus expression platform. Biotechnology Journal. 2021;16(4):2000021.

58. Monteiro F, Bernal V, Saelens X, Lozano AB, Bernal C, Sevilla A, et al. Metabolic profiling of insect cell lines: Unveiling cell line determinants behind system's productivity. Biotechnology and bioengineering. 2014;111(4):816–828.

59. Mietzsch M, Grasse S, Zurawski C, Weger S, Bennett A, Agbandje-McKenna M, et al. OneBac: platform for scalable and high-titer production of adeno-associated virus serotype 1–12 vectors for gene therapy. Human gene therapy. 2014;25(3):212–222.

60. Galibert L, Jacob A, Savy A, Dickx Y, Bonnin D, Lecomte C, et al. Monobac system–A single baculovirus for the production of rAAV. Microorganisms. 2021;9(9):1799.

61. Liu S, Li J, Peraramelli S, Luo N, Chen A, Dai M, et al. Systematic comparison of rAAV vectors manufactured using large-scale suspension cultures of Sf9 and HEK293 cells. Molecular Therapy. 2024;32(1):74–83.

62. Merten OW. Development of Stable Packaging and Producer Cell Lines for the Production of AAV Vectors. Microorganisms. 2024;12(2):384.

63. Széliová D, Ruckerbauer DE, Galleguillos SN, Petersen LB, Natter K, Hanscho M, et al. What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. Metabolic engineering. 2020;61:288–300.

64. Hefzi H, Martínez-Monge I, Marin de Mas I, Cowie NL, Toledo AG, Noh SM, et al. Multiplex genome editing eliminates lactate production without impacting growth rate in mammalian cells. Nature Metabolism. 2025; p. 1–16.

65. Cervelli T, Backovic A, Galli A. Formation of AAV single stranded DNA genome from a circular plasmid in Saccharomyces cerevisiae. PloS one. 2011;6(8):e23474.

66. Backovic A, Cervelli T, Salvetti A, Zentilin L, Giacca M, Galli A. Capsid protein expression and adeno-associated virus like particles assembly in Saccharomyces cerevisiae. Microbial cell factories. 2012;11:1–14.

67. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. Journal of general virology. 1977;36(1):59–72.

68. Bairoch A. The cellosaurus, a cell-line knowledge resource. Journal of biomolecular techniques: JBT. 2018;29(2):25.

69. Abaandou L, Quan D, Shiloach J. Affecting HEK293 cell growth and production performance by modifying the expression of specific genes. Cells. 2021;10(7):1667.

70. Malm M, Kuo CC, Barzadd MM, Mebrahtu A, Wistbacka N, Razavi R, et al. Harnessing secretory pathway differences between HEK293 and CHO to rescue production of difficult to express proteins. Metabolic engineering. 2022;72:171–187.

71. Jang M, Pete ES, Bruheim P. The impact of serum-free culture on HEK293 cells: From the establishment of suspension and adherent serum-free adaptation cultures to the investigation of growth and metabolic profiles. Frontiers in Bioengineering and Biotechnology. 2022;10:964397.

72. Tan E, Chin CSH, Lim ZFS, Ng SK. HEK293 cell line as a platform to produce recombinant proteins and viral vectors. Frontiers in bioengineering and biotechnology. 2021;9:796991.

73. 520472 RI. HEK293 Cell Line Market. Verified Market Reports. 2023;.

74. Bylund L, Kytölä S, Lui WO, Larsson C, Weber G. Analysis of the cytogenetic stability of the human embryonal kidney cell line 293 by cytogenetic and STR profiling approaches. Cytogenetic and genome research. 2004;106(1):28–32.

75. Lin YC, Boone M, Meuris L, Lemmens I, Van Roy N, Soete A, et al. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nature communications. 2014;5(1):4767.

76. Malm M, Saghaleyni R, Lundqvist M, Giudici M, Chotteau V, Field R, et al. Evolution from adherent to suspension: systems biology of HEK293 cell line development. Scientific reports. 2020;10(1):18996.

77. Pistek M, Kahlig CI, Hackl M, Unterthurner S, Kraus B, Grabherr R, et al. Comprehensive mRNA-sequencing-based characterization of three HEK-293 cell lines during an rAAV production process for gene therapy applications. Biotechnology Journal. 2023;18(8):2200513.

78. Arena TA, Chou B, Harms PD, Wong AW. An anti-apoptotic HEK293 cell line provides a robust and high titer platform for transient protein expression in bioreactors. In: MAbs. vol. 11. Taylor & Francis; 2019. p. 977–986.

79. Johari YB, Scarrott JM, Pohle TH, Liu P, Mayer A, Brown AJ, et al. Engineering of the CMV promoter for controlled expression of recombinant genes in HEK293 cells. Biotechnology Journal. 2022;17(8):2200062.

80. Carpentier E, Paris S, Kamen AA, Durocher Y. Limiting factors governing protein expression following polyethylenimine-mediated gene transfer in HEK293-EBNA1 cells. Journal of biotechnology. 2007;128(2):268–280.

81. Gurazada SGR, Kennedy HM, Braatz RD, Mehrman SJ, Polson SW, Rombel IT. HEK-omics: The promise of omics to optimize HEK293 for recombinant adeno-associated virus (rAAV) gene therapy manufacturing. Biotechnology Advances. 2024; p. 108506.

82. Rodrigues A, Guerreiro M, Formas-Oliveira A, Fernandes P, Blechert AK, Genzel Y, et al. Increased titer and reduced lactate accumulation in recombinant retrovirus production through the down-regulation of HIF1 and PDK. Biotechnology and bioengineering. 2016;113(1):150–162.

83. Zehetner L, Szeliova D, Kraus B, Hernandez Bort J, Zanghellini J. Improving HEK293-based AAV-production using GSMMs, and a multi-omics approach. bioRxiv. 2024; p. 2024–10.

84. Hayashi Y, Yokota A, Harada H, Huang G. Hypoxia/pseudohypoxia-mediated activation of hypoxia-inducible factor-1$\alpha$ in cancer. Cancer science. 2019;110(5):1510–1517.

85. Kelly B, O'neill LA. Metabolic reprogramming in macrophages and dendritic cells in innate immunity. Cell research. 2015;25(7):771–784.

86. Sharma M, Boytard L, Hadi T, Koelwyn G, Simon R, Ouimet M, et al. Enhanced glycolysis and HIF-1$\alpha$ activation in adipose tissue macrophages sustains local and systemic interleukin-1$\beta$ production in obesity. Scientific reports. 2020;10(1):5555.

87. Dey A, Prabhudesai S, Zhang Y, Rao G, Thirugnanam K, Hossen MN, et al. Cystathione $\beta$-synthase regulates HIF-1$\alpha$ stability through persulfidation of PHD2. Science advances. 2020;6(27):eaaz8534.

88. Semenza GL, et al. HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. The Journal of clinical investigation. 2013;123(9):3664–3671.

89. Patten DA, Lafleur VN, Robitaille GA, Chan DA, Giaccia AJ, Richard DE. Hypoxia-inducible factor-1 activation in nonhypoxic conditions: the essential role of mitochondrial-derived reactive oxygen species. Molecular biology of the cell. 2010;21(18):3247–3257.

90. Chandel NS, McClintock DS, Feliciano CE, Wood TM, Melendez JA, Rodriguez AM, et al. Reactive oxygen species generated at mitochondrial complex III stabilize hypoxia-inducible factor-1$\alpha$ during hypoxia: a mechanism of O2 sensing. Journal of Biological Chemistry. 2000;275(33):25130–25138.

91. Chua YL, Dufour E, Dassa EP, Rustin P, Jacobs HT, Taylor CT, et al. Stabilization of hypoxia-inducible factor-1$\alpha$ protein in hypoxia occurs independently of mitochondrial reactive oxygen species production. Journal of Biological Chemistry. 2010;285(41):31277–31284.

92. Stegen S, Van Gastel N, Eelen G, Ghesquière B, D'Anna F, Thienpont B, et al. HIF-1$\alpha$ promotes glutamine-mediated redox homeostasis and glycogen-dependent bioenergetics to support postimplantation bone cell survival. Cell metabolism. 2016;23(2):265–279.

93. Zhang J, Pavlova NN, Thompson CB. Cancer cell metabolism: the essential role of the nonessential amino acid, glutamine. The EMBO journal. 2017;36(10):1302–1315.

94. De Saedeleer CJ, Copetti T, Porporato PE, Verrax J, Feron O, Sonveaux P. Lactate activates HIF-1 in oxidative but not in Warburg-phenotype human tumor cells. PLoS One. 2012;.

95. Kierans S, Taylor C. Regulation of glycolysis by the hypoxia-inducible factor (HIF): implications for cellular physiology. The Journal of physiology. 2021;599(1):23–37.

96. Kim Jw, Tchernyshyov I, Semenza GL, Dang CV. HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. Cell metabolism. 2006;3(3):177–185.

97. Xiao W, Loscalzo J. Metabolic responses to reductive stress. Antioxidants & redox signaling. 2020;32(18):1330–1347.

98. Zhou J, Schmid T, Brune B. Tumor necrosis factor-$\alpha$ causes accumulation of a ubiquitinated form of hypoxia inducible factor-1$\alpha$ through a nuclear factor-$\kappa$B-dependent pathway. Molecular biology of the cell. 2003;14(6):2216–2225.

99. Sartori-Cintra AR, de Mara CS, Argolo DL, Coimbra IB. Regulation of hypoxia-inducible factor-1$\alpha$ (HIF-1$\alpha$) expression by interleukin-1$\beta$ (IL-1$\beta$), insulin-like growth factors I (IGF-I) and II (IGF-II) in human osteoarthritic chondrocytes. Clinics. 2012;67(1):35–40.

100. Imtiyaz HZ, Simon MC. Hypoxia-inducible factors as essential regulators of inflammation. Diverse Effects of Hypoxia on Tumor Progression. 2010; p. 105–120.

101. Chettimada S, Gupte R, Rawat D, Gebb SA, McMurtry IF, Gupte SA. Hypoxia-induced glucose-6-phosphate dehydrogenase overexpression and-activation in pulmonary artery smooth muscle cells: implication in pulmonary hypertension. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2015;308(3):L287–L300.

102. Jin L, Zhou Y. Crucial role of the pentose phosphate pathway in malignant tumors. Oncology letters. 2019;17(5):4213–4221.

103. Ghanem N, El-Baba C, Araji K, El-Khoury R, Usta J, Darwiche N. The pentose phosphate pathway in cancer: regulation and therapeutic opportunities. Chemotherapy. 2021;66(5-6):179–191.

104. Li X, Zhang KY, Zhang P, Chen LX, Wang L, Xie M, et al. Hydrogen sulfide inhibits formaldehyde-induced endoplasmic reticulum stress in PC12 cells by upregulation of SIRT-1. PloS one. 2014;9(2):e89856.

105. Oakes SA, Papa FR. The role of endoplasmic reticulum stress in human pathology. Annual Review of Pathology: Mechanisms of Disease. 2015;10(1):173–194.

106. Tang J, Ge Y, Yang L, Xu X, Sui T, Ge D, et al. ER stress via CHOP pathway is involved in FK506-induced apoptosis in rat fibroblasts. Cellular Physiology and Biochemistry. 2016;39(5):1965–1976.

107. Muñiz-García A, Romero M, Falcón-Perez JM, Murray P, Zorzano A, Mora S. Hypoxia-induced HIF1$\alpha$ activation regulates small extracellular vesicle release in human embryonic kidney cells. Scientific Reports. 2022;12(1):1443.

108. Zhang W, Zhou X, Yao Q, Liu Y, Zhang H, Dong Z. HIF-1-mediated production of exosomes during hypoxia is protective in renal tubular cells. American Journal of Physiology-Renal Physiology. 2017;313(4):F906–F913.

109. Buchsteiner M, Quek LE, Gray P, Nielsen LK. Improving culture performance and antibody production in CHO cell culture processes by reducing the Warburg effect. Biotechnology and Bioengineering. 2018;115(9):2315–2327.

110. Potter M, Newport E, Morten KJ. The Warburg effect: 80 years on. Biochemical Society Transactions. 2016;44(5):1499–1505.

111. Thompson CB, Vousden KH, Johnson RS, Koppenol WH, Sies H, Lu Z, et al. A century of the Warburg effect. Nature Metabolism. 2023;5(11):1840–1843.

112. Hardie DG. 100 years of the Warburg effect: a historical perspective. Endocrine-Related Cancer. 2022;29(12):T1–T13.

113. Henry O, Durocher Y. Enhanced glycoprotein production in HEK-293 cells expressing pyruvate carboxylase. Metabolic engineering. 2011;13(5):499–507.

114. Martínez-Monge I, Albiol J, Lecina M, Liste-Calleja L, Miret J, Solà C, et al. Metabolic flux balance analysis during lactate and glucose concomitant consumption in HEK293 cell cultures. Biotechnology and Bioengineering. 2019;116(2):388–404.

115. Elias CB, Carpentier E, Durocher Y, Bisson L, Wagner R, Kamen A. Improving glucose and glutamine metabolism of human HEK 293 and trichoplusiani insect cells engineered to express a cytosolic pyruvate carboxylase enzyme. Biotechnology progress. 2003;19(1):90–97.

116. Karengera E, Robotham A, Kelly J, Durocher Y, De Crescenzo G, Henry O. Altering the central carbon metabolism of HEK293 cells: Impact on recombinant glycoprotein quality. Journal of biotechnology. 2017;242:73–82.

117. Vallée C, Durocher Y, Henry O. Exploiting the metabolism of PYC expressing HEK293 cells in fed-batch cultures. Journal of biotechnology. 2014;169:63–70.

118. Splichal RC, Chen K, Walton SP, Chan C. The Role of Endoplasmic Reticulum Stress on Reducing Recombinant Protein Production in Mammalian Cells. Biochemical Engineering Journal. 2024; p. 109434.

119. Tokmakov AA, Kurotani A, Takagi T, Toyama M, Shirouzu M, Fukami Y, et al. Multiple post-translational modifications affect heterologous protein synthesis. Journal of Biological Chemistry. 2012;287(32):27106–27116.

120. Zhang L, Wang A. Virus-induced ER stress and the unfolded protein response. Frontiers in plant science. 2012;3:293.

121. Li S, Kong L, Yu X. The expanding roles of endoplasmic reticulum stress in virus replication and pathogenesis. Critical reviews in microbiology. 2015;41(2):150–164.

122. Maas NL, Diehl JA. Molecular pathways: the PERKs and pitfalls of targeting the unfolded protein response in cancer. Clinical Cancer Research. 2015;21(4):675–679.

123. Walter F, O'Brien A, Concannon CG, Düssmann H, Prehn JH. ER stress signaling has an activating transcription factor 6$\alpha$ (ATF6)-dependent "off-switch". Journal of Biological Chemistry. 2018;293(47):18270–18284.

124. Tsai YC, Weissman AM. The unfolded protein response, degradation from the endoplasmic reticulum, and cancer. Genes & cancer. 2010;1(7):764–778.

125. Formas-Oliveira AS, Basílio JS, Rodrigues AF, Coroadinha AS. Overexpression of ER protein processing and apoptosis regulator genes in human embryonic kidney 293 cells improves gene therapy vectors production. Biotechnology Journal. 2020;15(9):1900562.

126. Satkunanathan S, Thorpe R, Zhao Y. The function of DNA binding protein nucleophosmin in AAV replication. Virology. 2017;510:46–54.

127. Saluri M, Leppert A, Gese GV, Sahin C, Lama D, Kaldmäe M, et al. A "grappling hook" interaction connects self-assembly and chaperone activity of Nucleophosmin 1. PNAS nexus. 2023;2(2):pgac303.

128. Pérez-Rubio P, Lavado-García J, Bosch-Molist L, Romero EL, Cervera L, Gòdia F. Extracellular vesicle depletion and UGCG overexpression mitigate the cell density effect in HEK293 cell culture transfection. Molecular Therapy Methods & Clinical Development. 2024;32(1).

129. Lavado-García J, Pérez-Rubio P, Cervera L, Gòdia F. The cell density effect in animal cell-based bioprocessing: Questions, insights and perspectives. Biotechnology advances. 2022;60:108017.

130. Wosnitzka K, Pariag S, Trottin H, Nyamay'Antu A, Porte M, Hellal M, et al. Optimization of AAV process development: Transfection matters. Cell Gene Ther Insights. 2021;7:1–7.

131. Levy HC, Bowman VD, Govindasamy L, McKenna R, Nash K, Warrington K, et al. Heparin binding induces conformational changes in Adeno-associated virus serotype 2. Journal of structural biology. 2009;165(3):146–156.

132. Yang R, Tran NT, Chen T, Cui M, Wang Y, Sharma T, et al. AAVone: A Cost-Effective, Single-Plasmid Solution for Efficient AAV Production with Reduced DNA Impurities. bioRxiv. 2025; p. 2025–01.

133. Lavado-García J, Jorge I, Cervera L, Vázquez J, Gòdia F. Multiplexed quantitative proteomic analysis of HEK293 provides insights into molecular changes associated with the cell density effect, transient transfection, and virus-like particle production. Journal of proteome research. 2020;19(3):1085–1099.

134. Banik GG, Heath CA. Hybridoma growth and antibody production as a function of cell density and specific growth rate in perfusion culture. Biotechnology and bioengineering. 1995;48(3):289–300.

135. Lavado-García J, Jorge I, Boix-Besora A, Vázquez J, Gòdia F, Cervera L. Characterization of HIV-1 virus-like particles and determination of Gag stoichiometry for different production platforms. Biotechnology and bioengineering. 2021;118(7):2660–2675.

136. Su W, Seymour LW, Cawood R. AAV production in stable packaging cells requires expression of adenovirus 22/33K protein to allow episomal amplification of integrated rep/cap genes. Scientific Reports. 2023;13(1):21670.

137. Jalšić L, Lytvyn V, Elahi SM, Hrapovic S, Nassoury N, Chahal PS, et al. Inducible HEK293 AAV packaging cell lines expressing Rep proteins. Molecular Therapy-Methods & Clinical Development. 2023;30:259–275.

138. Lu M, Lee Z, Lin YC, Irfanullah I, Cai W, Hu WS. Enhancing the production of recombinant adeno-associated virus in synthetic cell lines through systematic characterization. Biotechnology and bioengineering. 2024;121(1):341–354.

139. Lee Z, Lu M, Irfanullah E, Soukup M, Hu WS. Construction of an rAAV producer cell line through synthetic biology. ACS Synthetic Biology. 2022;11(10):3285–3295.

140. Lewis NE, Schramm G, Bordbar A, Schellenberger J, Andersen MP, Cheng J, et al. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. Nature Biotechnology. 2010;28:1279–1285. doi:10.1038/nbt.1711.

141. Terzer M, Maynard ND, Covert MW, Stelling J. Genome-scale metabolic networks. WIREs Systems Biology and Medicine. 2009;1:285–297. doi:10.1002/wsbm.37.

142. Famili I, Palsson B. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. Biophysical Journal. 2003;85:16–26. doi:10.1016/s0006-3495(03)74450-6.

143. Nielsen J. Systems biology of lipid metabolism: from yeast to human. FEBS Letters. 2009;583:3905–3913. doi:10.1016/j.febslet.2009.10.054.

144. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics. 2014;15(2):107–120.

145. Yasemi M, Jolicoeur M. Modelling cell metabolism: a review on constraint-based steady-state and kinetic approaches. Processes. 2021;9(2):322.

146. Orth JD, Thiele I, Palsson B. What is flux balance analysis? Nature Biotechnology. 2010;28:245–248. doi:10.1038/nbt.1614.

147. Calmels C, Cann AM, Malphettes L, Andersen MR. Application of a curated genome-scale metabolic model of cho dg44 to an industrial fed-batch process. Metabolic Engineering. 2019;51:9–19. doi:10.1016/j.ymben.2018.09.009.

148. Ren S, Zeng B, Qian X. Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. BMC Bioinformatics. 2013;14. doi:10.1186/1471-2105-14-s2-s17.

149. Arieef MF, Kasim S, Choon YW, Mohamad MS, Deris S, Napis S. A hybrid of integer differential bees and flux balance analysis for improving succinate and lactate production. International Journal on Advanced Science, Engineering and Information Technology. 2017;7:1615. doi:10.18517/ijaseit.7.4-2.3398.

150. Huang Z, Xu J, Yongky A, Morris C, Polanco A, Reily MD, et al. Cho cell productivity improvement by genome-scale modeling and pathway analysis: application to feed supplements. Biochemical Engineering Journal. 2020;160:107638. doi:10.1016/j.bej.2020.107638.

151. Fouladiha H, Marashi S, Torkashvand F, Mahboudi F, Lewis NE, Vaziri B. A metabolic network-based approach for developing feeding strategies for cho cells to increase monoclonal antibody production. Bioprocess and Biosystems Engineering. 2020;43:1381–1389. doi:10.1007/s00449-020-02332-6.

152. Varma A, Palsson BO. Metabolic flux balancing: basic concepts, scientific and practical use. Bio/technology. 1994;12(10):994–998.

153. Raman K, Chandra N. Flux balance analysis of biological systems: applications and challenges. Briefings in bioinformatics. 2009;10(4):435–449.

154. Abedpour N, Kollmann M. Resource constrained flux balance analysis predicts selective pressure on the global structure of metabolic networks. BMC systems biology. 2015;9:1–13.

155. Feist AM, Palsson BO. The biomass objective function. Current opinion in microbiology. 2010;13(3):344–349.

156. Toro L, Pinilla L, Avignone-Rossa C, Ríos-Estepa R. An enhanced genome-scale metabolic reconstruction of Streptomyces clavuligerus identifies novel strain improvement strategies. Bioprocess and biosystems engineering. 2018;41:657–669.

157. Serrano-Bermúdez LM, González Barrios AF, Maranas CD, Montoya D. Clostridium butyricum maximizes growth while minimizing enzyme usage and ATP production: metabolic flux distribution of a strain cultured in glycerol. BMC systems biology. 2017;11:1–13.

158. Széliová D, Iurashev D, Ruckerbauer DE, Koellensperger G, Borth N, Melcher M, et al. Error propagation in constraint-based modeling of Chinese hamster ovary cells. Biotechnology Journal. 2021;16(4):2000320.

159. Krumholz EW, Libourel IG. Thermodynamic constraints improve metabolic networks. Biophysical journal. 2017;113(3):679–689.

160. Salvy P, Fengos G, Ataman M, Pathier T, Soh KC, Hatzimanikatis V. pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis. Bioinformatics. 2019;35(1):167–169.

161. Park JO, Rubin SA, Xu YF, Amador-Noguez D, Fan J, Shlomi T, et al. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nature chemical biology. 2016;12(7):482–489.

162. Suthers PF, Foster CJ, Sarkar D, Wang L, Maranas CD. Recent advances in constraint and machine learning-based metabolic modeling by leveraging stoichiometric balances, thermodynamic feasibility and kinetic law formalisms. Metabolic engineering. 2021;63:13–33.

163. De Martino A, De Martino D, Marinari E. The Essential Role of Thermodynamics in metabolic network modeling: physical insights and computational challenges. In: Chemical Kinetics: Beyond the Textbook. World Scientific; 2020. p. 455–471.

164. Beber ME, Gollub MG, Mozaffari D, Shebek KM, Flamholz AI, Milo R, et al. eQuilibrator 3.0: a database solution for thermodynamic constant estimation. Nucleic acids research. 2022;50(D1):D603–D609.

165. Mao Z, Niu J, Zhao J, Huang Y, Wu K, Yun L, et al. ECMpy 2.0: A Python package for automated construction and analysis of enzyme-constrained models. Synthetic and systems biotechnology. 2024;9(3):494–502.

166. Bulović A, Fischer S, Dinh M, Golib F, Liebermeister W, Poirier C, et al. Automated generation of bacterial resource allocation models. Metabolic engineering. 2019;55:12–22.

167. Wang J, Yang Z, Chen C, Yao G, Wan X, Bao S, et al. MPEK: a multitask deep learning framework based on pretrained language models for enzymatic reaction kinetic parameters prediction. Briefings in Bioinformatics. 2024;25(5):bbae387.

168. Li F, Yuan L, Lu H, Li G, Chen Y, Engqvist MK, et al. Deep learning-based k cat prediction enables improved enzyme-constrained model reconstruction. Nature Catalysis. 2022;5(8):662–672.

169. Kroll A, Rousset Y, Hu XP, Liebrand NA, Lercher MJ. Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. Nature communications. 2023;14(1):4139.

170. Matiolli CC, Soares RC, Alves HL, Abreu IA. Turning the knobs: the impact of post-translational modifications on carbon metabolism. Frontiers in Plant Science. 2022;12:781508.

171. Robinson PK. Enzymes: principles and biotechnological applications. Essays in biochemistry. 2015;59:1.

172. Strain B, Morrissey J, Antonakoudis A, Kontoravdi C. How reliable are Chinese hamster ovary (CHO) cell genome-scale metabolic models? Biotechnology and Bioengineering. 2023;120(9):2460–2478.

173. Robinson JL, Kocabaş P, Wang H, Cholley PE, Cook D, Nilsson A, et al. An atlas of human metabolism. Science signaling. 2020;13(624):eaaz1482.

174. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nature biotechnology. 2018;36(3):272–281.

175. Moškon M, Režen T. Context-specific genome-scale metabolic modelling and its application to the analysis of COVID-19 metabolic signatures. Metabolites. 2023;13(1):126.

176. Cho JS, Gu C, Han TH, Ryu JY, Lee SY. Reconstruction of context-specific genome-scale metabolic models using multiomics data to study metabolic rewiring. Current Opinion in Systems Biology. 2019;15:1–11.

177. Vieira V, Ferreira J, Rocha M. A pipeline for the reconstruction and evaluation of context-specific human metabolic models at a large-scale. PLoS computational biology. 2022;18(6):e1009294.

178. Correia S, Rocha M. A critical evaluation of methods for the reconstruction of tissue-specific models. In: Progress in Artificial Intelligence: 17th Portuguese Conference on Artificial Intelligence, EPIA 2015, Coimbra, Portugal, September 8-11, 2015. Proceedings 17. Springer; 2015. p. 340–352.

179. Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE. A systematic evaluation of methods for tailoring genome-scale metabolic models. Cell systems. 2017;4(3):318–329.

180. Walakira A, Rozman D, Režen T, Mraz M, Moškon M. Guided extraction of genome-scale metabolic models for the integration and analysis of omics data. Computational and structural biotechnology journal. 2021;19:3521–3530.

181. Richelle A, Chiang AW, Kuo CC, Lewis NE. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. PLoS computational biology. 2019;15(4):e1006867.

182. Gopalakrishnan S, Joshi CJ, Valderrama-Gómez MÁ, Icten E, Rolandi P, Johnson W, et al. Guidelines for extracting biologically relevant context-specific metabolic models using gene expression data. Metabolic engineering. 2023;75:181–191.

183. Pacheco MP, Pfau T, Sauter T. Benchmarking procedures for high-throughput context specific reconstruction algorithms. Frontiers in physiology. 2016;6:410.

184. Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Molecular systems biology. 2010;6(1):401.

185. Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. BMC systems biology. 2012;6:1–16.

186. Vlassis N, Pacheco MP, Sauter T. Fast reconstruction of compact context-specific metabolic network models. PLoS computational biology. 2014;10(1):e1003424.

187. Tefagh M, Boyd SP. SWIFTCORE: a tool for the context-specific reconstruction of genome-scale metabolic networks. BMC bioinformatics. 2020;21:1–14.

188. Schultz A, Qutub AA. Reconstruction of tissue-specific metabolic networks using CORDA. PLoS computational biology. 2016;12(3):e1004808.

189. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS computational biology. 2008;4(5):e1000082.

190. Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BØ, Hyduke DR. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics. 2013;29(22):2900–2908.

191. Jenior ML, Moutinho Jr TJ, Dougherty BV, Papin JA. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. PLoS computational biology. 2020;16(4):e1007099.

192. Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. Bioinformatics. 2010;26(24):3140–3142.

193. Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. PLoS computational biology. 2012;8(5):e1002518.

194. Robaina Estévez S, Nikoloski Z. Context-specific metabolic model extraction based on regularized least squares optimization. PloS one. 2015;10(7):e0131875.

195. Jensen PA, Papin JA. Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics. 2011;27(4):541–547.

196. Zhang C, Lee S, Bidkhori G, Benfeitas R, Lovric A, Chen S, et al. RMetD2: a tool for integration of relative transcriptomics data into Genome-scale metabolic models. BioRxiv. 2019; p. 663096.

197. Ravi S, Gunawan R. ΔFBA—Predicting metabolic flux alterations using genome-scale metabolic models and differential transcriptomic data. PLoS Computational Biology. 2021;17(11):e1009589.

198. Mirhakkak MH, Chen X, Ni Y, Heinekamp T, Sae-Ong T, Xu LL, et al. Genome-scale metabolic modeling of Aspergillus fumigatus strains reveals growth dependencies on the lung microbiome. Nature Communications. 2023;14(1):4369.

199. Ardalani O, Phaneuf PV, Mohie OS, Nielsen LK, Palsson B. Pangenome reconstruction of Lactobacillaceae metabolism predicts species-specific metabolic traits. bioRxiv. 2023; p. 2023–09.

200. Monk JM. Genome-scale metabolic network reconstructions of diverse Escherichia strains reveal strain-specific adaptations. Philosophical Transactions of the Royal Society B. 2022;377(1861):20210236.

201. Van der Maaten L, Hinton G. Visualizing non-metric similarities in multiple maps. Machine learning. 2012;87:33–55.

202. Jaccard P. The distribution of the flora in the alpine zone. 1. New phytologist. 1912;11(2):37–50.

203. Kenefake D, Armingol E, Lewis NE, Pistikopoulos EN. An improved algorithm for flux variability analysis. BMC bioinformatics. 2022;23(1):550.

204. Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. BMC bioinformatics. 2010;11:1–3.

205. Guebila MB. VFFVA: dynamic load balancing enables large-scale flux variability analysis. BMC bioinformatics. 2020;21:1–13.

206. San Román M, Cancela H, Acerenza L. Source and regulation of flux variability in Escherichia coli. BMC systems biology. 2014;8:1–11.

207. Abbate T, Dewasme L, Wouwer AV, Bogaerts P. Adaptive flux variability analysis of HEK cell cultures. Computers & Chemical Engineering. 2020;133:106633.

208. Price ND, Schellenberger J, Palsson BO. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. Biophysical journal. 2004;87(4):2172–2186.

209. Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. PloS one. 2014;9(2):e86587.

210. Galuzzi BG, Milazzo L, Damiani C. Best practices in flux sampling of constrained-based models. In: International Conference on Machine Learning, Optimization, and Data Science. Springer; 2022. p. 234–248.

211. Keaty TC, Jensen PA. Gapsplit: efficient random sampling for non-convex constraint-based models. Bioinformatics. 2020;36(8):2623–2625.

212. Haraldsdóttir HS, Cousins B, Thiele I, Fleming RM, Vempala S. CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. Bioinformatics. 2017;33(11):1741–1743.

213. Saa PA, Zapararte S, Drovandi CC, Nielsen LK. LooplessFluxSampler: an efficient toolbox for sampling the loopless flux solution space of metabolic models. BMC bioinformatics. 2024;25(1):3.

214. van Rosmalen R, Moreno-Paz S, Duman-Özdamar Z, Suarez-Diez M. CFSA: Comparative flux sampling analysis as a guide for strain design. Metabolic Engineering Communications. 2024;19:e00244.

215. Sharma K, Marucci L, Abdallah ZS. FluxGAT: Integrating Flux Sampling with Graph Neural Networks for Unbiased Gene Essentiality Classification. arXiv preprint arXiv:240318666. 2024;.

# 5 Thesis relevant publications

**Statement of Acceptance** A copy of the acceptance confirmation for the manuscript "Improving HEK293-based AAV-production using GSMMs, and a multi-omics approach" from Metabolic Engineering is provided. As the final, published version of the article was not available online at the time of printing this thesis the author's version of the accepted manuscript is provided.

Received: 20 December 2022 | Revised: 22 February 2023 | Accepted: 14 April 2023

DOI: 10.1002/biot.202200636

Biotechnology Journal

REVIEW

# Optimizing VLP production in gene therapy: Opportunities and challenges for *in silico* modeling

**Leopold Zehetner**[1,2] | **Diana Széliová**[2] | **Barbara Kraus**[1] | **Michael Graninger**[1] | **Jürgen Zanghellini**[2] | **Juan A. Hernandez Bort**[1]

[1]Biotherapeutics Process Development, Baxalta Innovations GmbH, a Part of Takeda Companies, Orth an der Donau, Austria

[2]Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria

**Correspondence**
Jürgen Zanghellini, Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria.
Email: juergen.zanghellini@univie.ac.at

**Funding information**
Baxalta Innovations GmbH, a Part of Takeda Companies

## Abstract

Over the past decades, virus-like particle (VLP)-based gene therapy (GT) evolved as a promising approach to cure inherited diseases or cancer. Tremendous costs due to inefficient production processes remain one of the key challenges despite considerable efforts to improve titers. This review aims to link genome-scale metabolic models (GSMMs) to cell lines used for VLP synthesis for the first time. We summarize recent advances and challenges of GSMMs for Chinese hamster ovary (CHO) cells and provide an overview of potential cell lines used in GT. Although GSMMs in CHO cells led to significant improvements in growth rates and recombinant protein (RP)-production, no GSMM has been established for VLP production so far. To facilitate the generation of GSMM for these cell lines we further provide an overview of existing omics data and the highest production titers so far reported.

**KEYWORDS**
gene therapy, genome-scale metabolic model, HEK-293, omics, virus-like particle

## 1 | INTRODUCTION

Genetic disorders, such as cancer or rare genetic inherited diseases, may be caused by mutation, alteration, deletion, or differential expression of specific genes. Gene therapy (GT) is a promising approach based on the use of nucleic acids harbored in so-called vectors, which are introduced into somatic cells to correct or prevent pathological processes in patients. The most popular GT-based application might be the replacement of defective genes by others that contain the information to produce the protein of interest correctly.[1,2] However, this emerging technology still brings some challenges during its application in patients, for example targeting with the appropriate vector the desired organ region. Therefore, numerous viral and non-viral vectors or even hybrids from both were developed or modified in the past to resolve this aspect. Some examples of vectors may include virus-like particle (VLPs),[3] artificial polymers,[4] or extracellular vesicles.[5]

VLPs form the most prominent group of vectors used in GT and can be produced by engineered host cell lines.[3] Although production platforms for VLP expression have been established and optimized in the last years, titers and biopotency are still insufficient to meet future demands.[6] Many efforts have been made to solve these limitations, for example, via cell culture media optimization,[7] and cell line engineering.[8] However, some of these efforts are time-consuming trial-and-error approaches; for instance, screening a pool of cells to pick out the best clones or using rational design approaches, like design-of-experiments (design-of-experiments (DoE)), to modify the medium composition to find the best possible medium formulation for boosting titers.

A more sophisticated methodology to shed light on the cellular black box may be based on *in silico* tools. It was already shown that

**Abbreviation:** CSD, computational strain design; GSMM, genome-scale metabolic model; GT, gene therapy; VLP, virus-like particle.

media composition could be optimized by analyzing metabolic flux distributions inside host cells and subsequently applying this knowledge to reduce the accumulation of byproducts while increasing cellular productivity.[9,10] However, additional factors may play a role in how the metabolism works; for example, it was demonstrated that some viruses could reprogram host cells' metabolism. These changes in the metabolism are even more significant than any changes triggered by the expression of recombinant protein (RPs)[11,12] and may lead to lower production titer due to activation of self-defense pathways of host cells.[13] This knowledge adds further complexity to elucidate the regulatory mechanisms behind VLP production.

To tackle the resulting derived limitations of low production titers, different strategies have been established in the past, including DoE,[14] extensive passaging of host cells,[15] hypothesis-driven genetic engineering,[8] the implementation of alternative host cell lines,[16] or media optimization.[7] It was observed that extensive passaging of Per.C6 cells resulted in an evolved cell line achieving 3-fold higher adenovirus (adenovirus (AdV)) production and a doubled growth rate compared to the parental strain leading to a reduction in the number of production batches, consequently.[15] In another approach, recombinant adeno-associated virus (recombinant adeno-associated virus (rAAV)) titers could be increased from $2.5 \times 10^9$ viral genomes (viral genomes (VG))/mL to $2.7 \times 10^{10}$ VG/mL in HEK-293F1 cells by media optimization,[7] although this is not the highest reported rAAV titer so far.[17] However, despite this progress cost of production are still high, and VLP-based drugs remain top among the list of the most expensive therapeutics in the world.[18] Consequently, novel strategies must be developed and implemented to increase titers and reduce costs while considering biosafety aspects and maintaining product quality.

In this review, we aim to give a condensed overview of genome-scale metabolic model (GSMM)- and omics-based bioprocess optimization with a strong focus on Chinese hamster ovary (CHO) cell lines since there have been no GSMMs established for cell lines used in GT so far. Nevertheless, similar approaches may be used for these cell lines. In the second part, we summarize cell lines and VLPs, most utilized in the industry for GT and optimization strategies as visualized in Figure 1.

## 2 | APPROACHES FOR PROCESS OPTIMIZATION USING CONSTRAINT-BASED MODELING

In the last two decades, systems biotechnological approaches have significantly altered the way industrial bioprocesses are designed and operated[19,20]–at least for microbial systems. A key method to better understand upstream processes is constraint-based modeling (constraint-based modeling (CBM)), which allows for analyzing intracellular metabolic processes. At its heart sits a GSMM, which is a mathematical representation of a cells metabolism and captures the stoichiometry of all metabolic transformation processes. Since the first GSMM was developed in 1999 for *Haemophilus influenzae*[21] a large number of metabolic models has been developed and successfully used.[22] For instance, GSMMs have been used to overcome various challenges in bioprocess optimization, including media optimization[23] or engineering target identification.[19]

Detailed protocols to reconstruct GSMMs from scratch,[24] supported by several automatic reconstruction pipelines, are available[25] resulting in several hundred GSMMs, which were reviewed in a recent GSMM-specific phylogenetic study.[22] Moreover, community-driven, highly manually curated global GSMMs for a few model organisms are currently available. These models include: *Homo sapiens*,[26,27] *Cricetulus griseus*,[28] *Mus musculus*, *Drosophila melanogaster*, *Danio rerio*, *Saccharomyces cerevisiae*,[29] or *Escherichia coli*[30], among others. Such global models allow cell- or tissue-specific reconstructions based on transcriptomes and metabolic exchange rates. However, the *de novo* reconstruction of global GSMMs for mammalian cells remains challenging and needs extensive manual curation to avoid incorrect transport reactions between organelles, or futile cycles, for example.[31]

Compared to global reconstructions, cell line-specific GSMMs allow for more accurate predictions[32] as their construction typically relies on the integration of cell line-specific transcriptomic and fluxomic data. Various algorithms for generating strain, tissue, or cell line-specific models exist.[33–39] However, certain issues remain, including the selection of transcriptomic thresholds[40] and the order of constraint introduction.[41–43] Thus, consensus pipelines remain to be established.

The use of cell line specific GSMMs for studying bioprocesses has been increasing over the last years. However, accurate dry weight measurements are often ignored in these models[44,45], despite solid evidence that time-resolved extracellular metabolome quantification[26–28] combined with dry weight determination of biomass to calculate exchange fluxes[46–49] improves cell line specific models even further. Most often, the measured exometabolome is limited to metabolites with well-established quantification protocols, like amino acids, glucose, lactate, and ammonia. With recent whole metabolome quantification approaches[50], more exchange fluxes might be available for integration in the future.

In general, bioprocesses can be optimized by improving growth rate or product formation. These cell-specific phenotypes can be assessed by flux balance analysis (flux balance analysis (FBA)), which is based on the stoichiometric matrix $S$, including metabolites as rows, reactions as columns, and the corresponding fluxes as a vector $v$ being constrained by lower ($l$) and upper ($u$) boundaries. Under a steady-state assumption of the system, so that influx and efflux of metabolites are in equilibrium:

$$Sv = 0 \tag{1a}$$

$$l \leq v \leq u \tag{1b}$$

The higher number of reactions compared to the number of metabolites in large networks results in a solution space with various possible solutions. Therefore, optimizing for a specific reaction, most often an artificial biomass reaction or a product-related reaction, leads to a specific solution.[51] GSMMs have been used as platforms to

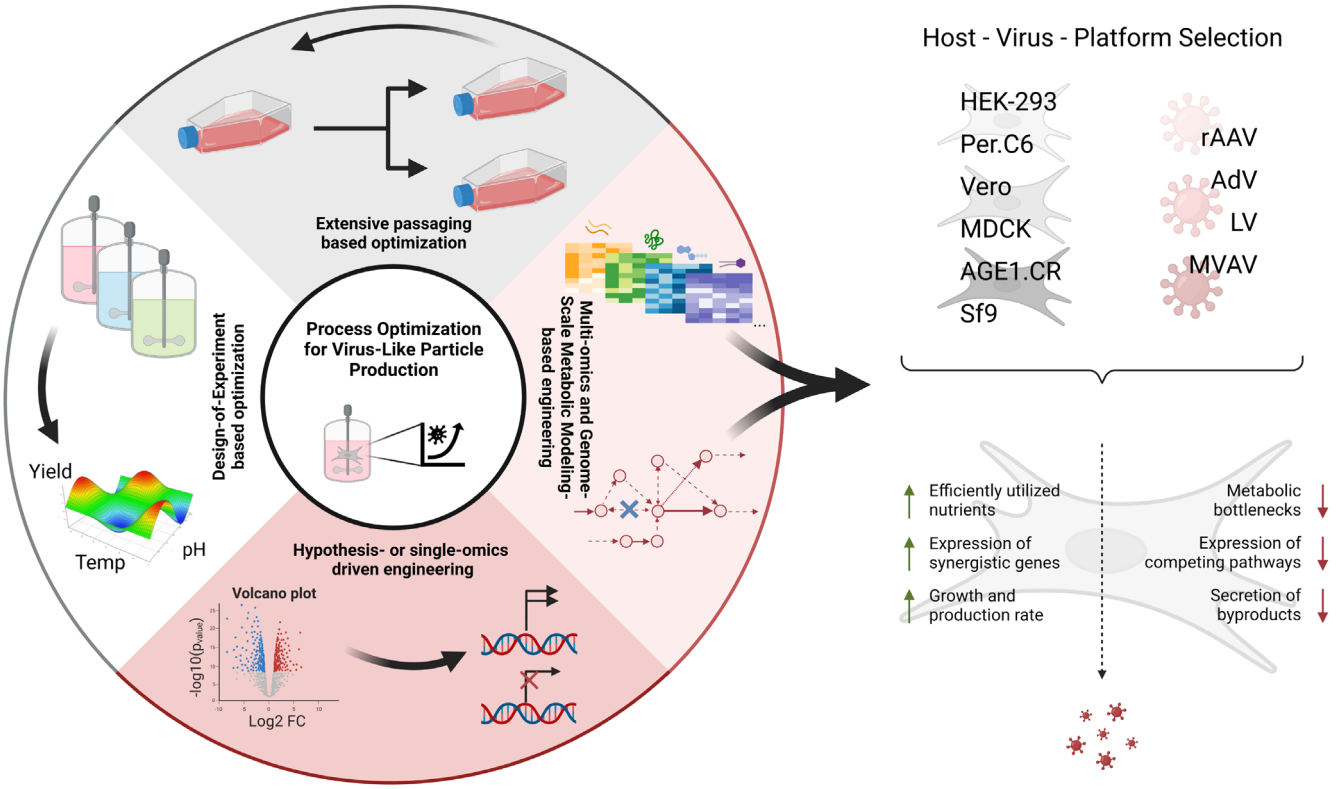


**FIGURE 1** Overview of current strategies for bioprocess optimization and host-virus platforms in gene therapy. Implementing genome-scale metabolic models for these cell lines based on multi-omics data may result in higher titers.

integrate experimentally generated multi-omics data and run simulations to identify metabolic bottlenecks under different conditions.[28] The most common constraints are metabolite exchange rates, calculated from concentration changes of extracellular metabolites.[46] Although FBA is still the most common approach for optimization of objective functions, it has been shown that FBA predicts growth rates in linear dependence on uptake rates without considering overflow metabolism.[52,53]. Consequently, alternative strategies have been developed to reproduce more accurate growth rates, such as hybrid-FBA,[54] or enzyme-constrained FBA.[52] For example, the overestimation of growth rates could be decreased from approx. 4-fold to 1.05-fold by applying hybrid-FBA.[54]

Alternative or additional constraints may be introduced to obtain more reliable simulation results. Several types of constraints can be used, including enzyme constraints,[55] thermodynamic constraints,[56] regulatory constraints[57] or hybrid versions thereof.[53,58] However, all constraining methods include certain disadvantages, which need to be considered when integrating them. Common types of CBM were summarized in an extensive review, discussing available tools, applications, and challenges in microbial cell factory design.[59]

In particular, enzyme-constrained genome-scale metabolic model (enzyme-constrained genome-scale metabolic models (ecGSMMs) s) have been shown to improve accuracy tremendously and allow condition-specific constraining of models since they rely on integrating quantitative proteomics measurements into models. Recently, software tools to introduce enzyme-constraints[55,60,61] have been published following the general equation: $v_i \leq k_{cat,i} \cdot [E_i]$ where $v_i$ is the flux of a certain reaction $i$, catalyzed by an enzyme $E_i$, $k_{cat,i}$ is the corresponding turnover number of the enzyme, obtained from databases, such as BRENDA[62] and $[E_i]$ is the concentration of enzyme(s) catalyzing the reaction.

Despite valuable insights and promising results from ecGSMMs, the lack of catalytic values of enzymes in databases limit the accuracy of these models. Several approaches have been developed to predict or improve the estimates of turnover numbers of enzymes. These include prediction of turnover numbers based on quantitative proteomics and fluxomics data under different conditions,[63] machine learning[64] and deep learning approaches,[65] and an approach based on metabolic control analysis to improve the turnover number estimates.[66] Until now, no comparative study has been published to determine which approach results in the most accurate ecGSMMs.

Recently, a new modeling framework – resource balance analysis (resource balance analysis (RBA)) – has been gaining popularity. In addition to a metabolic network, RBA-based models contain reactions for the synthesis of molecular machines (e.g., enzymes, ribosomes) and predict metabolic fluxes as well as the concentrations of the corresponding catalysts.[67] RBA-based models can be used to study resource reallocation in different media or after knock-out or differential expression and may lead to more realistic strain or medium designs in the future. Recently, a tool has been published that automatically builds RBA-based models from GSMMs.[68]

To efficiently design strains and predict engineering targets *in silico*, numerous tools for computational strain design (computational strain design (CSD)) have been developed as reviewed recently,[69] and condensed in workbenches.[70,71] CSD tools have been successfully applied to improve titers in microbial cell factories, such as *S. roseosporus*; three genes were predicted in the GSMM to have synergistic effects on daptomycin production and led to a titer improvement of 43.2%.[72] However, CSD tools have not been applied to mammalian cell lines so far.[69]

Numerous examples of successful bioprocess optimization using GSMMs are available. However, they typically focus on producing small molecules in well-studied microbial cells, such as *E. coli*[20] or *S. cerevisiae*,[19] which have less complex genomes or metabolic networks, and a higher amount of annotated genes than mammalian or insect cells. Moreover, due to the structural simplicity of small molecules, their production is less challenging than the production, secretion, and purification of RPs or VLPs. Additionally, host cell metabolism is influenced to a higher degree when synthesizing complex products compared to small molecules.

Surprisingly, GSMM of human tissues have been successfully used in disease modeling,[73] and drug target discovery,[74] whereas GSMMs of human cell lines are currently not used in bioprocess development. Except for small-scale models utilized for metabolic flux analysis (metabolic flux analysis (MFA)),[47] there are currently no GSMMs available for any cell line used for VLP production, even though highly curated reference models are available for certain species.[26,27] This is certainly a missed opportunity to optimize manufacturing in GT. Due to the lack of studies, we focus on the most prominently studied mammalian cell line in industry, CHO, even though production in these cells is limited to RPs so far and not to GT products.

## 2.1 | CHO models as opportunity to reconstruct mammalian cell models

Recently, comprehensive reviews summarized the efforts to optimize CHO-based bioprocesses utilizing *in silico* tools.[75,76] Therefore, we focus only on recent approaches to improve these bioprocesses.

The generation of a community-driven reference model for CHO cells,[28] iCHO1766, formed the basis for numerous studies, primarily focusing on optimizing media composition under batch,[10] fed-batch,[77] and continuous culturing conditions,[9] as well as for monoclonal antibody (monoclonal antibody (mAB)) production improvements,[78,79] and by integrating time-resolved transcriptomics data.[80]

For example, to optimize the medium composition, Huang et al.[23] integrated metabolite exchange rates collected during a fed-batch process into iCHO1766 and validated predicted fluxes with transcriptomics data. Based on the knowledge generated during their simulations, the amino acid concentration of leucine and valine was dramatically increased in the media leading to a production increase of 11.8%. Although their model was not specifically tailored to the cell line clone (neither cell-specific model nor cell-specific biomass composition), it was a promising example of how GSMMs facilitated process optimization and avoided time-consuming DoEs. Even though artificial biomass reactions are commonly used for growth simulation as an objective function,[46] this concept might be questionable: Schinn et al.[81] showed much more accurate growth predictions when using reactive oxygen species (reactive oxygen species (ROS)) synthesis or mitochondrial reduced nicotinamide-adenine dinucleotide (reduced nicotinamide-adenine dinucleotide (NADH)) regeneration as objective function for example. This aspect needs to be studied in more detail since selecting an appropriate objective function is one of the most fundamental pillars of the simulation of GSMMs.

Apart from metabolism, protein secretion machinery is an additional bottleneck for producing RPs in CHO cells. Hence, Gutierrez et al.[82] coupled a model of a secretory pathway to a CHO-GSMM and estimated the energy cost of secretion in terms of adenosine triphosphate (adenosine triphosphate (ATP)) consumption. Furthermore, the model was verified by replicating experimental data of mAB titer improvement by knockdown of the selection marker gene *neoR*. In contrast, Morrissey et al. identified that secretion of proteins has no significant impact on the energy metabolism in CHO-GSMMs.[32] However, Kol et al., showed experimentally that mAB titers, purity, and growth rates increased by knocking out sets of secreted host cell proteins in different CHO clones.[83] The most recent version of CHO-GSMM[32] was generated by combining previous models. By comparing the updated global CHO-GSMM, two cell line specific CHO-GSMMs and a recently published reduced CHO-model,[84] it was shown that the reduced models predict core metabolic functions more precisely, whereas essential gene identification is more accurate in the cell line specific models.[32] Despite the comprehensive comparison, the impact of selecting the most suitable objective function for phenotype prediction remains to be investigated.

Currently, only one ecGSMM for CHO cells has been published based on iCHO2296, which resulted in more accurate prediction of overflow metabolism.[52] Additionally, lactate consumption and utilization as energy source after entering the tricarboxylic acid cycle (tricarboxylic acid cycle (TCA)) were identified,[52] as already examined for HEK-293 cells.[49] Despite the availability of various updates[32,52,82] and omics datasets available for different CHO cell lines,[28,85] CSD-tools have so far not been applied to CHO-GSMMs, which limits improvement strategies.

Besides *in silico* strain optimization, GSMMs may be a cornerstone in the next decade as part of the digitalization, and computational optimization of bioprocesses as described for HEK-293 cells during rAAV production[86] and for CHO cells.[87] To realize the full potential of digital tools for process control and monitoring, robust biomarkers for cellular states throughout the process need to be identified. Established concepts for biomarker identification for diagnosis can be adapted for bioprocesses based on multi-omics data.[88] For example, comprehensive quantitative metabolomics analyses, as performed for CHO[45] and HEK-293 cells,[44] could form starting points to more efficient and sustainable bioprocesses.

## 2.2 | Specific difficulties in reconstructing VLP-producing cell lines

After GSMM reconstruction, the target production reactions need to be added to the model. While the addition of biosynthesis pathways for small molecules is usually straightforward, there is currently no standardized way to add VLP production into a GSMM. Several approaches are possible, depending on the desired level of detail.

The simplest approach is to add an artificial VLP production reaction with an average composition, similar to a biomass reaction. While for some VLPs the main components are proteins and nucleic acids (rAAV, AdV), for VLPs with a membrane (lentivirus [LV], modified vaccinia ankara virus [MvaV]), lipids need to be included as well. The average composition of nucleic acids can be calculated from the genome sequence and the protein composition from the sequences and stoichiometries of the capsid proteins. However, the lipid composition needs to be measured as it was shown that viruses can reprogram the host cell's lipid metabolism and may have a different membrane composition than the host.[89,90]

We can also model the VLP production in more detail and add separate reactions for the expression, assembly, modifications and secretion of the VLPs (similarly to the mammalian secretory model[82]). This would allow the integration of enzyme constraints for each reaction (i.e., kinetic constants and enzyme concentrations). If available, kinetic models can be used to obtain the necessary parameters. For example, Nguyen et al.[6] generated a mechanistic model for rAAV 5 production in HEK-293 cells consisting of 14 reactions, and estimated kinetic parameters using data from literature and their transient transfection experiments. However, mechanistic models still need to be developed for other viruses and host cells. A detailed overview of the biology of commonly used viruses in GT was recently published by Bulcha et al.[3] which can serve as a starting point for model construction. However, additional experiments are needed to obtain the necessary parameters (plasmid uptake rates, VLP production rates etc.) is needed to tackle common challenges in VLP production. Long-term passaging or cross-contaminations may lead to wrongly packaged VLPs or decrease biosafety profiles.[93,94] Furthermore, stably integrated viral genes necessary for VLP expression may lose their function after numerous passaging steps which reduces VLP titers.[95] These obstacles can be caused by changes in the genome, but could be detected by de novo sequencing of the production cell lines. For example, de novo sequencing of three Vero genomes led to the discovery of 60 potentially intact retroviral genes, which were not identified in a reference genome.[95] Although these genes have not been characterized as harmful to humans, they comprise potential tumorigenicity.[96] Furthermore, epigenomic changes in cell lines have been shown to occur after a certain number of passages, which might be responsible for altered metabolism and production profiles over time. With the invention of third-generation sequencing platforms, it is possible to detect these changes over time and to guarantee the long-term reproducibility of bioprocesses.[97]

Other than detecting the challenges described above, *de novo* sequencing can provide new information that might be useful for genetic engineering. For example, assembled genome can be used to predict unannotated genes' functions. In a recent study, the concept of identifying gene functions is described in detail by combining different databases. The identification of novel enzymes allows the expansion of metabolic networks, which is especially important for mammalian cell lines, and offers novel targets for genetic engineering.[98] Additionally, updated genomes allow more reliable genetic engineering through more accurate primer design and avoiding incorrect cleavages.

Finally, cell line-specific genomes (and transcriptomes) can be used to reconstruct cell line-specific GSMMs. These form a basis for more accurate predictions of bioprocess and cell line engineering strategies, as discussed in the previous section. Currently, no GSMMs are available for the cell lines in Table 1. However, several small-scale metabolic models for MFA have been generated and used for process optimization.[48,99]

# 3 | POTENTIAL CELL LINES USED IN GENE THERAPY SUITABLE FOR FUTURE GSMMs

Numerous cell lines have been screened for VLP production as alternative production platforms with improved biosafety profiles.[16] However, only a few are currently used in industrial processes for producing the most commonly used VLPs, rAAVs, AdVs, LVs[3], or MvaV.[91,92] Therefore, only selected cell lines, listed in Table 1, are discussed in the following sections.

## 3.1 | Advantages of generating cell line-specific genomes

While reference genomes are available for most cell lines listed in Table 1, generating cell line specific genomes and other omics datasets

## 3.2 | HEK-293

HEK-293 cells and derivated clones are the workhorses for producing VLPs in GT. Several strains have been optimized to grow in suspension and under serum-free conditions, the most appropriate conditions in industrial bioprocesses. To achieve these settings, adherent cell lines have been adapted through numerous passaging steps by reducing the presence of serum by each passage[115] and optimized for the production of rAAV[116,117] or RPs which are difficult to express in CHO cells.[118] However, the highest rAAV-titers in HEK-293 cells have been reported using adherently grown HEK-293H in fetal bovine serum (FBS)-containing media,[17] which is not the best choice for manufactured products as recommended by the health authorities due to potential presence of prions or any other compounds, and complicates scale up. Additionally, these cell lines often lack a deep metabolic investigation.

**TABLE 1** Selected cell lines utilized for VLP production in industrial bioprocesses. Currently highest titers reported in literature are presented, as well as genomes, reference-GSMMs, and omics-based optimization approaches for VLP production.

| Cell line | Genome | Reference-GSMM | VLP | Titer (VG/mL) | Omics Data |
|---|---|---|---|---|---|
| HEK-293 | *PRJNA565658*,[100] *PRJEB3209*[101] | Human-Gem,[26] Recon3D[27] | rAAV | $10^{15}$[17] | Proteome[102] |
| | | | AdV | $2 \times 10^{11}$[103] | – |
| | | | LV | $4.5 \times 10^{9}$[104] | Proteome[105] |
| PER.C6 | – | Human-Gem[26], Recon3D[27] | AdV | $10^{5}$VG/cell[106] | – |
| Vero | *PRJNA847024*,[107] *PRJDB2865*[95] | – | AdV | (not quantified)[108] | – |
| | | | rAAV | 75,000 VG/cell[109] | – |
| MDCK | *PRJNA905218*[110] | – | AdV | $2.1 \times 10^{8} ip/mL$[111] | – |
| AGE1.CR.pIX | *PRJNA669953* | – | MvaV | $2.5 \times 10^{10} TCID50/1/day$[91] | – |
| Sf9 | *PRJNA380964*[112] | – | rAAV | $2.4 \times 10^{12}$[113] | Proteome[114] |

Abbreviations: AdV, adenovirus; GSMM, genome-scale metabolic model; LV, lentivirus; MvaV, modified vaccinia ankara virus; rAAV, recombinant adeno-associated virus; VG, viral genome; VLP, virus-like particle.

### 3.2.1 | Growth optimization of HEK-293 cells

This issue may be addressed by investigating the cellular behavior under various conditions using omics experiments to allow the identification of engineering targets for faster optimization and metabolic analyses.[100,101] Recently, a differential analysis of three adherent strains, including HEK-293E and HEK-293T, and two suspended strains, HEK-293H, and HEK-293F, was performed based on transcriptomic and genomic data. After pairwise comparison, five potential target genes were identified to be consistently upregulated in adherent cell lines, with all involved in cellular adhesion, and DHRS3 and LOX among them.[100] Since adherently grown cell lines complicate scale up, dispensable transmembrane proteins, such as DHRS3, have become promising engineering targets. Although DHRS3 consists of fewer amino acid (amino acids (AAs) s) (302), fewer transmembrane domain (4), and can be expected to be expressed only in small amounts than other transmembrane proteins, such as the monocarbohydrate transporter GLUT1 (492 AA, 12 TMD),[119] it should be considered additionally, that DHRS3 consumes NADPH,[120] which decreases reduction potential for metabolic redox reactions and for maintaining a reducing environment.[121] Though this is just exemplary, and the influence of one enzyme may be negligible, metabolic profiling and how resources are allocated and utilized by cells is one of the most important aspects during both early process development and optimization attempts of established processes.

Therefore, a model for HEK-293SF-3FS cells was established and used to analyze the formation of lactate based on 13C-labeled glucose and glutamine substrates during exponential growth.[122] Interestingly, the proportion of lactate generated from glutamine is negligible, whereas 77% of glucose is catabolized to lactate and 15% of glucose is metabolized in the pentose phosphate pathway. The main problem arises from the amount of lactate reintroduced (only 22%) into the TCA cycle as substrate.[123] Based on these findings, Dietmair et al.[124] examined the influence of RP production in HEK-293F cells using time-resolved transcriptome, metabolome, and fluxome data. Despite similar growth rates of producing and non-producing clones, glucose consumption and lactate production were 20% higher in non-producer cells, whereas AA transporters were upregulated in producer cells. Supported by differential expression analysis, the authors concluded that protein assembly and -folding might be bottlenecks during RP production instead of metabolic limitations. Nevertheless, a more efficient strategy to utilize lactate as a carbon source remained unresolved until recently: Martinez-Monge et al.[49] investigated the influence of pH-controlled and -uncontrolled growth phases based on metabolic changes of HEK-293 cells. Besides the two main growth phases, consuming glucose first, followed by lactate consumption, a third phase was observed where glucose and lactate were co-consumed under non-pH-controlled conditions. During this phase, a 3-fold increase in the growth rate was recorded. Despite these promising results, pH-control of good manufacturing practice (good manufacturing practice (GMP))-compliant bioprocesses is one of the cornerstones, especially with a focus on batch-to-batch consistency and reproducibility.

Furthermore, it must be considered that engineering metabolic enzymes is often more challenging. On the other hand, it is still difficult and laborious to diminish gene expression or enzymatic activity, though possible.[125] One opportunity to expand the number of potential metabolic engineering targets is the implementation of HEK-specific GSMMs. Surprisingly, the most recent and most extensive model for HEK-293 cells was curated by manual reduction of the Recon2.0,[126] and consists of only 357 reactions.[47] Although reduced models were recently shown to capture core metabolic functions more precisely,[32] the number of potential targets for engineering is limited.

### 3.2.2 | Adeno-associated virus production in HEK-293 cells

Recently, reviews have been published describing the production cycles of rAAVs, AdVs, and LVs in general,[3] and in HEK-293 cells, specifically.[8]

rAAV has been one of the most prominent vectors for gene delivery in recent decades. There is a large number of studies that describe various strategies to increase productivity. Zhao et al.[14] chose a DoE approach to identify the best parameter setting during the production phase in HEK-293T cells for multiple serotypes, and lead to a maximum of $> 10^{14}$ VG/mL. However, DoE-approaches are often laborious and are limited to specific process parameters. In contrast, the identification of engineering targets is so far based on only one or a few omics datasets under a limited number of conditions.[8] Nevertheless, recently published studies based on omics techniques could form a starting point for omics-based process optimization on a cellular level.

Strasser et al.[102] used FreeStyle 293-F as host cells for rAAV 5 production and examined the metabolic differences based on proteomics data. The pathway analysis revealed changes in endosomal-lysosomal protein levels, translation machinery, and energy metabolism between transfected and not-transfected cells. Based on their investigation, chloroquinone was added to alter endocytosis, which led to a titer improvement of 35%. In addition, by generating double knockouts of STAT1, and BAX genes, which are involved in the immune response to viral infection, titers for rAAV production could be increased by 80% compared to a wild-type strain.[13] Despite different study designs, both omics-data-based and hypothesis-driven approaches significantly improve rAAV titers. However, these studies only report improvements based on a limited number of genetic targets and neglect computational optimization approaches.

In contrast, Nguyen et al.[6] predicted improvements of rAAV 5 production by temporally staggering transfection of the plasmids due to different expressions of the corresponding viral genes. The established mechanistic model of plasmid uptake and viral particle production allows fundamental insights into time-staggered bottlenecks of rAAV production in HEK-293 cells. Though these results remain to be verified, it could potentially overcome the premature depletion of capsid proteins.[127,128] Furthermore, secretion rates of rAAVs remain dependent on serotype. Certain rAAVs interact with host cell proteins and are not secreted, which complicates purification and bears the risk of potential contaminations.[129] However, this issue needs to be addressed by alterations on a molecular level of viral protein since the cellular interaction proteins are essential for host cells.

### 3.2.3 | Adenovirus production in HEK-293 cells

Although AdV has been extensively used for gene delivery in GT, or vaccination,[103] only a few studies investigated the production process in HEK-293 cells: Negrete et al.[130] adapted a HEK-293 substrain to grow in suspension and compared adenoviral vector titers in dependency of different multiplicity of infection (multiplicity of infection (MOI)), with a maximum productivity of $3 \times 10^{10}$ plagueforming units (PFG/mL) at a MOI of 10. Joe et al.[103] further decreased the MOI, resulting in $2 \times 10^{11}$ VG/mL at a 1000 L scale for the production of a AdV-based Covid-19 vaccine. Using a DoE approach, Chen et al.[131] could identify optimized parameters for the production of AdV in a perfusion reactor.

A model for MFA studies, covering 40 reactions, was generated by Henry et al.[48] based on 293SF-3F6 cells. By comparing six different infection states, including different cell densities and perfusion rates, it was possible to identify the best settings for ATP and AdV production. To the best of our knowledge, there are currently no studies describing process optimization for AdV production in HEK-293 cells based on omics data. However, since AdVs are reliable and safe vectors for gene delivery and were used as vaccines during the recent Covid-19 pandemic, more effort should be put into optimizing these bioprocesses.

### 3.2.4 | Retrovirus production in HEK-293 cells

Retrovirus-like particles are of particular interest in GT due to their ability to stably integrate target genes into patients' genomes. However, these vectors can potentially cause tumors, as was shown in several clinical trials, using MvaV.[132] Therefore, despite their limited application in vivo, these vectors are commonly used for ex vivo GT applications, such as the generation of CAR-T cells.[133] Besides gamma-retrovirus (gamma-retrovirus (gRV)), LVs became more important during the last decade, particularly HIV-1-based vectors.[134]

LV production processes and optimization strategies have been established using transient transfection,[135] stable expression,[104,136] flow electroporation,[137] various media supplements,[138] or DoE approaches[139] to reach maximum titers of $4.5 \times 10^9$ VG/mL.[104] An important milestone was set by Lavado-Garcia et al.[105] by investigating metabolic changes between transfected, mock-transfected, and non-transfected HEK-293SF-3F6 cells in suspension during HIV-1 production based on quantitative, labeled time-resolved proteomes. Besides changes caused by virus production, cell density effects were examined to find the optimal cell density for transfection. By considering both growth and production phases, it was possible to identify more engineering targets to overcome bottlenecks such as cell density effects. Based on these data transfection efficiency was increased by overexpression of UGCG by 17%, and VLP production was raised 3.3-fold by overexpressing NEDD4L, which is involved in endosomal sorting.[140] Nevertheless, it would have been possible to generate time-resolved ecGSMMs for each state by adding measurements for usually measured metabolites for bioprocess control, such as AA, glucose, lactate, and comparatively inexpensive ammonia. Combining proteomics and fluxomics data would allow the prediction of catalytic rates of enzymes under different conditions.[63]

## 3.3 | Per.C6

An adherent Per.C6 cell line, derived from human fetal retinoblasts, was adapted to grow in suspension and under serum-free conditions. Additionally, the adenoviral E1-gene was inserted to produce E1-deficient recombinant AdV-like particles.[141] Metabolic behavior of Per.C6 cells was studied after infection with AdVs and media exchange,[142] following numerous passages of the original cell line,[15]

under various pH conditions,[143] and under low levels of glucose and glutamine.[144] Based on these studies, the authors identified a pH of 7.3 as the optimum for AdV production in Per.C6 cells. Additionally, they found that cell growth was reduced after media exchange and production decreased. In the same study, the authors described the increased consumption of glucose, oxygen, and certain aliphatic AAs.

In contrast to other producing cells, such as HEK-293 cells, Per.C6 cells produce high titers at different cell densities, although the cell-specific productivity is lower. Based on that, AdV production could be increased by adapting the cell line to suspension culture. In contrast to the original cell line, the high-producer cells grew after infection with AdVs, which increased the available amount of cells for infection.[15] Currently, the highest titer of AdV particles in Per.C6 cells was reported by Vellinga, et al.[106] Despite the availability of numerous exchange rate data under different conditions, no metabolic network or omics-data-based optimization studies for Per.C6 cells have been published. In recent years, the platform was used to produce an AdV-based Covid-19 vaccine. Besides AdV production, Per.C6 cells were used to manufacture poliovirus-based vaccines.[145] Although these candidates have not yet been used for GT, they offer potential alternatives for future gene delivery vectors.

### 3.4 | Vero

Vero cells are derived from *C. sabaeus* and have been used as a model to investigate metabolic influence after viral infection and as production hosts for antigen-presenting VLPs, or deliver genes, as recently summarized.[146] Despite lower growth rates and cell densities in manufacturing processes compared to other cell lines, Vero cells can be considered for high-yield viral vector production.

Recently, the genomes of substrains were sequenced, assembled, and annotated[95,107] as a basis for strain design. Furthermore, differences between adherent and suspended Vero cells on transcriptome level were investigated and revealed differentially expressed pathways as potential engineering targets.[147] As in HEK-293 and MDCK cells, knocking out STAT1 increased titers of Dengue 2 virus vaccines 30-fold.[13] Metabolic analyses during cell culture processes lead to the identification of essential metabolites for growth,[148] under various conditions, including batch,[149] or fed-batch.[150] In the latter study, three kinetic models were established for flux analysis by measuring specific uptake and secretion rates to enable control strategies for fed-batch processes. However, further omics-based, time-resolved studies would allow more detailed analyses and may address problems such as low cell densities. Although Vero cells are mainly used for antigen-presenting VLPs, they can be used as a production platform of various VLPs for gene delivery, including AdVs,[108] rAAVs,[109] or vesicular stomatitis virus[151] for example.

### 3.5 | MDCK

MDCK cells, derived from kidney cells from *C. familiaris*, were extracted and adapted for cultivation.[152] By integrating the human SIAT7E gene, the cell line could be converted to grow in suspension and under serum-free conditions for influenza vaccine production.[153] Alterations between adherent and suspension cultures were investigated based on label-free proteomes and metabolomes in combination with enzyme activity assays.[154] The generated data were integrated into a MFA model, which was generated based on different media compositions.[155] Besides differentially expressed intracellular proteins, alterations in catalytic activities were observed for enzymes, including Pirin, CAV1, or AMPK, which are known to influence cellular metabolism directly. Furthermore, glucose uptake rate and lactate secretion rate decreased by 20% and 40%, respectively, after adaptation to suspension. Consequently, growth rates in suspension culture were reduced by up to 30%.[154] A proteome-based comparative study of adherent and suspension cell lines reveiled differently expressed proteins involved in intercellular adhesion, junction, and adhesion to the extracellular matrix.[156] These data may form the basis for enzyme-constrained MDCK cell modeling on genome scale as soon as a reference model is available. However, experimental biomass composition should be measured additionally to improve predictions.

Besides metabolic flux profiling during growth and influenza production,[157] MDCK cells are mainly used for AdV production, such as canine AdV vectors type 2.[111,158] Carinhas et al.[159] performed a label-based metabolic flux analysis to observe changes in E1-transformed MDCK cells during growth, and production phases, by comparing infected and mock-infected cells. After infection with AdV, central metabolic fluxes were increased. Based on that, several strategies for titer improvement were suggested, such as supplementation of lipids, cholesterol, or nucleic acid precursors. However, a more extensive set of engineering targets could be identified by measuring changes in transcriptome or proteome levels.

Currently, there is a need for GSMM established for MDCK cells, and genome, transcriptome, or proteome-based studies investigating VLP production for GT. However, Tao et al.[160] compared whole transcriptomic profiles after infection with different influenza strains, but not with a focus on productivity, and Capellini et al.[161] investigated differential expression of 41 genes by mimicking bacterial infection, and cadmium exposure. Although the cell line-specific genome has not been sequenced so far, a reference genome from the original species is available and could serve as a starting point for developing a GSMM.[110] However, the scientific community should consider closing these knowledge gaps further to increase the production of influenza and AdV vectors.

### 3.6 | AGE1.CR.pIX

AGE1.CR cells are derived from the fetal retina of *C. moschata*. The cell line was adapted to grow in suspension and to produce different VLPs, including MvaV.[92] After adaption to suspension growth under serum-free conditions, a substrain was developed by integrating the adenoviral *pIX* gene to increase the capsid stability during production.[162] In a detailed study, these two cell lines were ana-

lyzed regarding growth, metabolism, and VLP production from the same research group.[163] Lohr et al.[164] observed that these clones behave the same way as other mammalian cell lines regarding the Warburg effect. However, AGE1.CR cell lines showed a higher independency from glutamine compared to mammalian cell lines, which was also confirmed by MFA simulation. In recent years, these cell lines were optimized for high-density cultivation, using perfusion or continuous processes for MvaV production.[91,92,165,166] However, neither genomes nor transcriptomes or proteomes are currently published to our knowledge to generate a cell line specific GSMM. Nevertheless, a sequenced genome of *C. moschata* was provided, which may serve as a starting point to generate a reference GSMM (GenBank Assembly Accession: *GCA*_018104995.1).

### 3.7 | Sf9

Sf9 cell lines, derived from ovarian cells of *S. frugiperda*, have gained much attention during the last century due to the well-established baculovirus-system for stable and high-yield production of RPs and VLPs.[167,168] The cell line was adapted to grow in suspension culture, under serum-free conditions, and on a large scale.[169] Recently, the genome was sequenced and annotated.[112] Quantitative proteomics studies of Sf9 cells during growth and 6 hours post-infection with baculoviruses shed light on cellular behavior during growth and early-stage infection. Differential expression analysis revealed lower expression of viral proteins at higher cell densities and differentially expressed enzymes involved in metabolism, translation, protein processing, and signaling.[114] Currently, the highest titer of $2.4 \times 10^{12}$ VG/mL rAAV particles produced in Sf9 cells was reported, in addition to a decreased contamination rate with baculovirus genes.[113]

Moreover, the metabolic potential was examined during growth,[170] and production of VLPs.[171] Both studies compared the metabolic behavior of Sf9 cells to other insect-derived cell lines. They mentioned a lower lactate and ammonia production and a higher VLP production in Sf9 cells. For further specification and simulation, a small-scale metabolic model for MFA analyses was developed, containing 51 reactions.[99]

One of the most critical issues when using different platforms for VLP production was reported in different studies recently.[172,173] By comparing rAAV production in HEK-293 and Sf9 cells, it was identified that posttranslational modifications of capsids and methylation patterns of VGs differed between the two platforms.[172] Consequently, the change of production platform should be done carefully since influences in biopotency and safety profiles may be likely due to these changes.

Most of the mentioned studies end with suggestions for genetic engineering but need to verify their findings. Furthermore, there is currently no GSMM established, despite transcriptomes and proteomes available. Combined with the sequenced genome, these resources may form a sufficient starting point to generate a draft-GSMM.

## 4 | CONCLUSION

VLP-based GT emerged as one of the most promising approaches to cure genetic diseases. However, numerous challenges remain, even for FDA-approved drugs, including safety profiles, or titers. Despite the availability of high-throughput techniques to generate multi-omics datasets of host cells, optimization approaches still rely mainly on long-term passaging,[15] hypothesis-driven genetic engineering,[8] or DoE.[14] In recent years, a few studies initiated omics-based strain design approaches for VLP production in GT.[17,102] Currently, HEK-293 cell lines are the most commonly used platforms for VLP production, though other cell lines surpass HEK-293 cells in growth rates,[15] auxotrophy,[164] or productivity.[106] In turn, available global GSMMs[26,27] of human cells are advantageous for CSD approaches, while GSMMs need to be established first for non-human cell lines used in GT. The lack of global GSMM reconstructions for these cell lines is surprising, even though tremendous advances based on GSMMs have been reported in disease modeling[73] or bioprocess optimization.[23] Nevertheless, numerous challenges must be addressed when implementing GSMMs for these cell lines, including experimental and computational approaches.

### AUTHOR CONTRIBUTIONS

Leopold Zehetner: Conceptualization-equal, formal analysis-equal, investigation-equal, visualization-equal, writing–original draft-equal, writing–review & editing-equal. Diana Széliová: Conceptualization-equal, investigation-equal, visualization-equal, writing–review & editing-equal. Barbara Kraus: Writing–review & editing-equal. Michael Graninger: Writing–review & editing-equal. Jürgen Zanghellini: Conceptualization-equal, supervision-equal, writing–review & editing-equal. Juan Antonio Hernandez Bort: Conceptualization-equal, funding acquisition-equal, project administration-equal, supervision-equal, writing–review & editing-equal.

### ACKNOWLEDGMENTS

Leopold Zehetner, Barbara Kraus, Michael Graninger and Juan A. Hernandez Bort are all employees of Baxalta Innovations GmbH, a part of Takeda companies, which is involved in the development of gene therapy products and may be owner of stock options. Diana Széliová and Jürgen Zanghellini are full employees of University of Vienna. The presented work was partially funded by Baxalta Innovations GmbH.

### CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

### DATA AVAILABILITY STATEMENT

Data sharing not applicable–no new data generated.

### ORCID

*Leopold Zehetner* https://orcid.org/0009-0003-2677-3685
*Diana Széliová* https://orcid.org/0000-0002-9885-9758

*Jürgen Zanghellini* https://orcid.org/0000-0002-1964-2455
*Juan A. Hernandez Bort* https://orcid.org/0000-0003-2348-4864

## REFERENCES

1. Friedmann, T. (1992). A brief history of gene therapy. *Nature Genetics*, *2*(2), 93–98.
2. Gonçalves GAR, P. R. (2017). Gene therapy: Advances, challenges and perspectives. *Einstein (Sao Paulo)*, *15*, 369–375.
3. Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W., & Gao, G. (2021). Viral vector platforms within the gene therapy landscape. *Signal Transduction and Targeted Therapy*, *6*(1), 1–24.
4. Sung, Y. K., & Kim, S. (2019). Recent advances in the development of gene delivery systems. *Biomaterials Research*, *23*(1), 1–7.
5. Jiang, L., Vader, P., & Schiffelers, R. (2017). Extracellular vesicles for nucleic acid delivery: Progress and prospects for safe RNA-based gene therapy. *Gene Therapy*, *24*(3), 157–166.
6. Nguyen, T. N., Sha, S., Hong, M. S., Maloney, A. J., Barone, P. W., Neufeld, C., Wolfrum, J., Springs, S. L., Sinskey, A. J., & Braatz, R. D. (2021). Mechanistic model for production of recombinant adeno-associated virus via triple transfection of HEK293 cells. *Molecular Therapy-Methods & Clinical Development*, *21*, 642–655.
7. Martin, C., & Zatina, J. (2021). Utilizing a media panel rapidly accelerates media optimization for AAV manufacturers. *Cell and Gene Therapy Insights*, *7*(9), 1065–1072.
8. Abaandou, L., Quan, D., & Shiloach, J. (2021). Affecting HEK293 cell growth and production performance by modifying the expression of specific genes. *Cells*, *10*(7), 1667.
9. Pérez-Fernández, B. A., Cossio-Diaz, F. -dJ., Boggiano, T., Leon, K., & Mulet, R. (2021). In-silico media optimization for continuous cultures using genome scale metabolic networks: The case of CHO-K1. *Biotechnology and Bioengineering*, *118*(5), 1884–1897.
10. Hong, J. K., Choi, D. H., Park, S. Y., Silberberg, Y. R., Shozui, F., Nakamura, E., Kayahara, T., & Lee, D. Y. (2022). Data-driven and model-guided systematic framework for media development in CHO cell culture. *Metabolic Engineering*, *73*, 114–123.
11. Girdhar, K., Powis, A., Raisingani, A., Chrudinová, M., Huang, R., Tran, T., Sevgi, K., Dogus, D. Y., & Altindis, E. (2021). Viruses and Metabolism: The Effects of Viral Infections and Viral Insulins on Host Metabolism. *Annual Review of Virology*, *8*, 373–391.
12. Sumbria, D., Berber, E., Mathayan, M., & Rouse, B. T. (2021). Virus infections and host metabolism–can we manage the interactions? *Frontiers in Immunology*, *11*, 594963.
13. Turner-Gillies, E., Shapiro, B., & Tian, F. (2021). Generation of cell lines capable of producing high-titer viral stocks for use in vaccine manufacture and gene therapy. *ATCC Application Note AP-082021*, *1*, 1–8.
14. Zhao, H., Lee, K. J., Daris, M., Lin, Y., Wolfe, T., Sheng, J., Plewa, C., Wang, S., & Meisen, W. H. (2020). Creation of a high-yield AAV vector production platform in suspension cells using a design-of-experiment approach. *Molecular Therapy-Methods & Clinical Development*, *18*, 312–320.
15. Berdichevsky, M., Gentile, M. P., Hughes, B., Meis, P., Peltier, J., Blumentals, I., Aunins, J., & Altaras, N. E. (2008). Establishment of higher passage PER. C6 cells for adenovirus manufacture. *Biotechnology Progress*, *24*(1), 158–165.
16. Dotti, S., Lombardo, T., Villa, R., Cacciamali, A., Zanotti, C., Andreani, N. A., Cinotti, S., & Ferrari, M. (2017). Transformation and tumorigenicity testing of simian cell lines and evaluation of poliovirus replication. *Plos One*, *12*(1), e0169391.
17. Strobel, B., Zuckschwerdt, K., Zimmermann, G., Mayer, C., Eytner, R., Rechtsteiner, P., Kreuz, S., & Lamla, T. (2019). Standardized, scalable, and timely flexible adeno-associated virus vector production using frozen high-density HEK-293 cell stocks and CELLdiscs. *Human Gene Therapy Methods*, *30*(1), 23–33.
18. Garrison Jr, L. P., Jiao, B., & Dabbous, O. (2021). Gene therapy may not be as expensive as people think: challenges in assessing the value of single and short-term therapies. *Journal of Managed Care & Specialty Pharmacy*, *27*(5), 674–681.
19. Agren, R., Otero, J. M., & Nielsen, J. (2013). Genome-scale modeling enables metabolic engineering of Saccharomyces cerevisiae for succinic acid production. *Journal of Industrial Microbiology and Biotechnology*, *40*(7), 735–747.
20. Brunk, E., George, K. W., Alonso-Gutierrez, J., Brunk, E., George, K. W., Alonso-Gutierrez, J., Thompson, M., Baidoo, E., Wang, G., Petzold, C. J., McCloskey, D., Monk, J., Yang, L., O'Brien, E. J., Batth, T. S., Martin, H. G., Feist, A., Adams, P. D., Keasling, J. D., & Lee, T. S. (2016). Characterizing strain variation in engineered E. coli using a multi-omics-based workflow. *Cell Systems*, *2*(5), 335–346.
21. Edwards, J. S., & Palsson, B. O. (1999). Systems properties of the Haemophilus influenzae Rd metabolic genotype. *Journal of Biological Chemistry*, *274*(25), 17410–17416.
22. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., & Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biology*, *20*(1), 1–18.
23. Huang, Z., Xu, J., Yongky, A., Morrisa, C. S., Polancoa, A. L., Reilyc, M., Borysb, M. C., Lib, Z. J., & Yoon, S. (2020). CHO cell productivity improvement by genome-scale modeling and pathway analysis: Application to feed supplements. *Biochemical Engineering Journal*, *160*, 107638.
24. Thiele, I., & Palsson, BØ. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, *5*(1), 93–121.
25. Faria, J. P., Rocha, M., Rocha, I., & Henry, C. S. (2018). Methods for automated genome-scale metabolic model reconstruction. *Biochemical Society Transactions*, *46*(4), 931–936.
26. Robinson, J. L., Kocabaş, P., Wang, H., Cholley, P. E., Cook, D., Nilsson, A., Anton, M., Ferreira, R., Domenzain, I., Billa, V., Limeta, A., Hedin, A., Gustafsson, J., Kerkhoven, E. J., Svensson, L. T., Palsson, B. O., Mardinoglu, A., Hansson, L., Uhlén, M., & Nielsen, J. (2020). An atlas of human metabolism. *Science Signaling*, *13*(624), eaaz1482.
27. Brunk, E., Sahoo, S., Zielinski, D. C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G. A., Aurich, M. K., Prlic, A., Sastry, A., Danielsdottir, A. D., Heinken, A., Noronha, A., Rose, P. W., Burley, S. K., Fleming, R. M. T., Nielsen, J., … Palsson, B. O. (2018). Recon 3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology*, *36*(3), 272–281.
28. Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C. A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V. S., Kyriakopoulos, S., Jiménez, N. E., Zielinski, D. C., Quek, L. E., Wulff, T., Arnsdorf, J., Li, S., Lee, J. S., … Lewis, N. E. (2016). A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism. *Cell Systems*, *3*(5), 434–443.
29. Wang, H., Robinson, J. L., Kocabas, P., Gustafsson, J., Anton, M., Cholley, P. E., Huang, S., Gobom, J., Svensson, T., Uhlen, M., Zetterberg, H., & Nielsen, J. (2021). Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proceedings of the National Academy of Sciences*, *118*(30), e2102344118.
30. Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A. M., & Palsson, B. O. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. *Nature Biotechnology*, *35*(10), 904–908.
31. Strain, B., Morrissey, J., Antonakoudis, A., & Kontoravdi, C. (2023). Genome-scale models as a vehicle for knowledge transfer from microbial to mammalian cell systems. *Computational and Structural Biotechnology Journal*, *21*, 1543–1549.
32. Morrissey, J., Strain, B., Antonakoudis, A., & Kontoravdi, C. (2023). How reliable are CHO cell Genome-scale metabolic models? A comprehensive analysis. *Biotechnology and Bionengineering*.

33. Moškon, M., & Režen, T. (2023). Context-specific genome-scale metabolic modelling and its application to the analysis of COVID-19 metabolic signatures. *Metabolites*, *13*(1), 126.
34. Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., & Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Computational Biology*, *8*(5), e1002518.
35. Zur, H., Ruppin, E., & Shlomi, T. (2010). iMAT: An integrative metabolic analysis tool. *Bioinformatics*, *26*(24), 3140–3142.
36. Becker, S. A., & Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology*, *4*(5), e1000082.
37. Jerby, L., Shlomi, T., & Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism. *Molecular Systems Biology*, *6*(1), 401.
38. Vlassis, N., Pacheco, M. P., & Sauter, T. (2014). Fast reconstruction of compact context-specific metabolic network models. *PLoS Computational Biology*, *10*(1), e1003424.
39. Schultz, A., & Qutub, A. A. (2016). Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Computational Biology*, *12*(3), e1004808.
40. Vieira, V., Ferreira, J., & Rocha, M. (2022). A pipeline for the reconstruction and evaluation of context-specific human metabolic models at a large-scale. *PLoS Computational Biology*, *18*(6), e1009294.
41. Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D. C., & Lewis, N. E. (2017). A systematic evaluation of methods for tailoring genome-scale metabolic models. *Cell Systems*, *4*(3), 318–329.
42. Richelle, A., Chiang, A. W., Kuo, C. C., & Lewis, N. E. (2019). Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLoS Computational Biology*, *15*(4), e1006867.
43. Walakira, A., Rozman, D., Režen, T., Mraz, M., & Moškon, M. (2021). Guided extraction of genome-scale metabolic models for the integration and analysis of omics data. *Computational and Structural Biotechnology Journal*, *19*, 3521–3530.
44. Daskalaki, E., Pillon, N. J., Krook, A., Wheelock, C. E., & Checa, A. (2018). The influence of culture media upon observed cell secretome metabolite profiles: The balance between cell viability and data interpretability. *Analytica Chimica Acta*, *1037*, 338–350.
45. Yao, G., Aron, K., Borys, M., Li, Z., Pendse, G., & Lee, K. (2021). A metabolomics approach to increasing Chinese hamster ovary (CHO) cell productivity. *Metabolites*, *11*(12), 823.
46. Széliová, D., Ruckerbauer, D. E., Galleguillos, S. N., Petersen, L. B., Natter, K., Hanscho, M., Troyer, C., Causon, T., Schoeny, H., Christensen, H. B., Lee, D. Y., Lewis, N. E., Koellensperger, G., Hann, S., Nielsen, L. K., Borth, N., & Zanghellini, J. (2020). What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metabolic Engineering*, *61*, 288–300.
47. Quek, L. E., Dietmair, S., Hanscho, M., Martínez, V. S., Borth, N., & Nielsen, L. K. (2014). Reducing Recon 2 for steady-state flux analysis of HEK cell culture. *Journal of Biotechnology*, *184*, 172–178.
48. Henry, O., Perrier, M., & Kamen, A. (2005). Metabolic flux analysis of HEK-293 cells in perfusion cultures for the production of adenoviral vectors. *Metabolic Engineering*, *7*(5-6), 467–476.
49. Martínez-Monge, I., Albiol, J., Lecina, M., Liste-Calleja, L., Miret, J., Solá, C., & Cairó, J. J. (2019). Metabolic flux balance analysis during lactate and glucose concomitant consumption in HEK293 cell cultures. *Biotechnology and Bioengineering*, *116*(2), 388–404.
50. Schwaiger, M., Schoeny, H., El Abieadm, Y., Hermann, G., Rampler, E., & Koellensperger, G. (2019). Merging metabolomics and lipidomics into one analytical run. *Analyst*, *144*(1), 220–229.
51. Orth, J. D., Thiele, I., & Palsson, BØ. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248.
52. Yeo, H. C., Hong, J., Lakshmanan, M., & Lee, D. Y. (2020). Enzyme capacity-based genome scale modelling of CHO cells. *Metabolic Engineering*, *60*, 138–147.
53. Oftadeh, O., Salvy, P., Masid, M., Curvat, M., Miskovic, L., & Hatzimanikatis, V. (2021). A genome-scale metabolic model of Saccharomyces cerevisiae that integrates expression constraints and reaction thermodynamics. *Nature Communications*, *12*(1), 1–10.
54. Ramos, J. R., Oliveira, G. P., Dumas, P., & Oliveira, R. (2022). Genome-scale modeling of Chinese hamster ovary cells by hybrid semi-parametric flux balance analysis. *Bioprocess and Biosystems Engineering*, *45*(11), 1–16.
55. Bekiaris, P. S., & Klamt, S. (2020). Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics*, *21*(1), 1–13.
56. Krumholz, E. W., & Libourel, I. G. (2017). Thermodynamic constraints improve metabolic networks. *Biophysical Journal*, *113*(3), 679–689.
57. O'brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., & Palsson, BØ. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, *9*(1), 693.
58. Yang, X., Mao, Z., Zhao, X., Wang, R., Zhang, P., Cai, J., Xue, C., & Ma, H. (2021). Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. *Metabolic Engineering*, *67*, 133–144.
59. Lu, J., Bi, X., Liu, Y., Lv, X., Li, J., Du, G., & Liu, L. (2022). In silico cell factory design driven by comprehensive genome-scale metabolic models: Development and challenges. *Systems Microbiology and Biomanufacturing*, 1–16.
60. Domenzain, I., Sánchez, B., Anton, M., Kerkhoven, E. J., Millán-Oropeza, A., Henry, C., Siewers, V., Morrissey, J. P., Sonnenschein, N., & Nielsen, J. (2022). Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nature Communications*, *13*(1), 1–13.
61. Mao, Z., Zhao, X., Yang, X., Zhang, P., Du, J., Yuan, Q., & Ma, H. (2022). ECMpy, a simplified workflow for constructing enzymatic constrained metabolic network model. *Biomolecules*, *12*(1), 65.
62. Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F., & Schomburg, D. (2002). BRENDA: A resource for enzyme data and metabolic information. *Trends in Biochemical Sciences*, *27*(1), 54–56.
63. Davidi, D., Noor, E., Liebermeister, W., Bar-Even, A., Flamholz, A., Tummler, K., Barenholz, U., Goldenfeld, M., Shlomi, T., & Milo, R. (2016). Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k cat measurements. *Proceedings of the National Academy of Sciences*, *113*(12), 3401–3406.
64. Heckmann, D., Lloyd, C. J., Mih, N., Ha, Y., Zielinski, D. C., Haiman, Z. B., Desouki, A. A., Lercher, M. J., & Palsson, B. O. (2018). Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nature Communications*, *9*(1), 1–10.
65. Li, F., Yuan, L., Lu, H., Li, G., Chen, Y., Engqvist, M. K. M., Kerkhoven, E. J., & Nielsen, J. (2022). Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis*, *5*, 662–672.
66. Besancon, M., Kratochvil, M., Kuate, C. A. F., Wilken, S. E., Besanạon, M., Kratochvíl, M., Foko Kuate, C. A., Trefois, C., Gu, W., & Ebenhöh, O. (2022). Interrogating the effect of enzyme kinetics on metabolism using differentiable constraint-based models. *Metabolic Engineering*, *74*, 72–82.
67. Goelzer, A., Muntel, J., Chubukov, V., Jules, M., Prestel, E., Nölker, R., Mariadassou, M., Aymerich, S., Hecker, M., Noirot, P., Becher, D., & Fromion, V. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, *32*, 232–243.
68. Bulović, A., Fischer, S., Dinh, M., Golib, F., Liebermeister, W., Poirier, C., Tournier, L., Klipp, E., Fromion, V., & Goelzer, A. (2019). Automated generation of bacterial resource allocation models. *Metabolic Engineering*, *55*, 12–22.

69. Antonakoudis, A., Barbosa, R., Kotidis, P., & Kontoravdi, C. (2020). The era of big data: Genome-scale modelling meets machine learning. *Computational and Structural Biotechnology Journal*, *18*, 3287–3300.
70. Schneider, P., Bekiaris, P. S., Kamp, V. A., & Klamt, S. (2022). StrainDesign: A comprehensive Python package for computational design of metabolic networks. *Bioinformatics*, *38*(21), 4981–4983.
71. Pereira, V., Cruz, F., & Rocha, M. (2021). MEWpy: A computational strain optimization workbench in Python. *Bioinformatics*, *37*(16), 2494–2496.
72. Huang, D., Wen, J., Wang, G., Yu, G., Jia, X., & Chen, Y. (2012). In silico aided metabolic engineering of Streptomyces roseosporus for daptomycin yield improvement. *Applied Microbiology and Biotechnology*, *94*(3), 637–649.
73. Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., & Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature Communications*, *5*(1), 1–11.
74. Cheng, K., Martin-Sancho, L., Pal, L. R., Pu, Y., Riva, L., Yin, X., Sinha, S., Nair, N. U., Chanda, S. K., & Ruppin, E. (2021). Genome-scale metabolic modeling reveals SARS-CoV-2-induced metabolic changes and antiviral targets. *Molecular Systems Biology*, *17*(11), e10260.
75. Sha, S., Huang, Z., Wang, Z., & Yoon, S. (2018). Mechanistic modeling and applications for CHO cell culture development and production. *Current Opinion in Chemical Engineering*, *22*, 54–61.
76. Rejc, Ž., Magdevska, L., Tršelič, T., Osolin, T., Vodopivec, R., Mraz, J., Pavliha, E., Zimic, N., Cvitanovic, T., Rozman, D., Mosǩon, M., & Mraz, M. (2017). Computational modelling of genome-scale metabolic networks and its application to CHO cell cultures. *Computers in Biology and Medicine*, *88*, 150–160.
77. Selvarasu, S., Ho, Y. S., Chong, W. P., Wong, N. S., Yusufi, F. N., Lee, Y. Y., Yap, M. G., & 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.
78. Fouladiha, H., Marashi, S. A., Torkashvand, F., Mahboudi, F., Lewis, N. E., & Vaziri, B. (2020). A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. *Bioprocess and Biosystems Engineering*, *43*(8), 1381–1389.
79. Calmels, C., McCann, A., Malphettes, L., & Andersen, M. R. (2019). Application of a curated genome-scale metabolic model of CHO DG44 to an industrial fed-batch process. *Metabolic Engineering*, *51*, 9–19.
80. Huang, Z., & Yoon, S. (2020). Integration of time-series transcriptomic data with genome-scale CHO metabolic models for mAb engineering. *Processes*, *8*(3), 331.
81. Schinn, S. M., Morrison, C., Wei, W., Zhang, L., & Lewis, N. E. (2021). Systematic evaluation of parameters for genome-scale metabolic models of cultured mammalian cells. *Metabolic Engineering*, *66*, 21–30.
82. Gutierrez, J. M., Feizi, A., Li, S., Kallehauge, T. B., Hefzi, H., Grav, L. M., Ley, D., Baycin Hizal, D., Betenbaugh, M. J., Voldborg, B., Faustrup Kildegaard, H., Min Lee, G., Palsson, B. O., Nielsen, J., & Lewis, N. E. (2020). Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nature Communications*, *11*(1), 1–10.
83. Kol, S., Ley, D., Wulff, T., Decker, M., Arnsdorf, J., Schoffelen, S., Hansen, A. H., Jensen, T. L., Gutierrez, J. M., Chiang, A. W. T., Masson, H. O., Palsson, B. O., Voldborg, B. G., Pedersen, L. E., Kildegaard, H. F., Lee, G. M., & Lewis, N. E. (2020). Multiplex secretome engineering enhances recombinant protein production and purity. *Nature Communications*, *11*(1), 1–10.
84. Val, d. I. J., Kyriakopoulos, S., Albrecht, S., Stockmann, H., Rudd, P. M., Polizzi, K. M., & Kontoravdi, C. (2021). CHOmpact: a reduced metabolic model of Chinese hamster ovary cells with enhanced interpretability. *bioRxiv*.
85. Lund, A. M., Kaas, C. S., Brandl, J., Pedersen, L. E., Kildegaard, H. F., Kristensen, C., & Andersen, M. R. (2017). Network reconstruction of the mouse secretory pathway applied on CHO cell transcriptome data. *BMC Systems Biology*, *11*(1), 1–17.
86. Joiner, J., Huang, Z., McHugh, K., Stebbins, M., Aron, K., Borys, M., & Khetan, A. (2022). Process modeling of recombinant adeno-associated virus production in HEK293 cells. *Current Opinion in Chemical Engineering*, *36*, 100823.
87. Helgers, H., Schmidt, A., & Strube, J. (2022). Towards Autonomous Process Control–Digital Twin for CHO Cell-Based Antibody Manufacturing Using a Dynamic Metabolic Model. *Processes*, *10*(2), 316.
88. Amer, B., & Baidoo, E. E. (2021). Omics-driven biotechnology for industrial applications. *Frontiers in Bioengineering and Biotechnology*, *9*, 613307.
89. Ketter, E., & Randall, G. (2019). Virus impact on lipids and membranes. *Annual Review of Virology*, *6*, 319–340.
90. Ivanova, P. T., Myers, D. S., Milne, S. B., McClaren, J. L., Thomas, P. G., & Brown, H. A. (2015). Lipid Composition of the viral envelope of three strains of influenza virus – Not all viruses are created equal. *ACS Infectious Diseases*, *1*(9), 435–442.
91. Gränicher, G., Babakhani, M., Göbel, S., Jordan, I., Marichal-Gallardo, P., Genzel, Y., & Reichl, U. (2021). A high cell density perfusion process for Modified Vaccinia virus Ankara production: Process integration with inline DNA digestion and cost analysis. *Biotechnology and Bioengineering*, *118*(12), 4720–4734.
92. Gränicher, G., Tapia, F., Behrendt, I., Jordan, I., Genzel, Y., & Reichl, U. (2021). Production of Modified Vaccinia Ankara virus by intensified cell cultures: A comparison of platform technologies for viral vector production. *Biotechnology Journal*, *16*(1), 2000024.
93. Mouriaux, F., Zaniolo, K., Bergeron, M. A., Weidmann, C., De La Fouchardiáre, A., Fournier, F., Droit, A., Morcos, M. W., Landreville, S., & Guérin, S. L. (2016). Effects of long-term serial passaging on the characteristics and properties of cell lines derived from uveal melanoma primary tumors. *Investigative Ophthalmology & Visual Science*, *57*(13), 5288–5301.
94. Hughes, P., Marshall, D., Reid, Y., Parkes, H., & Gelber, C. (2007). The costs of using unauthenticated, over-passaged cell lines: How much more data do we need? *Biotechniques*, *43*(5), 575–586.
95. Sakuma, C., Sekizuka, T., Kuroda, M., Kasai, F., Saito, K., Ikeda, M., Yamaji, T., Osada, N., & Hanada, K. (2018). Novel endogenous simian retroviral integrations in Vero cells: Implications for quality control of a human vaccine cell substrate. *Scientific Reports*, *8*(1), 1–10.
96. Victoria, J. G., Wang, C., Jones, M. S., Jaing, C., McLoughlin, K., Gardner, S., & Delwart, E. L. (2010). Viral nucleic acids in live-attenuated vaccines: Detection of minority variants and an adventitious virus. *Journal of Virology*, *84*(12), 6033–6040.
97. Franzen, J., Georgomanolis, T., Selich, A., Kuo, C. C., Stöger, R., Brant, L., Mulabdic, M. S., Fernandez-Rebollo, E., Grezella, C., Ostrowska, A., Begemann, M., Nikolic, M., Rath, B., Ho, A. D., Rothe, M., Schambach, A., Papantonis, A., & Wagner, W. (2021). DNA methylation changes during long-term in vitro cell culture are caused by epigenetic drift. *Communications Biology*, *4*(1), 1–12.
98. Vayena, E., Chiappino-Pepe, A., MohammadiPeyhani, H., Vayena, E., Chiappino-Pepe, A., MohammadiPeyhani, H., Francioli, Y., Hadadi, N., Ataman, M., Hafner, J., Pavlou, S., & Hatzimanikatis, V. (2022). NICEgame: A workflow for annotating the knowledge gaps in metabolic reconstructions, using known and hypothetical reactions. *bioRxiv*.
99. Carinhas, N., Bernal, V., Teixeira, A. P., Carrondo, M. J., 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*(1), 1–13.

100. Malm, M., Saghaleyni, R., Lundqvist, M., Giudici, M., Chotteau, V., Field, R., Varley, P. G., Hatton, D., Grassi, L., Svensson, T., Nielsen, J., & Rockberg, J. (2020). Evolution from adherent to suspension: systems biology of HEK293 cell line development. *Scientific Reports*, *10*(1), 1–15.
101. Lin, Y. C., Boone, M., Meuris, L., Lemmens, I., van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., van de Peer, Y., Tavernier, J., & Callewaert, N. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications*, *5*(1), 4767.
102. Strasser, L., Boi, S., Guapo, F., Donohue, N., Barron, N., Rainbow-Fletcher, A., & Bones, J. (2021). Proteomic Landscape of Adeno-Associated Virus (AAV)-Producing HEK293 Cells. *International Journal of Molecular Sciences*, *22*(21), 11499.
103. Joe, C. C. D., Jiang, J., Linke, T., Li, Y., Fedosyuk, S., Gupta, G., Berg, A., Segireddy, R. R., Mainwaring, D., Joshi, A., Cashen, P., Rees, B., Chopra, N., Nestola, P., Humphreys, J., Davies, S., Smith, N., Bruce, S., Bormans, D., … Douglas, A. D. (2022). Manufacturinga chimpanzee adenovirus-vectored SARS-CoV-2 vaccine to meet global needs. *Biotechnology and Bioengineering*, *119*(1), 48.
104. Tran, M. Y., & Kamen, A. A. (2022). Production of lentiviral vectors using a HEK-293 producer cell line and advanced perfusion processing. *Frontiers in Bioengineering and Biotechnology*, *10*, 887716.
105. Lavado-García, J., Jorge, I., Cervera, L., Vazquez, J., & Gòdia, F. (2020). Multiplexed quantitative proteomic analysis of HEK293 provides insights into molecular changes associated with the cell density effect, transient transfection, and virus-like particle production. *Journal of Proteome Research*, *19*(3), 1085–1099.
106. Vellinga, J., Smith, J. P., Lipiec, A., Majhen, D., Lemckert, A., van Ooij, M., … & Havenga, M. (2014). Challenges in manufacturing adenoviral vectors for global vaccine product deployment. *Human gene therapy*, *25*(4), 318–327.
107. Sène, M. A., Kiesslich, S., Djambazian, H., Ragoussis, J., Xia, Y., & Kamen, A. A. (2021). Haplotype-resolved de novo assembly of the Vero cell line genome. *NPJ Vaccines*, *6*(1), 1–10.
108. Lee, D. K., Park, J., & Seo, D. W. (2020). Suspension culture of Vero cells for the production of adenovirus type 5. *Clinical and Experimental Vaccine Research*, *9*(1), 48–55.
109. Zhang, H., Xie, J., Xie, Q., Wilson, J. M., & Gao, G. (2009). Adenovirus–adeno-associated virus hybrid for large-scale recombinant adeno-associated virus production. *Human Gene Therapy*, *20*(9), 922–929.
110. Field, M. A., Rosen, B. D., Dudchenko, O., Chan, E. K. F., Minoche, A. E., Edwards, R. J., Barton, K., Lyons, R. J., Tuipulotu, D. E., Hayes, V. M., D Omer, A., Colaric, Z., Keilwagen, J., Skvortsova, K., Bogdanovic, O., Smith, M. A., Aiden, E. L., Smith, T. P. L., Zammit, R. A., & Ballard, J. W. O. (2020). Canfam_GSD: de novo chromosome-length genome assembly of the German Shepherd Dog (Canis lupus familiaris) using a combination of long reads, optical mapping, and Hi-C. *GigaScience*, *9*(4), giaa027.
111. Castro, R., Fernandes, P., Laske, T., Sousa, M. F., Genzel, Y., Scharfenberg, K., Alves, P. M., & Coroadinha, A. S. (2015). Production of canine adenovirus type 2 in serum-free suspension cultures of MDCK cells. *Applied Microbiology and Biotechnology*, *99*(17), 7059–7068.
112. Nandakumar, S., Ma, H., & Khan, A. S. (2017). Whole-genome sequence of the Spodoptera frugiperda Sf9 insect cell line. *Genome Announcements*, *5*(34), e00829–17.
113. Galibert, L., Jacob, A., Savy, A., Dickx, Y., Bonnin, D., Lecomte, C., Rivollet, L., Sanatine, P., Boutin Fontaine, M., Le Bec, C., & Merten, O. W. (2021). Monobac system–A single baculovirus for the production of rAAV. *Microorganisms*, *9*(9), 1799.
114. Carinhas, N., Robitaille, A. M., Moes, S., Carrondo, M. J., Jenoe, P., Oliveira, R., & Alves, P. M. (2011). Quantitative proteomics of Spodoptera frugiperda cells during growth and baculovirus infection. *PloS One*, *6*(10), e26444.
115. Tsao, Y. S., Condon, R., Schaefer, E., Lio, P., & Liu, Z. (2001). Development and improvement of a serum-free suspension process for the production of recombinant adenoviral vectors using HEK293 cells. *Cytotechnology*, *37*(3), 189–198.
116. Grieger, J. C., Soltys, S. M., & Samulski, R. J. (2016). Production of recombinant adeno-associated virus vectors using suspension HEK293 cells and continuous harvest of vector from the culture media for GMP FIX and FLT1 clinical vector. *Molecular Therapy*, *24*(2), 287–297.
117. Blessing, D., Vachey, G., Pythoud, C., Rey, M., Padrun, V., Wurm, F. M., Schneider, B. L., & Déglon, N. (2019). Scalable production of AAV vectors in orbitally shaken HEK293 cells. *Molecular Therapy-Methods & Clinical Development*, *13*, 14–26.
118. Malm, M., Kuo, C. C., Barzadd, M. M., Mebrahtu, A., Wistbacka, N., Razavi, R., Volk, A. L., Lundqvist, M., Kotol, D., Tegel, H., Hober, S., Edfors, F., Gräslund, T., Chotteau, V., Field, R., Varley, P. G., Roth, R. G., Lewis, N. E., Hatton, D., & Rockberg, J. (2022). Harnessing secretory pathway differences between HEK293 and CHO to rescue production of difficult to express proteins. *Metabolic Engineering*, *72*, 171–187.
119. Thorens, B., & Mueckler, M. (2010). Glucose transporters in the 21st Century. *American Journal of Physiology-Endocrinology and Metabolism*, *298*(2), E141–E145.
120. Haeseleer, F., Huang, J., Lebioda, L., Saari, J. C., & Palczewski, K. (1998). Molecular characterization of a novel short-chain dehydrogenase/reductase that reduces all-trans-retinal. *Journal of Biological Chemistry*, *273*(34), 21790–21799.
121. Hosios, A. M., & Vander Heiden, M. G. (2018). The redox requirements of proliferating mammalian cells. *Journal of Biological Chemistry*, *293*(20), 7490–7498.
122. Nadeau, I., Gilbert, P., Jacob, D., Perrier, M., & Kamen, A. (2002). Low-protein medium affects the 293SF central metabolism during growth and infection with adenovirus. *Biotechnology and Bioengineering*, *77*(1), 91–104.
123. Henry, O., Jolicoeur, M., & Kamen, A. (2011). Unraveling the metabolism of HEK-293 cells using lactate isotopomer analysis. *Bioprocess and Biosystems Engineering*, *34*(3), 263–273.
124. Dietmair, S., Hodson, M. P., Quek, L. E., Timmins, N. E., Gray, P., & Nielsen, L. K. (2012). A multi-omics analysis of recombinant protein production in Hek293 cells. *PLOS ONE*.
125. Noguchi, S., Ishikawa, H., Wakita, K., Matsuda, F., & Shimizu, H. (2020). Direct and quantitative analysis of altered metabolic flux distributions and cellular ATP production pathway in fumarate hydratase-diminished cells. *Scientific Reports*, *10*(1), 1–12.
126. Thiele, I., Swainston, N., Fleming, R. M., Hoppe, A., Sahoo, S., Aurich, M. K., Haraldsdottir, H., Mo, M. L., Rolfsson, O., Stobbe, M. D., Thorleifsson, S. G., Agren, R., Bölling, C., Bordel, S., Chavali, A. K., Dobson, P., Dunn, W. B., Endler, L., Hala, D., … Palsson, B. ØO. (2013). A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, *31*(5), 419–425.
127. Li, J., Samulski, R. J., & Xiao, X. (1997). Role for highly regulated rep gene expression in adeno-associated virus vector production. *Journal of Virology*, *71*(7), 5236–5243.
128. Mietzsch, M., Eddington, C., Jose, A., Hsi, J., Chipman, P., Henley, T., Choudhry, M., McKenna, R., & Agbandje-McKenna, M. (2021). Improved genome packaging efficiency of adeno-associated virus vectors using rep hybrids. *Journal of Virology*, *95*(19), e00773–21.
129. Nonnenmacher, M., & Weber, T. (2012). Intracellular transport of recombinant adeno-associated virus vectors. *Gene Therapy*, *19*(6), 649–658.

130. Negrete, A., Ling, T. C., & Lyddiatt, A. (2007). Production of adenoviral vectors and its recovery. *Process Biochemistry*, *42*(7), 1107–1113.
131. Chen, K. D., Wu, X. X., Yu, D. S., Ou, H. L., Li, Y. H., Zhou, Y. Q., & Li, L. J. (2018). Process optimization for the rapid production of adenoviral vectors for clinical trials in a disposable bioreactor system. *Applied Microbiology and Biotechnology*, *102*(15), 6469–6477.
132. Bulaklak, K., & Gersbach, C. A. (2020). The once and future gene therapy. *Nature Communications*, *11*(1), 1–4.
133. Bendjama, K., & Quemeneur, E. (2017). Modified vaccinia virus Ankara-based vaccines in the era of personalized immunotherapy of cancer. *Human Vaccines & Immunotherapeutics*, *13*(9), 1997–2003.
134. Tang, Q. L., Gu, L. X., Xu, Y., Liao, X. H., Zhou, Y., & Zhang, T. C. (2021). Establishing functional lentiviral vector production in a stirred bioreactor for CAR-T cell therapy. *Bioengineered*, *12*(1), 2095–2105.
135. Ansorge, S., Lanthier, S., Transfiguracion, J., Durocher, Y., Henry, O., & Kamen, A. (2009). Development of a scalable process for high-yield lentiviral vector production by transient transfection of HEK293 suspension cultures. *The Journal of Gene Medicine: A Cross-Disciplinary Journal for Research on the Science of Gene Transfer and its Clinical Applications*, *11*(10), 868–876.
136. Manceur, A. P., Kim, H., Misic, V., Andreev, N., Dorion-Thibaudeau, J., Lanthier, S., Bernier, A., Tremblay, S., Gélinas, A. M., Broussau, S., Gilbert, R., & Ansorge, S. (2017). Scalable lentiviral vector production using stable HEK293SF producer cell lines. *Human Gene Therapy Methods*, *28*(6), 330–339.
137. Witting, S. R., Li, L. H., Jasti, A., Allen, C., Cornetta, K., Brady, J., Shivakumar, R., & Peshwa, M. V. (2012). Efficient large volume lentiviral vector production using flow electroporation. *Human Gene Therapy*, *23*(2), 243–249.
138. Gélinas, J. F., Davies, L. A., Gill, D. R., & Hyde, S. C. (2017). Assessment of selected media supplements to improve F/HN lentiviral vector production yields. *Scientific Reports*, *7*(1), 1–12.
139. Labisch, J. J., Bollmann, F., Wolff, M. W., & Pflanz, K. (2021). A new simplified clarification approach for lentiviral vectors using diatomaceous earth improves throughput and safe handling. *Journal of Biotechnology*, *326*, 11–20.
140. Lavado-García, J., Díaz-Maneh, A., Canal-Paulí, N., Pérez-Rubio, P., Gòdia, F., & Cervera, L. (2021). Metabolic engineering of HEK293 cells to improve transient transfection and cell budding of HIV-1 virus-like particles. *Biotechnology and Bioengineering*, *118*(4), 1630–1644.
141. Lewis, J., Brown, E., & Duncan, P. (2006). Approaches to the release of a master cell bank of PER.C6 cells; a novel cell substrate for the manufacture of human vaccines. *Developments in Biologicals*, *123*, 165–176.
142. Maranga, L., Auniņš, J. G., & Zhou, W. (2005). Characterization of changes in PER. C6® cellular metabolism during growth and propagation of a replication-deficient adenovirus vector. *Biotechnology and bioengineering*, *90*(5), 645–655.
143. Xie, L., Pilbrough, W., Metallo, C., Zhong, T., Pikus, L., Leung, J., Aunins, J. G., & Zhou, W. (2002). Serum-free suspension cultivation of PER. C6® cells and recombinant adenovirus production under different pH conditions. *Biotechnology and Bioengineering*, *80*(5), 569–579.
144. Maranga, L., & Goochee, C. F. (2006). Metabolism of PER. C6TM cells cultivated under fed-batch conditions at low glucose and glutamine levels. *Biotechnology and Bioengineering*, *94*(1), 139–150.
145. Sanders, B. P., Edo-Matas, D., Custers, J. H., Koldijk, M. H., Klaren, V., Turk, M., Luitjens, A., Bakker, W. A., Uytdehaag, F., Goudsmit, J., Lewis, J. A., & Schuitemaker, H. (2013). PER. C6® cells as a serum-free suspension cell platform for the production of high titer poliovirus: A potential low cost of goods option for world supply of inactivated poliovirus vaccine. *Vaccine*, *31*(5), 850–856.
146. Kiesslich, S., & Kamen, A. A. (2020). Vero cell upstream bioprocess development for the production of viral vectors and vaccines. *Biotechnology Advances*, *44*, 107608.
147. Sène, M. A., Xia, Y., & Kamen, A. A. (2021). Comparative transcriptomics analyses of a Vero cell line in suspension versus adherent culture conditions. Research Square.
148. Quesney, S., Marc, A., Gerdil, C., Gimenez, C., Marvel, J., Richard, Y., & Meignier, B. (2003). Kinetics and metabolic specificities of Vero cells in bioreactor cultures with serum-free medium. *Cytotechnology*, *42*(1), 1–11.
149. Petiot, E., Guedon, E., Blanchard, F., Gény, C., Pinton, H., & Marc, A. (2010). Kinetic characterization of vero cell metabolism in a serum-free batch culture process. *Biotechnology and Bioengineering*, *107*(1), 143–153.
150. Bastin, G., Chotteau, V., & Vande Wouwer, A. (2021). Metabolic flux analysis of VERO cells under various culture conditions. *Processes*, *9*(12), 2097.
151. Kiesslich, S., Kim, G. N., Shen, C. F., Kang, C. Y., & Kamen, A. A. (2021). Bioreactor production of rVSV-based vectors in Vero cell suspension cultures. *Biotechnology and Bioengineering*, *118*(7), 2649–2659.
152. Taub, M., Chuman, L., Saier Jr, M. H., & Sato, G. (1979). Growth of Madin-Darby canine kidney epithelial cell (MDCK) line in hormone-supplemented, serum-free medium. *Proceedings of the National Academy of Sciences*, *76*(7), 3338–3342.
153. Chu, C., Lugovtsev, V., Golding, H., Betenbaugh, M., & Shiloach, J. (2009). Conversion of MDCK cell line to suspension culture by transfecting with human siat7e gene and its application for influenza virus production. *Proceedings of the National Academy of Sciences*, *106*(35), 14802–14807.
154. Pech, S., Rehberg, M., Janke, R., Benndorf, D., Genzel, Y., Muth, T., Sickmann, A., Rapp, E., & Reichl, U. (2021). Tracking changes in adaptation to suspension growth for MDCK cells: cell growth correlates with levels of metabolites, enzymes and proteins. *Applied Microbiology and Biotechnology*, *105*(5), 1861–1874.
155. Rehberg, M., Rath, A., Ritter, J., Genzel, Y., & Reichl, U. (2014). Changes in intracellular metabolite pools during growth of adherent MDCK cells in two different media. *Applied Microbiology and Biotechnology*, *98*(1), 385–397.
156. Ye, X., Wang, J., Qiao, Z., Yang, D., Wang, J., Abudureyimu, A., Yang, K., Feng, Y., Ma, Z., & Liu, Z. (2021). Quantitative proteomic analysis of MDCK cell adhesion. *Molecular Omics*, *17*(1), 121–129.
157. Wahl, A., Sidorenko, Y., Dauner, M., Genzel, Y., & Reichl, U. (2008). Metabolic flux model for an anchorage-dependent MDCK cell line: Characteristic growth phases and minimum substrate consumption flux distribution. *Biotechnology and Bioengineering*, *101*(1), 135–152.
158. Fernandes, P., Peixoto, C., Santiago, V., Kremer, E., Coroadinha, A., & Alves, P. (2013). Bioprocess development for canine adenovirus type 2 vectors. *Gene Therapy*, *20*(4), 353–360.
159. Carinhas, N., Pais, D. A., Koshkin, A., Fernandes, P., Coroadinha, A. S., Carrondo, M. J., Alves, P. M., & Teixeira, A. P. (2016). Metabolic flux profiling of MDCK cells during growth and canine adenovirus vector production. *Scientific Reports*, *6*(1), 1–11.
160. Tao, P., Ning, Z., Hao, X., Lin, X., Zheng, Q., & Li, S. (2019). Comparative analysis of whole-transcriptome RNA expression in MDCK cells infected with the H3N2 and H5N1 canine influenza viruses. *Frontiers in Cellular and Infection Microbiology*, *9*, 76.
161. Capellini, F. M., Vencia, W., Amadori, M., Mignone, G., Parisi, E., Masiello, L., Vivaldi, B., Ferrari, A., & Razzuoli, E. Characterization of MDCK cells and evaluation of their ability to respond to infectious and non-infectious stressors. *Cytotechnology*, *72*(1), 97–109.
162. Jordan, I., Vos, A., Beilfuß, S., Neubert, A., Breul, S., & Sandig, V. (2009). An avian cell line designed for production of highly attenuated viruses. *Vaccine*, *27*(5), 748–756.

163. Lohr, V., Rath, A., Genzel, Y., Jordan, I., Sandig, V., & Reichl, U. (2009). New avian suspension cell lines provide production of influenza virus and MVA in serum-free media: studies on growth, metabolism and virus propagation. *Vaccine*, *27*(36), 4975–4982.
164. Lohr, V., Hädicke, O., Genzel, Y., Jordan, I., Büntemeyer, H., Klamt, S., & Reichl, U. (2014). The avian cell line AGE1. CR. pIX characterized by metabolic flux analysis. *BMC Biotechnology*, *14*(1), 1–14.
165. Tapia, F., Jordan, I., Genzel, Y., & Reichl, U. (2017). Efficient and stable production of Modified Vaccinia Ankara virus in two-stage semi-continuous and in continuous stirred tank cultivation systems. *PLoS One*, *12*(8), e0182553.
166. Vázquez-Ramírez, D., Genzel, Y., Jordan, I., Sandig, V., & Reichl, U. (2018). High-cell-density cultivations to increase MVA virus production. *Vaccine*, *36*(22), 3124–3133.
167. Nguyen, Q., Nielsen, L. K., & Reid, S. (2013). Genome scale transcriptomics of baculovirus-insect interactions. *Viruses*, *5*(11), 2721–2747.
168. 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.
169. Cruz, P., Moreira, J., & Carrondo, M. (1997). Insect cell growth evaluation during serum-free adaptation in stirred suspension cultures. *Biotechnology Techniques*, *11*(2), 117–120.
170. Benslimane, C., Elias, C. B., Hawari, J., & Kamen, A. (2005). Insights into the central metabolism of Spodoptera frugiperda (Sf-9) and Trichoplusia ni BTI-Tn-5B1-4 (Tn-5) insect cells by radiolabeling studies. *Biotechnology Progress*, *21*(1), 78–86.
171. Monteiro, F., Bernal, V., Saelens, X., Lozano, A. B., Bernal, C., Sevilla, A., Carrondo, M. J., & Alves, P. M. (2014). Metabolic profiling of insect cell lines: Unveiling cell line determinants behind system's productivity. *Biotechnology and Bioengineering*, *111*(4), 816–828.
172. Kondratov, O., Marsic, D., Crosson, S. M., Mendez-Gomez, H. R., Moskalenko, O., Mietzsch, M., Heilbronn, R., Allison, J. R., Green, K. B., Agbandje-McKenna, M., & Zolotukhin, S. (2017). Direct head-to-head evaluation of recombinant adeno-associated viral vectors manufactured in human versus insect cells. *Molecular Therapy*, *25*(12), 2661–2675.
173. Rumachik, N. G., Malaker, S. A., & Paulk, N. K. (2021). VectorMOD: Method for bottom-up proteomic characterization of rAAV capsid post-translational modifications and vector impurities. *Frontiers in Immunology*, *12*, 657795.


**How to cite this article:** Zehetner, L., Széliová, D., Kraus, B., Graninger, M., Zanghellini, J., & Hernandez Bort, J. A. (2023). Optimizing VLP production in gene therapy: Opportunities and challenges for *in silico* modeling. *Biotechnology Journal*, *18*, e2200636. https://doi.org/10.1002/biot.202200636

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

# Logistic PCA explains differences between genome-scale metabolic models in terms of metabolic pathways

**Leopold Zehetner**[1,2,3], **Diana Széliová**[1], **Barbara Kraus**[3], **Juan A. Hernandez Bort**[3]*, **Jürgen Zanghellini**[1]*

**1** Department of Analytical Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria, **2** Vienna Doctoral School in Chemistry (DoSChem), University of Vienna, Vienna, Austria, **3** Gene Therapy Process Development, Baxalta Innovations GmbH, a Part of Takeda Companies, Orth an der Donau, Austria

* juergen.zanghellini@univie.ac.at (JZ); hernandezj76@univie.ac.at (JAHB)





**Citation:** Zehetner L, Széliová D, Kraus B, Hernandez Bort JA, Zanghellini J (2024) Logistic PCA explains differences between genome-scale metabolic models in terms of metabolic pathways. PLoS Comput Biol 20(6): e1012236. https://doi.org/10.1371/journal.pcbi.1012236

**Editor:** Christoph Kaleta, Christian Albrechts Universitat zu Kiel, GERMANY

**Received:** December 9, 2023

**Accepted:** June 7, 2024

**Published:** June 24, 2024



**Data Availability Statement:** All data and codes underlying this article are freely available in a GitHub repository, at https://github.com/LeoZ93/lpca4gsmm.

**Funding:** This work was supported by Baxalta Innovation GmbH (to BK, JAHB, LZ), and by University of Vienna (to DS, JZ). The presented work was partly funded by Baxalta Innovations GmbH. Open access funding was provided by University of Vienna. The funders had no role in

## Abstract

Genome-scale metabolic models (GSMMs) offer a holistic view of biochemical reaction networks, enabling in-depth analyses of metabolism across species and tissues in multiple conditions. However, comparing GSMMs Against each other poses challenges as current dimensionality reduction algorithms or clustering methods lack mechanistic interpretability, and often rely on subjective assumptions. Here, we propose a new approach utilizing logisitic principal component analysis (LPCA) that efficiently clusters GSMMs while singling out mechanistic differences in terms of reactions and pathways that drive the categorization. We applied LPCA to multiple diverse datasets, including GSMMs of 222 *Escherichia*-strains, 343 budding yeasts (*Saccharomycotina*), 80 human tissues, and 2943 *Firmicutes* strains. Our findings demonstrate LPCA's effectiveness in preserving microbial phylogenetic relationships and discerning human tissue-specific metabolic profiles, exhibiting comparable performance to traditional methods like t-distributed stochastic neighborhood embedding (t-SNE) and Jaccard coefficients. Moreover, the subsystems and associated reactions identified by LPCA align with existing knowledge, underscoring its reliability in dissecting GSMMs and uncovering the underlying drivers of separation.

## Author's summary

GSMMs are comprehensive representations of all the biochemical reactions that occur within an organism, enabling insights into cellular processes. Our study introduces LPCA to explore and compare these biochemical networks across different species and tissues only based on the presence or absence of reactions, summarized in a binary matrix. LPCA analyzes these binary matrices of specific biochemical reactions, identifying significant differences and similarities. We applied LPCA to a range of datasets, including bacterial strains, fungi, and human tissues. Our findings demonstrate LPCA's effectiveness in distinguishing microbial phylogenetic relationships and discerning tissue-specific profiles in humans. LPCA also offers precise information on the biochemical drivers of these

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: LZ, BK, and JAHB are employees of Baxalta Innovation GmbH. Employees of Baxalta Innovations GmbH may be owners of stock and/or stock options. DS and JZ are employees of the University of Vienna. The presented work was partly funded by Baxalta Innovations GmbH. The funder had no role in the study design. The funder approved the decision to publish.

differences, contributing to a deeper understanding of metabolic subsystems. This research showcases LPCA as a valuable method for examining the complex interplay of reactions within GSMMs, offering insights that could support further scientific investigation into metabolic processes.

## Introduction

GSMMs are mathematical representations of species- or context-specific metabolic reaction networks [1] and have been successfully applied to study diseases [2–4], optimize bioprocesses [5–8], or investigate metabolic differences across species [9–11]. To compare different GSMMs, dimensionality reduction techniques can be applied to the results of simulations, such as flux balance analysis [12]. The flux distributions or growth rates can be predicted for different environmental conditions and serve as input for principal component analysis (PCA). PCA of growth rates was used to suggest potential auxotrophies [9, 11]. While this approach has become a well established method for comparing GSMMs, it necessitates the incorporation of preassumed environmental data. Another challenge is that the predicted growth rates can be similar even in different environments, which might lead to a reduced discriminative capacity. One potential remediation is to focus solely on simulated growth rates from selected environmental conditions exhibiting significant variation across GSMMs; however, this mandates the integration of subjective parameters or thresholds.

Alternative methods analyze binary matrices derived from the presence or absence of reactions in the models. Jaccard coefficients [13] are often utilized to measure similarity between GSMMs, visualized through heatmaps to identify clusters of similar models. However, this method lacks insight into specific pathways driving heterogeneity. t-SNE analysis [10, 14], while effective in clustering GSMMs consistently, requires specific prerequisites, such as distance metrics, and hyperparameters, which might result in erroneous clustering outcomes [15–17], and may lack reproducibility due to its non-deterministic nature. Additionally, it does not provide a straightforward identification of key variables driving clustering. In contrast, PCA allows the calculation of loadings representing reactions contributing most to principal components. This enables further analysis by grouping loadings based on gene ontology terms, pathways, or subsystems. However, PCA's direct application to binary data is not suitable [18] because it relies on variance and covariance calculations, assumptions of continuous distribution, and linear relationships, which are better suited for continuous datasets rather than binary datasets [19]. Thus, as GSMMs become increasingly complex, the need for automated tools to identify the underlying reactions and pathways driving clustering grows, making simultaneous comparison and factor identification essential.

To address the limitations of existing methods, we use LPCA for classifying GSMMs (see Fig 1). LPCA is an adaption of classic PCA to analyze heterogeneity in binary data [18, 20]. In a classic PCA, continuous data is transformed to a new coordinate system such that the greatest variance lies on the first coordinate. Landgraf and Lee introduced LPCA [21], which extends traditional PCA to handle binary data. They reinterpret PCA as a method for projecting data into a lower-dimensional space while maintaining proximity to the original data. For Gaussian data, this involves a straightforward projection, and minimizing squared error. However, for binary data, LPCA projects natural parameters derived from a Bernoulli model while minimizing Bernoulli deviance.

LPCA has been used with other biological data, such as binary genomics data [18], but not for the comparison of GSMMs. Here we show that LPCA enables efficient clustering based

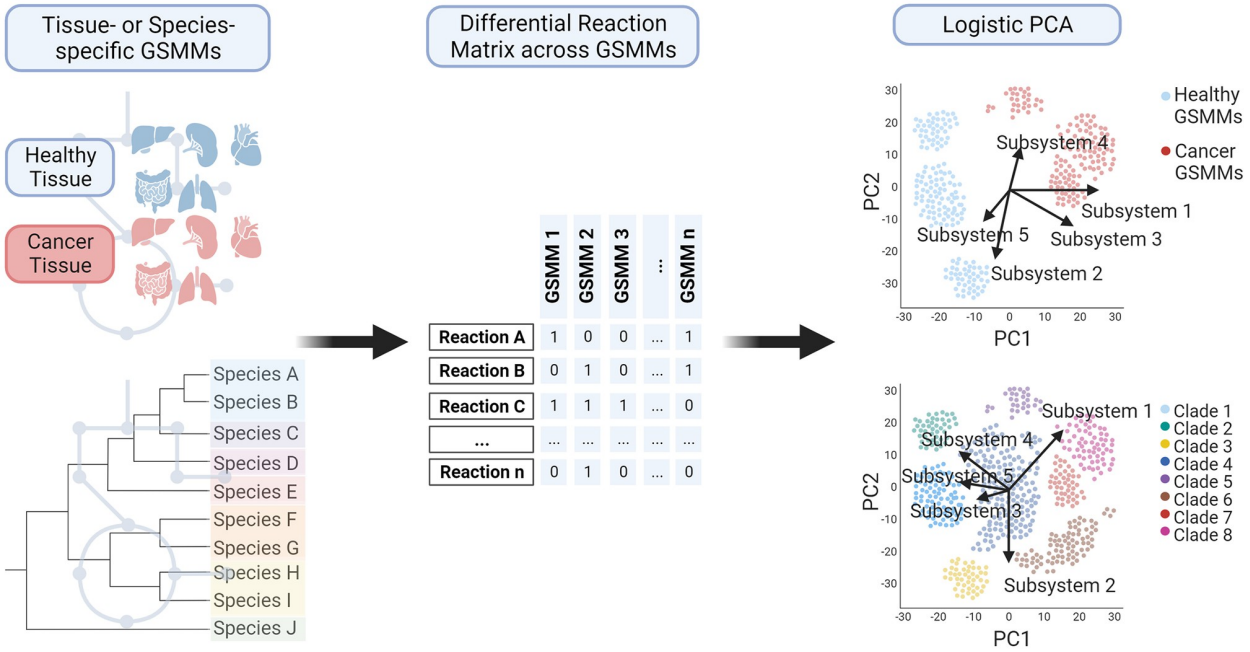


**Fig 1. Schematic workflow of applying LPCA to binary reaction matrices derived from GSMMs.** (Created with BioRender.com).

https://doi.org/10.1371/journal.pcbi.1012236.g001

solely on the presence or absence of reactions. Additionally, using LPCA, reactions were identified as contributing most to clustering, demonstrating its advantages over existing methods like t-SNE and Jaccard coefficients. Furthermore, we identify key subsystems that differentiate GSMMs clusters, a feature not achievable with current methods. We validate our approach by reconstructing phylogenetic associations. Overall, we provide an alternative for streamlined subsystem analysis that elucidates variations across GSMMs.

## Methods

### LPCA on binary reaction matrices

Unless otherwise stated, LPCA was performed on the differential binary pan-reaction matrix, $\Delta\boldsymbol{X}$, using the `logisticPCA` function from the "logisticPCA" package (v0.2) [21] in R (v4.3.1). Due to the large dimensions of the binary matrices, partial decomposition was chosen, as recommended in the "logisticPCA" package documentation. This approach involves computing only a few eigenvalues instead of all of them, thereby speeding up the computation process.

The binary data matrix $\Delta\boldsymbol{X}$ ($N \times R$) consists of $N$ rows, where every row represents one context- or species-specific GSMM, and $R$ columns represent the reactions. LPCA minimizes the Bernoulli deviance $D$, which quantifies the difference between the binary data matrix $\Delta\boldsymbol{X}$, and the LPCA reconstruction $\widehat{\Theta}$, representing the estimated natural parameters of the Bernoulli model,

$$\min_{\boldsymbol{\mu},\boldsymbol{U}} D(\Delta\boldsymbol{X}|\widehat{\boldsymbol{\Theta}}) \quad = -2\sum_{n=1}^{N}\sum_{r=1}^{R}\Delta X_{n,r}\widehat{\Theta}_{n,r} - \log[1+\exp(\widehat{\Theta}_{n,r})] \tag{1a}$$

with

$$\widehat{\Theta} = 1_N \boldsymbol{\mu}^T + (\boldsymbol{\Theta} - 1_N \boldsymbol{\mu}^T) U U^T, \tag{1b}$$

$$\boldsymbol{\Theta} = m(2\Delta \boldsymbol{X} - \mathbf{1}_N \mathbf{1}_R^T), \tag{1c}$$

with a tunable parameter $m$ that, in default mode, is automatically selected by the `logisticPCA` function. $\boldsymbol{\Theta}$ ($N \times R$) denotes the matrix of natural parameters from the saturated model. A saturated model in the context of binary data and Bernoulli distributions is one in which the estimated probabilities exactly match the observed data $\Delta \boldsymbol{X}$. The results from the `logisticPCA` function include the mean parameter vector $\boldsymbol{\mu}$ ($R \times \mathbf{1}$) and the loading matrix $\boldsymbol{U}$ ($R \times i$), where $i$ belongs to the number of principal components. $\boldsymbol{U}$ and $\boldsymbol{\mu}$ are solved such that the Bernoulli deviance $D$ is minimized.

The principal component scores $\boldsymbol{S}$ ($N \times i$) are then obtained from $\boldsymbol{U}$ and $\boldsymbol{\mu}$ within the `logisticPCA` by:

$$\boldsymbol{S} = (\boldsymbol{\Theta} - \mathbf{1}_N \mu^T) \boldsymbol{U} \tag{2}$$

The set of loadings $\boldsymbol{U}_{*,\boldsymbol{r}}$ across all principal components $i$ for a reaction $r$ represents a loading vector.

We refer to a loading vector containing loadings for the first two principal components as reaction-centric loading vector. These reaction-centric loading vectors can be added to LPCA score plots as arrows to visualize how each reaction contributes to the clustering.

To avoid overwhelming complexity in visualization, we introduce subsystem-centric loading vectors as proxies. For each principal component $i$, we compute the average loading across all reactions $r$ within subsystem $j$, denoted as

$$\mathrm{avg}\left(U_i^j\right) = \frac{1}{R_j} \sum_{r=1}^{R_j} U_{i,r}^{(j)}. \tag{3}$$

Here, $R_j$ is the number of reactions in subsystem $j$, and $U_{i,r}^{(j)}$ is the loading of principal component $i$ for reaction $r$ within the subsystem $j$.

The average loading vectors for the first two principal components were visualized in the LPCA plots.

To rank the contributions of the subsystems to the principal components, we compute the Euclidian norm of the subsystem-centric loading vectors across the first two principal components

$$|\mathrm{avg}(U^j)| = \sqrt{\sum_{i=1}^{2} \mathrm{avg}(U_i^j)^2} = \sqrt{\sum_{i=1}^{2} \frac{1}{R_j} \sum_{r=1}^{R_j} U_{i,r}^{(j)}}. \tag{4}$$

As an alternative ranking measure, we first calculate the Euclidian norm of every reaction-centric loading vector, and then the average as

$$\mathrm{avg}|U^j| = \frac{1}{R_j} \sum_{r=1}^{R_j} \sqrt{\sum_{i=1}^{2} \left(U_{i,r}^{(j)}\right)^2}, \tag{5}$$

where the inner sum traverses over the first two principal components.

Both measures were employed to identify key subsystems responsible for driving the observed differentiation.

## Data sources, collection and pre-processing

**Differential binary pan-reaction matrix, ΔX.** We acquired GSMMs and metabolic reconstructions from four distinct sources as outlined below. For each dataset, we generated a binary "pan-reaction" matrix $\boldsymbol{X} \in \{0, 1\}^{N \times R}$, where each of the $N$ rows corresponds to a GSMM and each of the $R$ columns represents a reaction, using COBRApy (v0.26.3) [22]. These matrices indicate whether individual reactions are present (1) or absent (0) within species-specific GSMMs or context-specific reconstructions and provide a comprehensive overview of the reactions present across all the GSMMs. To simplify the matrices, we removed columns containing only 1 or 0. We refer to the resulting data as the differential binary reaction matrix $\Delta\boldsymbol{X}$ of the pan-GSMM or the pan-reconstruction.

- ***Escherichia***. We used 222 GSMMs of *Escherichia* species reconstructed by Monk [9], grown across 570 environmental conditions. The resulting pan-GSMM contains 3342 reactions, each present in at least one strain. Of these, 1688 reactions are not consistently present across all strains, which thus form the differential matrix $\Delta\boldsymbol{X}_{Escherichia} \in \{0, 1\}^{222 \times 1688}$.
  During our analysis, we discovered that the reaction labeled as "PRCOA1" is associated with "Cholesterol degradation" [23] rather than its initial association with "Histidine Metabolism" [9]. Consequently, we established a new subsystem named "Cholesterol degradation", reassigning "PRCOA1" and related reactions based on BiGG annotations. Subsequently, we recalculated the LPCA and subsystem-centric loadings to reflect this adjustment.
- ***Firmicutes***. We used 2943 GSMMs of the phylum *Firmicutes* from the Agora2 dataset [14] forming a differential matrix $\Delta\boldsymbol{X}_{Firmicutes} \in \{0, 1\}^{2943 \times 5267}$.
- ***Fungi***. We used 343 GSMMs of fungi reconstructed by Lu et al. [10] based on previous sequencing data [24]. The resulting pan-GSMM contains 4599 reactions, each present in at least one strain. Of these, 2519 reactions are not consistently present across all strains, forming the differential matrix $\Delta\boldsymbol{X}_{Fungi} \in \{0, 1\}^{343 \times 2519}$.
- ***Human***. We used normalized gene expression data from 50 healthy tissue samples [25] and 30 cancer tissue samples [26] sourced from the Human Protein Atlas [25, 26] to generate context-specific genome-scale metabolic reconstructions. Reactions were taken from the latest iteration of the human metabolic model, "Human1" [27], with gene expression considered for normalized transcripts per million (nTPM) values exceeding 0.2. To avoid bias from gap-filling, only reactions with gene assignments were included in the context-specific reconstruction. The resulting pan-reconstructions contained 7975 reactions. Of these, 2390 reactions are not consistently present across all reconstructions, which thus form the differential matrix $\Delta\boldsymbol{X}_{\text{Human}} \in \{1, 0\}^{80 \times 2390}$.

## Additional computational analyses

**t-SNE on binary reaction matrices.** Binary reaction datasets from context-specific GSMMs were analyzed by Hamming-distance based t-SNE calculation using the R package "Rtsne" (v0.16) [28]. For plotting, only the first two t-SNE values were considered.

**Jaccard coefficients on binary reaction matrices [29].** Jaccard coefficients were calculated based on binary reaction pairwise between context-specific GSMMs using:

$$J(A, B) = \frac{|\Delta\boldsymbol{X}_{A,*} \cap \Delta\boldsymbol{X}_{B,*}|}{|\Delta\boldsymbol{X}_{A,*} \cup \Delta\boldsymbol{X}_{B,*}|} \tag{6}$$

where $J(A, B)$ is the Jaccard coefficient between differential reactions of GSMMs $A$ and $B$, consisting of reactions $\Delta \boldsymbol{X}_{A,*}$, and $\Delta \boldsymbol{X}_{B,*}$, respectively. $|\Delta \boldsymbol{X}_{A,*} \cap \Delta \boldsymbol{X}_{B,*}|$ is the intersection of reactions between GSMMs $A$ and $B$, and $|\Delta \boldsymbol{X}_{A,*} \cup \Delta \boldsymbol{X}_{B,*}|$ is union of reactions between GSMMs $A$ and $B$.

**Principal component analysis.** ***Escherichia***. For visual comparison, the PCA plot based on simulated growth rates from environmental conditions was reproduced for *Escherichia* strains from [9]. To obtain a similar clustering, a correlation matrix was computed, and PCA was performed using the `princomp` function in R.

**Phylogenetic analysis.** ***Escherichia***. Proteins from genomic sequences (downloaded from Enterobase—v1.1.5 [30]) were predicted using "prodigal" (v2.6.3) [31], followed by phylogenetic comparison using "OrthoFinder" (v2.5.5) [32]. The phylogenetic tree was then plotted using the "ape" package (v5.7–1) [33] in R.

***Fungi***. The phylogenetic tree was reproduced using the "ape" package [33] based on a Newick file published by Shen *et al* [24].

**Cophenetic correlation coefficient.** We used the cophenetic correlation coefficient [34] to examine similarity between LPCA scores, and phylogenetic trees, where possible. Initially, we computed the pairwise distances among LPCA scores using both Euclidean and Manhattan metrics. These distances were then subjected to hierarchical clustering, resulting in the formation of dendrograms. The obtained dendrograms were compared using the cophenetic correlation coefficient by applying the `cophenetic`, and `cor` functions in R. Due to the inherent non-deterministic nature of t-SNE, a comparison based on the cophenetic correlation coefficient is not meaningful.

**Subsystem analysis using multinomial logistic regression.** A multinomial logistic regression (MLR) model, using the `multinom` function ("nnet", v7.3–19), was applied to calculate the contribution of specific reactions to pre-defined clusters of GSMMs (i.e. phylogenetic clades/tissue types). Parameter estimates were derived for each reaction, reflecting their relative contributions to the cluster membership. To quantify the aggregate influence of reactions within biological subsystems, we calculated the Euclidean norm of the parameter estimates for each reaction, which provides a measure of the reaction's importance. These values were then averaged by subsystem, offering a subsystem-level perspective on the factors driving the clustering of metabolic models following:

$$\text{avg}|\beta^j| = \frac{1}{R_j}\sum_{r=1}^{R_j}\sqrt{\sum_{c=1}^{12}\left(\beta_{c,r}^j\right)^2}, \tag{7}$$

where $\beta$ is the coefficient for reaction $r$ in subsystem $j$, for one of the 12 clades $c$. $R_j$ is the number of reactions in subsystem $j$. Before plotting, subsystems were normalized by dividing every subsystem by the highest value from MLR and LPCA.

## Results and discussion

### Comparison of 222 strain-specific GSMMs from the genus *Escherichia*

We used LPCA to analyze 222 strain-specific GSMMs of the genus *Escherichia* [9]. Both the differential binary pan-reaction matrix $\Delta \boldsymbol{X}$ (Fig 2A) and the full binary pan-reaction matrix (S1 Fig) were subjected to LPCA, and showed similar clustering behavior. Overall, seven subclusters were identified, three of which were exclusively associated with *E. albertii* strains (blue in Fig 2A), two with *E. fergusonii* strains (red in Fig 2A, red), and the remaining two consist of *E. coli*, *S. dysenteriae*, *S. flexneri*, as well as Clades II to VIII (orange in Fig 2A). While *E. coli* strains did not segregate from *Shigella* species, they tended to accumulate on one side of the

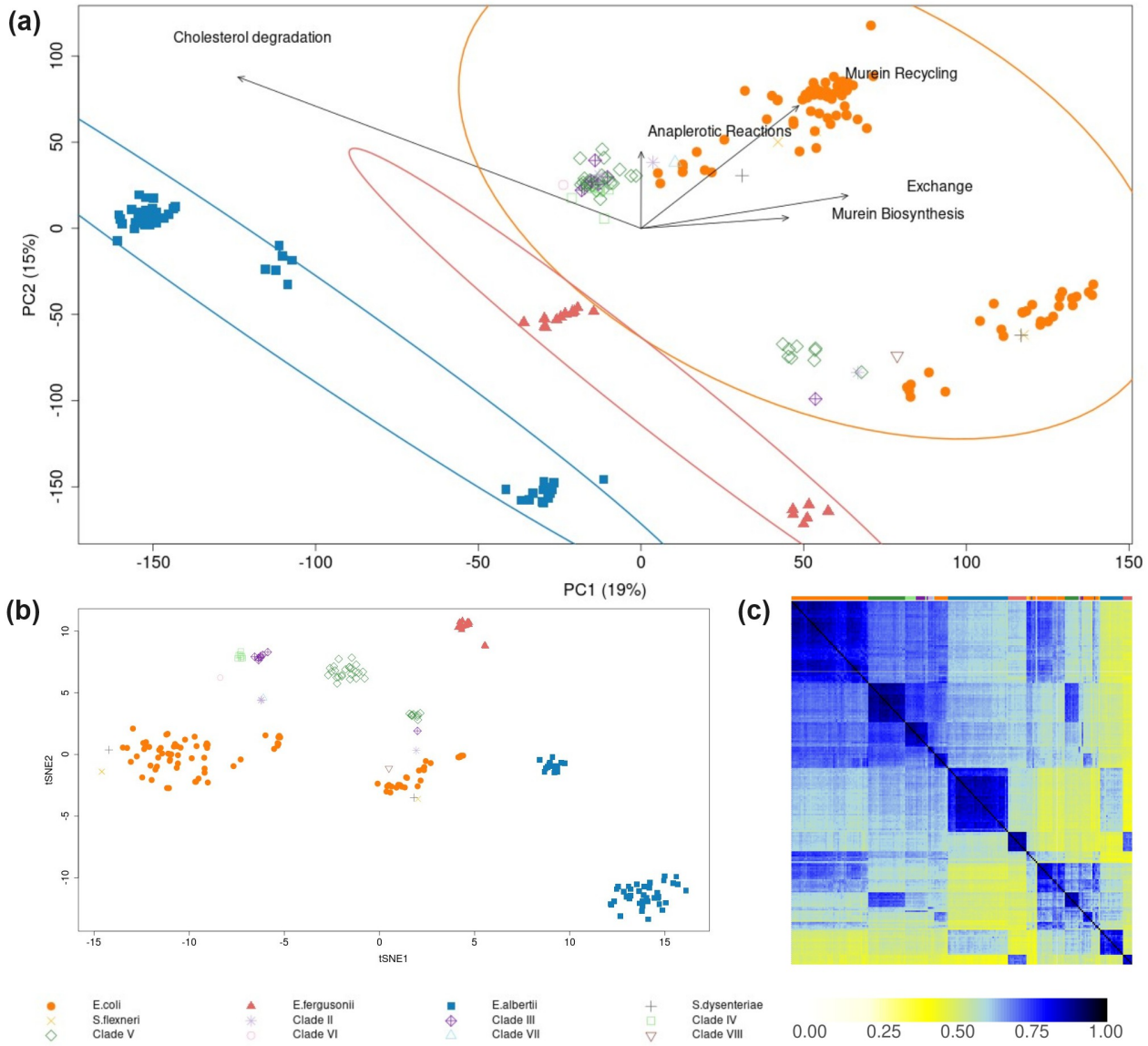


**Fig 2. LPCA (a), t-SNE (b) and Jaccard coefficients (c) derived from a binary reaction matrix from differential reactions in 222 *Escherichia* GSMMs.** In panels (a) and (b), points represent individual GSMMs, with different genera indicated by distinct symbols and colors. The top row in panel (c) uses these same colors to indicate the corresponding genera. Circles in panel (a) highlight clusters of *E. albertii* strains (blue), *E. fergusonii* strains (red), and a mixed cluster of *E. coli*, *S. dysenteriae*, *S. flexneri*, and Clades II to VIII (orange). Labeled arrows in panel (a) denote subsystem-centric loading vectors from LPCA (refer to the results and methods section for definitions).

https://doi.org/10.1371/journal.pcbi.1012236.g002

two subclusters within the orange ellipse. In contrast, Clades II to VIII gathered on the opposite side.

**LPCA aligns with established GSMM classification methods and phylogenetic analysis.** To validate our LPCA analysis, we compared it with t-SNE analysis (see Fig 2B) and a Jaccard coefficient analysis (see Fig 2C). Overall the three methods showed similar clustering behavior.

- t-SNE analysis found two distinct clusters for *E. albertii* (blue in Fig 2B), contrasting the three clusters identified by LPCA. Additionally, t-SNE identified only one cluster for *E. fergusonii* (red in Fig 2B), as opposed to two clusters observed with LPCA. Both LPCA and t-SNE identified two identical clusters for *E. coli* (orange in Fig 2B). However, t-SNE provided better resolution for the remaining clades compared to LPCA.
- Ten permutations of the sample order resulted in substantial variability in the t-SNE-derived coordinates, highlighting the sensitivity of this method. In contrast, the scores obtained from the LPCA approach remained remarkably stable across the same number of permutations (compare S5 Fig).
- Hierarchical clustering of pairwise Jaccard coefficients revealed two distinct clusters for *E. albertii*, two clusters for *E. fergusonii*, and two mixed clusters comprising *E. coli* and the remaining clades (Fig 2C).

Again, clustering patterns remained essentially consistent across all methods when using the full binary pan-reaction matrix instead of the differential matrix (compare Fig 2 with S1 Fig).

Next, we tested if the clustering by LPCA aligns with a phylogenetic analysis (S3 Fig). We compared LPCA scores with a whole-genome based phylogenetic tree using the cophenetic correlation coefficient [34]. LPCA scores showed a good cophenetic correlation with the phylogenetic tree when Manhattan distances (0.61) were used to calculate the pairwise distance between scores. The correlation weakened when using Euclidean distance. This drop in correlation with Euclidian distance (0.37) is expected as it is more suited for continuous data rather than binary data. Consequently, we suggest that Manhattan distances derived from LPCA scores effectively conserve phylogenetic relations.

Finally, when comparing LPCA to standard PCA using simulated growth rates under different environmental conditions, we observe a better separation of GSMMs, especially between *E. coli* and *E. albertii* (S2 Fig).

**LPCA pinpoints key metabolic subsystems distinguishing GSMM clusters.** To identify the factors driving the observed clusters, we analyzed subsystem-centric loading vectors derived from LPCA, as described in Eq (2) in the Methods section. These vectors aim to measure the contribution of the entire metabolic subsystem (like glycolysis, etc.) rather than individual reactions. We find that cluster separation was predominantly influenced by (the top) five subsystems: "Murein Biosynthesis", "Murein Recycling", "Anaplerotic Reactions", "Exchange", and "Histidine Metabolism" (Fig 2A). Notably, when employing a different ranking method based on (4) rather than (3), four out of these five driver subsystems remained consistent (see S1 Table), underscoring their robust impact on the clustering.

The subsystem "Exchange" consists of only one reaction: Methylmalonate-semialdehyde dehydrogenase ("MMSAD3"). Its presence in this list is unexpected given that the *Escherichia* GSMMs were reconstructed to compare 570 environmental conditions. Thus, the corresponding exchange systems should be universally present in all models, regardless of genetic evidence. According to BiGG annotation, "MMSAD3" is indeed categorized as exchange but is associated with propanoate metabolism according to KEGG. We suspect that "MMSAD3" might have been incorrectly assigned.

"Murein Recycling" and "Murein Biosynthesis" are known to be highly conserved subsystems among *Escherichia* strains, with 88% of the reactions being shared across all *Escherichia* strains [9]. The highest loading values were obtained for D-alanyl-D-alanine dipeptidase ("ALAALAD" from the "Murein Recycling" subsystem), which is conserved in 82% *E. coli*s, in 100% *Shigella*s, and in 2% Clades II to VIII. In contrast, "ALAALAD" is completely absent in

the more distinct clades, *E. albertii* and *E. fergusonii*. D-alanyl-D-alanine dipeptidase (*vanX*) is responsible for cleaving D-alanyl-D-alanine dipeptide. This enzyme allows bacteria to use D-alanine as a carbon source and modify peptidoglycan structures [35]. Intriguingly, modifications in these peptidoglycan precursors, such as the substitution of D-alanyl-D-alanine with D-alanyl-D-lactate, confer resistance to the antibiotic vancomycin, which targets terminal D-alanyl-D-alanine residues to prevent crosslinking [35, 36].

In further exploring the capabilities of LPCA, we analyzed the set of differential reactions within the "Histidine Metabolism" subsystem when contrasting the clades *E. albertii* and *E. coli*, as an example. We noted virtually identical loading values for the reactions "HISDr", "URCN", and "IZPN" (S2 Table). These reactions detail the stepwise breakdown of histidine into N-formimidoyl-L-glutamate via urocanate and 4-imidazolone-5-propanoate. This pathway was found to be twice as prevalent in *E. albertii* strains (25%) compared to *E. coli* strains (12%). This differential trait suggests that at least certain *E. albertii* strains might possess the metabolic capability to utilize histidine as a source of carbon and nitrogen.

One particular reaction, "PRCOA1", prevalent in 61% of *E. albertii*-specific GSMMs, appeared as a significant difference, see S2 Table. However, we found that this reaction might be misallocated to "Histidine Metabolism". According to the BiGG database [37], "PRCOA1" converts CoA-20-hydroxy-cholest-4-en-3-one C5 side chain to Androst-4-ene-3,17-dione, and belongs to "Cholesterol degradation", rather than "Histidine Metabolism" (see reaction-centric loadings in S2 Table). Thus, we newly introduced the so far missing subsystem "Cholesterol degradation", reassigned "PRCOA1" (and associated reactions according to KEGG), and repeated LPCA (see Fig 2A) as well as computing the subsystem loadings new. This time "Cholesterol degradation" replaced "Histidine metabolism" among the top five driving subsystems. The results from Fig 2 suggest that LPCA is capable of distinguishing between phylogenetically distant species by analyzing their sets of differential reactions. Additionally, the loading values from LPCA help identify the key drivers for the observed separation between species and may also hint at misannotations. Unlike other methods, this identification happens simultaneously with the computation of LPCA scores, offering insights into both the extent of metabolic differences and the specific distinctions driving them.

**MLR reveals subsystems contributing to phylogenetic classification.** To validate our interpretation of the subsystem-centric loadings from LPCA, we used a MLR model to independently evaluate the contribution of each reaction to phylogenetic clades (see Methods for details). By grouping the reaction-centric parameters by subsystem (see Methods 6), we observed different subsystems to be the most influential drivers for separation in comparison to LPCA-derived subsystem-centric loadings. As displayed in Fig 3, "Pentose Phosphate Pathway", "Lipopolysaccharide Biosynthesis / Recycling", "Pentose and Glucuronate Interconversions", "Murein Biosynthesis", and "Anaplerotic Reactions" are the five most important subsystems, responsible for the separation of clades. Two subsystems are shared between MLR and LPCA within the top five subsystems: "Murein Biosynthesis", and "Anaplerotic Reactions". Moreover, "ALAALAD" is again the most influential reaction within the subsystem "Murein Biosynthesis" in MLR. Additionally, it seems that a lower number of subsystems may impact the clustering, since 8 subsystems have a normalized value above 0.5 in MLR, while only 4 subsystems have a normalized value above 0.5 in LPCA. Normalized values in MLR and LPCA were obtained by dividing every subsystem-centric value by the maximum value. While MLR can pinpoint key determinants of a pre-defined categorization, LPCA provides a notable benefit through the detection of possibly new (sub)clusters. Such subclusters could reveal metabolic distinctions that might be missed when exclusively depending on the primary phylogenetic categorization. Moreover, LPCA allows for the identification of the orientation of reaction- or subsystem-centric loadings, an analysis unachievable with MLR.

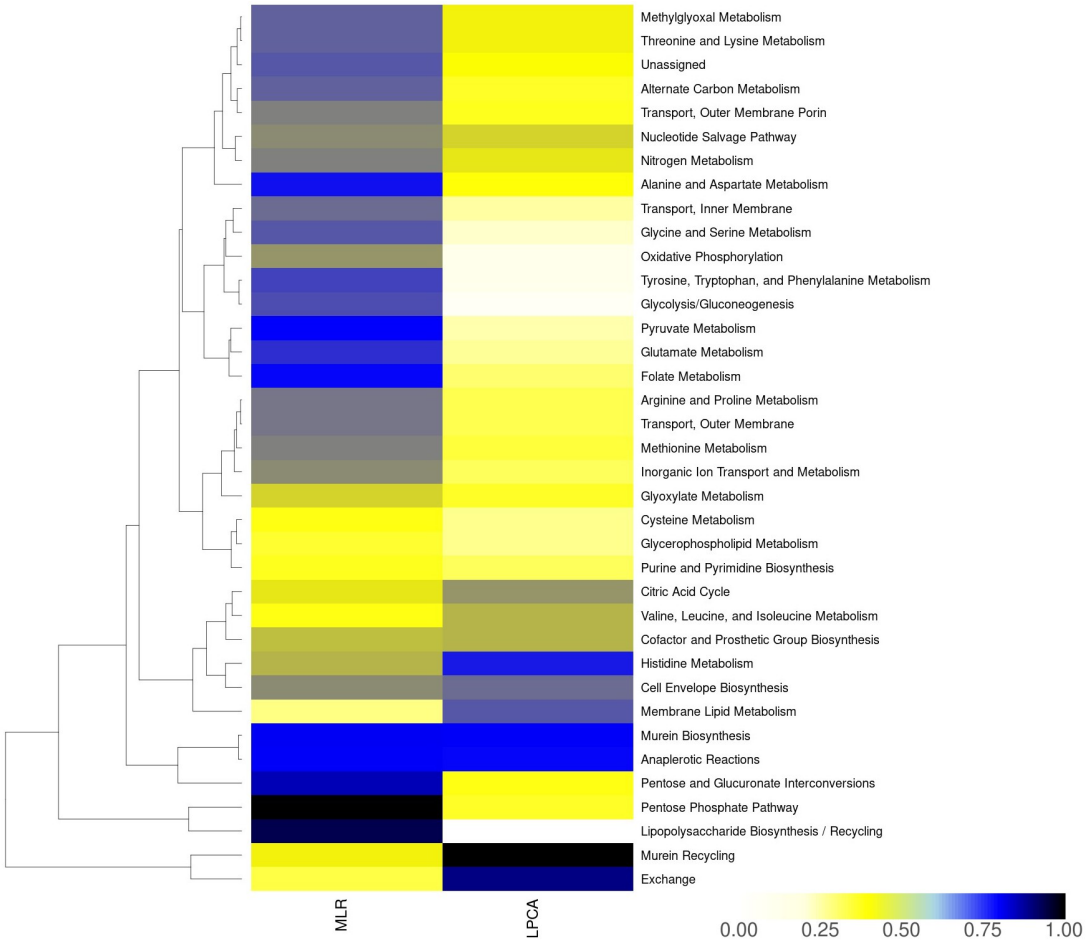


**Fig 3. Impact of subsystems derived from LPCA and MLR for *Escherichia* GSMMs.** MLR: contribution of subsystems to phylogenetic classification normalized to the maximum value. LPCA: subsystem-centric loadings normalized to maximum loading (refer to methods section for details).

https://doi.org/10.1371/journal.pcbi.1012236.g003

## Comparison of 343 yeast-specific GSMMs from the subphylum *Saccharomycotina*

We applied LPCA to the set of differential reactions extracted from 343 yeast-specific GSMMs [10], mostly obtained through a recent sequencing effort [24], see Fig 4a. Similar to the LPCA analysis of GSMMs for *Escherichia* (see above), our findings reveal a clustering of species with closer phylogenetic relationships. However, when considering the first two principal components, the separation is less pronounced for yeast, accounting for only 27% of the variance, in contrast to 34% for *Escherichia* species. In the phylogenetic tree S6 Fig, similar clades were observed to group into closely related clusters as in the LPCA plot Fig 4A. Notably, the *Lipomycetaceae* clade (purple) was distinctly isolated from other clades. In contrast, clades with closer phylogenetic relationships formed subclusters. In particular, the *Saccharomycodaceae*-specific GSMMs (green) constituted a distinct cluster, despite their close phylogenetic relation to *Saccharomycetaceae*.

Again, our analysis identified the top five subsystems critically contributing to cluster separation: "Glycerolipid metabolism", "Phagosome", "Nitrogen metabolism", "Fructose and mannose metabolism", and "Peroxisome". However, comprehensively examining these pathways

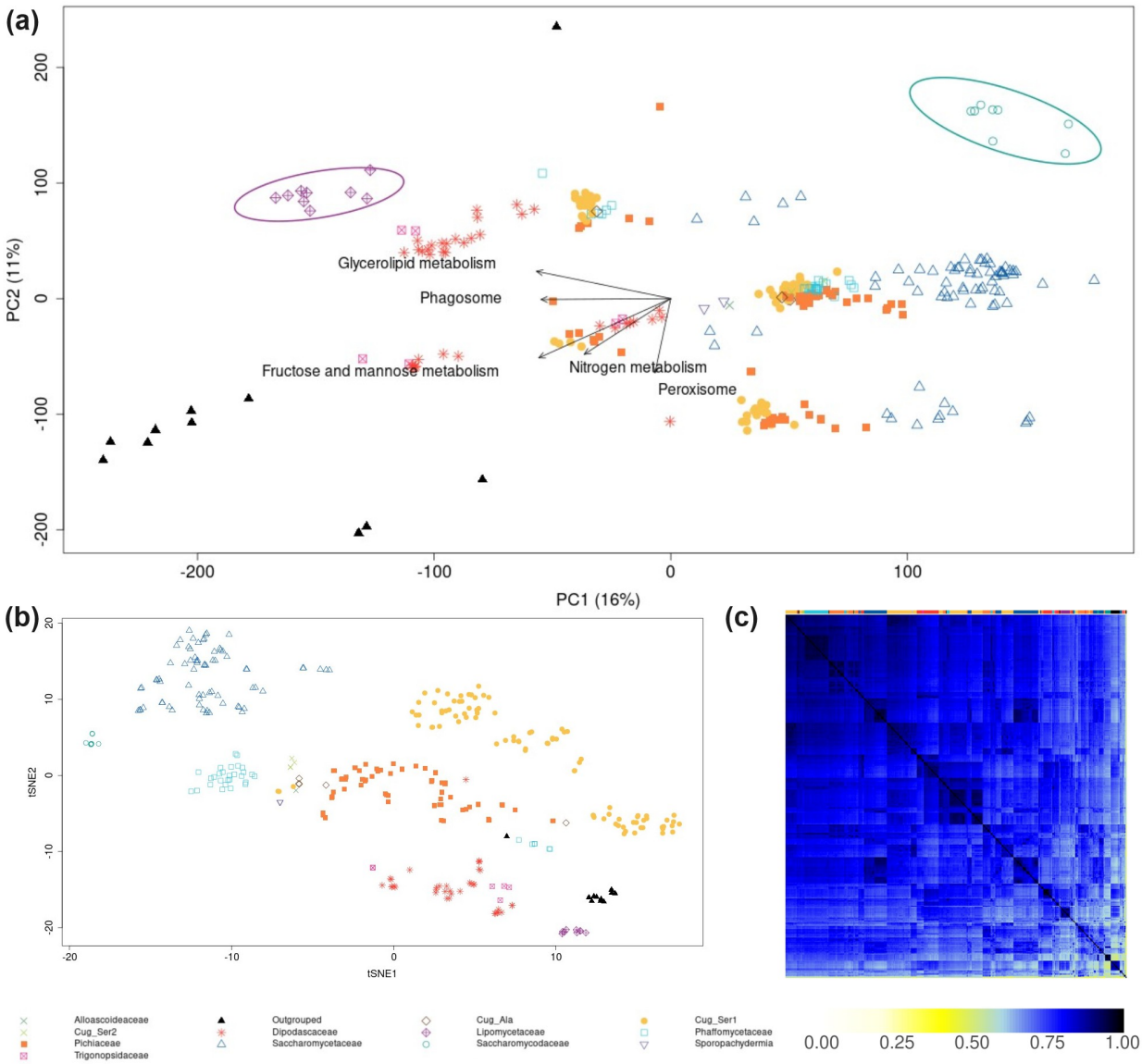


**Fig 4. LPCA (a), t-SNE (b), and Jaccard coefficients (c) derived from a binary reaction matrix from yeast-specific GSMMs.** In panels (a) and (b), points represent individual GSMMs, with different genera indicated by distinct symbols and colors. The top row in panel (c) uses these same colors to indicate the corresponding genera. Circles in panel (a) highlight a cluster of the *Lipomycetaceae* clade (purple), and the *Saccharomycodaceae* (green). Labeled arrows in panel (a) denote subsystem-centric loading vectors from LPCA (refer to the results and methods section for definitions).

https://doi.org/10.1371/journal.pcbi.1012236.g004

in yeast (as we did above for *Escherichia*) faces two major challenges challenges (i) limited knowledge on fungal metabolism—yeast and other fungi remain relatively understudied compared to mammalian and prokaryotic cells [38]; and (ii) a significant proportion of reactions (>50%) are in the subsystem "Unassigned" in the current dataset.

## Comparison of healthy and cancerous tissues

Previous studies have demonstrated the effectiveness of PCA in distinguishing between healthy and cancerous tissues using transcriptomic data [39, 40]. Here, we replicated this finding using

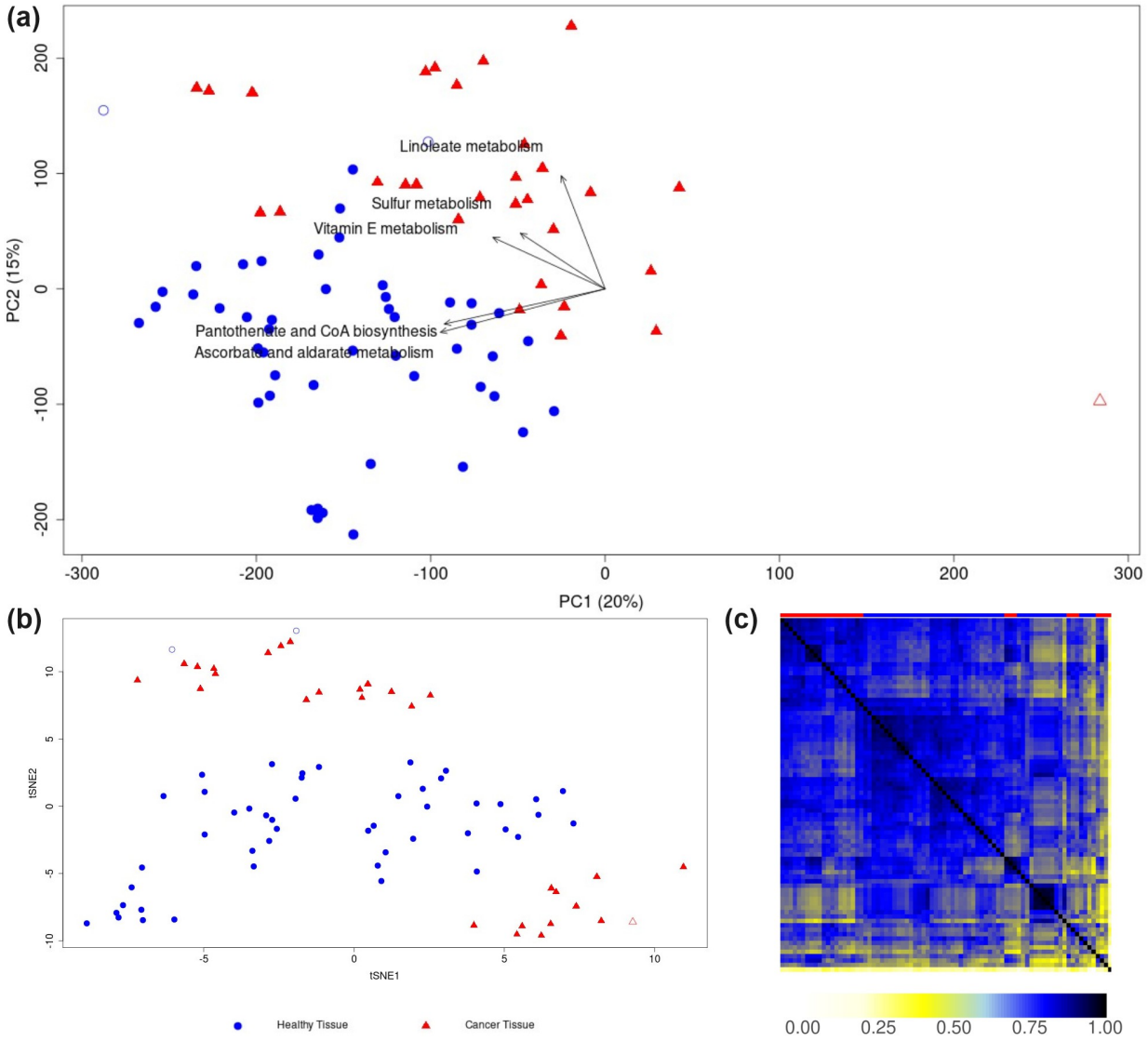


**Fig 5. LPCA (a), t-SNE (b), and Jaccard coefficients (c) derived from a binary reaction matrix from context-specific reconstructions from healthy and cancerous tissues.** In panels (a) and (b), points represent individual reconstructions, with reconstructions from healthy (blue squares) and cancerous tissues (red triangles). The top row in panel (c) uses these same colors to indicate the corresponding tissue. t-SNE appears less effective in identifying outliers (red open triangle, and blue open circles) compared to LPCA and Jaccard coefficients. Labeled arrows in panel (a) denote subsystem-centric loading vectors from LPCA (refer to the results and methods section for definitions).

https://doi.org/10.1371/journal.pcbi.1012236.g005

normalized gene expression data taken from the Human Protein Atlas [25, 26], see S7 Fig. When restricting the analysis to genes annotated within the generic "Human1" reconstruction, the distinction became less evident, see S7 Fig. Further narrowing the analysis to genes derived from the annotations in context-specific GSMMs (see Methods for details) resulted in the loss of differentiation (S7 Fig). This outcome may be attributed to the notably fewer genes (561) in these models compared to the broader set in "Human1" (containing 2897 metabolic genes). We explored if LPCA could recover the differentiation between healthy and cancerous tissues using the differential set of reactions in these GSMMs. Indeed, Fig 5A indicates the feasibility of such recovery via LPCA.

Again, in comparison to t-SNE (Fig 5B) and Jaccard coefficients (Fig 5C) suggests that both t-SNE and LPCA can distinguish healthy from cancerous reconstructions. However, t-SNE appears less effective in identifying outliers (Fig 5, red, unfilled triangle) compared to LPCA and Jaccard coefficients. In addition, two healthy tissue samples (Fig 5, blue, unfilled circles) tended to cluster with reconstructions from cancer tissues, showing a more pronounced clustering when using LPCA compared to t-SNE. LPCA facilitates a clearer interpretation of the underlying factors, aiding in the identification of differences between context-specific GSMMs.

When analyzing the LPCA-derived subsystem-centric loadings, "Linoleate Metabolism" emerged as a crucial subsystem for distinguishing between context-specific GSMMs. Within this subsystem, the reaction "MAR02438" exhibited the highest loading value and is associated with the genes *PTGS1* or *PTGS2*. These genes play a pivotal role in prostaglandin synthesis and are linked to the vascular endothelial-derived growth factor (VEGF) signaling pathway, critical for angiogenesis [41]. *PTGS2* holds significant importance in cancer research. Inhibitors targeting the enzyme COX2, encoded by the *PTGS2* gene, such as celecoxib, have demonstrated cancer-retarding properties [42].

Another important insight came from the subsystem "Ascorbate and alderate metabolism", which was found to be top-ranked in subsystem-centric loadings. It consists of only one reaction-centric loading ("MAR08346"), pointing towards healthy tissue samples (Fig 5A), which indicates a higher presence in healthy tissues than cancerous tissues. The reaction describes the reversible conversion of "L-gulonate" to "L-gulono-1,4-lactone" in the endoplasmic reticulum and is catalyzed by the enzyme "Regulacin" (*RGN*). Besides its metabolic role, *RGN* is involved in calcium homeostasis, antioxidant defense, apoptosis, and cell proliferation [43]. Recently, *RGN* has been identified to be downregulated in several cancer cells [44] and that survival of cancer patients is positively correlated with a higher expression level of *RGN* [45].

In the "Human1" GSMM [27], "L-gulonate" can be converted to "L-gulono-1,4-lactone" (*RGN*, endoplasmic reticulum), "glucuronate" (*AKR1A1*, endoplasmic reticulum), or to "3-dehydr-L-gulonate" (*CRYL1*, cytoplasm). High *CRYL1*-expression has been shown to increase survival rate at least in clear cell renal cell carcinoma patients, while *CRYL1* silencing led to increased cell migration and proliferation [46]. In contrast, *AKR1A1* was found to be upregulated in many cancer cells and is associated with drug resistance [47]. Since *AKR1A1* catalyzes the final conversion step from "D-glucose" to "L-gulonate" via "glucuronate", while *RGN*, and *CRYL1* are downregulated, an accumulation of "L-gulonate" might emerge in cancerous cells and could be further investigated. This finding is supported by the subsystem analysis using MLR (Fig 6), where the subsystem "Ascorbate and alderate metabolism" is ranked on top as well. Further top-ranked subsystems from MLR include "Terpenoid backbone biosynthesis", "Metabolism of other amino acids", "Tricarboxylic acid cycle and glyoxylate/dicarboxylate metabolism", and "Phosphatidylinositol phosphate metabolism", none of them shared with the top-ranked subsystems from LPCA, being "Linoleate metabolism", "Pantothenate and CoA biosynthesis", "Vitamin E metabolism", and "Sulfur metabolism".

## Analysis of 2943 Agora2-derived *Firmicutes* species

Finally, we applied LPCA to a subset (*Firmicutes*) of the Agora2 dataset, after creating a binary reaction matrix, containing 5267 differential reactions from 2943 species-specific GSMMs. The three main orders (*Bacillales*, *Eubacteriales*, and *Lactobacillales*) formed subclusters regardless of the applied method (Fig 7). Incomplete subsystem assignments within the GSMMs prevented the determination of meaningful subsystem-centric loadings. This underscores the significance of accurate subsystem assignments when utilizing LPCA for GSMMs.

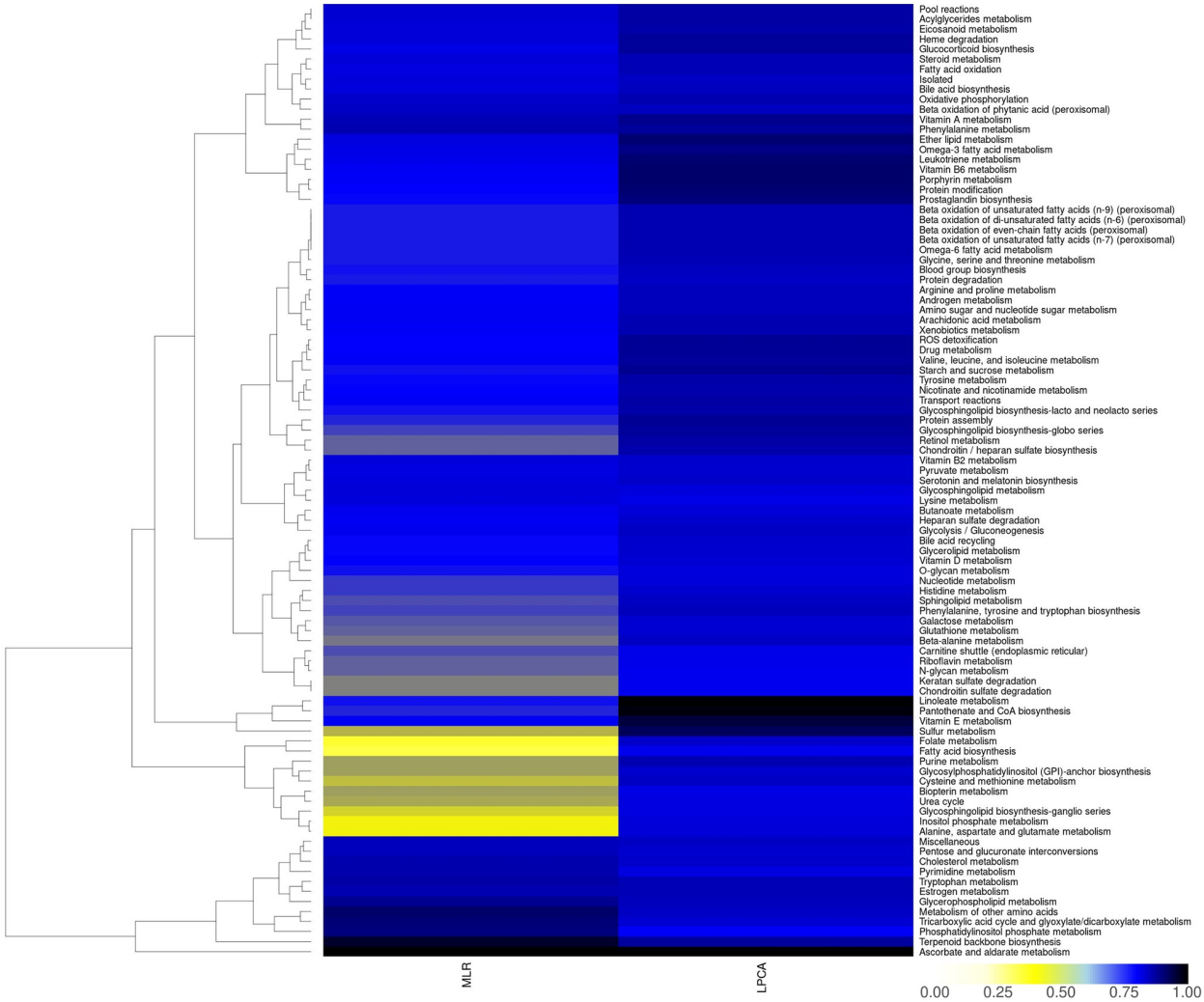


**Fig 6. Impact of subsystems from LPCA and MLR for human reconstructions.** MLR: contribution of subsystems to phylogenetic classification normalized to the maximum value. LPCA: subsystem-centric loadings normalized to maximum loading (refer to methods section for details).

https://doi.org/10.1371/journal.pcbi.1012236.g006

We assessed the run time for LPCA, t-SNE, and Jaccard coefficients. While t-SNE and Jaccard coefficients finished in 0.31 h and 1.26 h, respectively, LPCA took 20 h on an "AMD EPYC 7542 32-Core Processor", 96 cores, 406 GB RAM. While LPCA offers valuable insights into GSMMs at the subsystem and reaction levels, its longer runtime may pose challenges for huge datasets and could benefit from further optimization.

## Conclusion

Here we introduced LPCA to simultaneously compare and analyze multiple GSMMs to identify similarities and differences in metabolic capabilities across multiple species and strains. LPCA extends standard PCA to binary data sets. Thus, it allows us to analyze the presence or absence of biochemical reactions across GSMMs. Our approach not only confirmed the established phylogenetic relationships but also demonstrated the robustness of LPCA in delineating

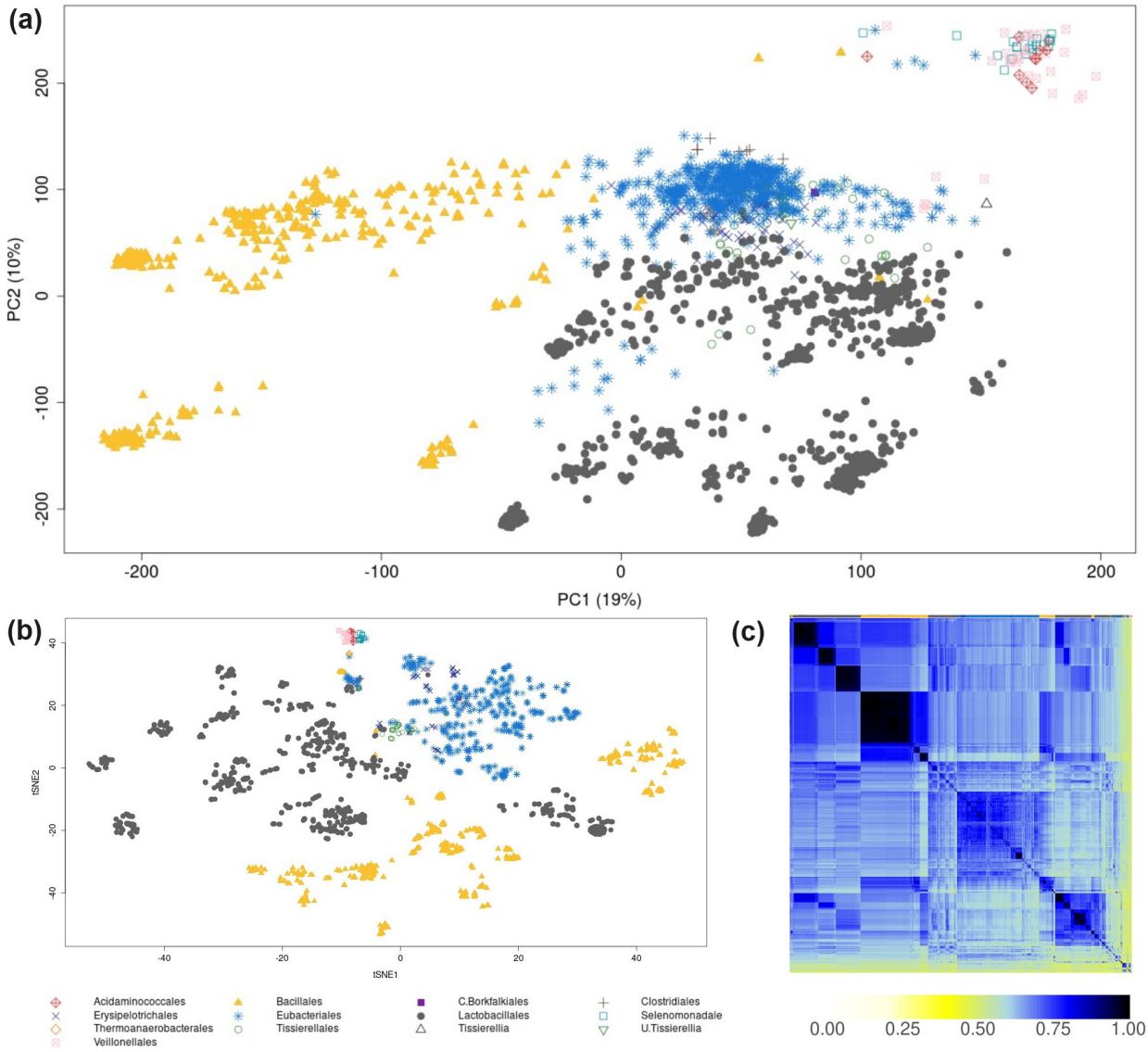


**Fig 7. LPCA (a), t-SNE (b) and Jaccard coefficients (c) derived from a binary reaction matrix from 2943 *Firmicutes* species-GSMMs.** In panels (a) and (b), points represent individual GSMMs, with different genera indicated by distinct symbols and colors. The top row in panel (c) uses these same colors to indicate the corresponding genera. The clustering of rows and columns in panel (c) was performed using the default hierarchical clustering settings (refer to methods section for details).

https://doi.org/10.1371/journal.pcbi.1012236.g007

clustering patterns. Utilizing LPCA on the set of differential reactions across GSMMs provides distinct advantages:

1. Enhanced Clade Discrimination: Unlike PCA relying on simulated growth rates from varied environmental conditions, LPCA exhibited clear separation of distinct clades. This method mitigates bias by solely assessing reaction presence in GSMMs, rather than subjectively selecting environmental conditions for growth rate simulations.
2. Subsystem Identification: In contrast to t-SNE and Jaccard coefficients, our LPCA approach provides precise information on drivers that govern separation and enables an efficient

subsystem analysis by grouping reactions based on their loading values. This may also support curation efforts and metabolic subsystem analysis.

3. Comparison to MLR: While MLR can identify driving factors, LPCA offers a significant advantage by simultaneously identifying potential subclusters. These subclusters may indicate metabolic variations that could be overlooked when relying solely on the initial phylogenetic classification. Additionally, the direction of reaction- or subsystem-centric loadings can be determined with LPCA, which is not possible with MLR.
4. Transcriptome-based Equivalence: Our findings showcased the ability of LPCA to recover cluster patterns observed in transcriptome-based PCA plots solely from genome-scale metabolic reconstruction and/or GSMMs.

Here, our focus was on grouping reactions by standard biochemical subsystems such as glycolysis, etc. However, LPCA can be extended to group reaction-centric loadings by any other pathways of interest. This expansion could complement recently developed tools designed to facilitate metabolic pathway analysis [48, 49]. In addition to other methods for feature selection from omics datasets [50–52], PCA loadings have been established as an effective method for this purpose, facilitating the identification of biologically significant features with high variance across different conditions or phenotypes [53]. LPCA loadings could be used in a similar way with respect to GSMMs. While LPCA is limited to binary datasets, our study demonstrates its effectiveness with binary reaction datasets derived from GSMMs, highlighting its potential to guide further research and pathway analysis in GSMMs. By identifying reactions with high LPCA loading values, we have pinpointed those that play pivotal roles in GSMMs, suggesting targets for further experimental and computational investigation. Integrating this approach with prior to other omics analyses could ultimately provide a more comprehensive understanding of metabolic pathways.

## Supporting information

**S1 Fig. LPCA (a), t-SNE (b) and Jaccard coefficients (c) derived from a binary reaction matrix from pan-reactions in 222 *Escherichia* GSMMs.** In panels (a) and (b), points represent individual GSMMs, with different genera indicated by distinct symbols and colors. The top row in panel (c) uses these same colors to indicate the corresponding genera. Circles in panel (a) highlight clusters of *E. albertii* strains (blue), *E. fergusonii* strains (red), and a mixed cluster of *E. coli*, *S. dysenteriae*, *S. flexneri*, and Clades II to VIII (orange). Labeled arrows in panel (a) denote subsystem-centric loading vectors from LPCA (refer to the results and methods section for definitions). The clustering of rows and columns in panel (c) was performed using the default hierarchical clustering settings (refer to methods section for details.
(TIF)

**S2 Fig. PCA based on simulated growth rates from 222 *Escherichia* species-specific GSMMs across 570 different environmental conditions [9].** While *E.fergusonii* GSMMs could be well separated based on simulated growth rates, the remaining clades seem to be less separatable.
(TIF)

**S3 Fig. Phylogenetic tree of 222 *Escherichia* species based on whole genomes.** Genomes were obtained from Enterobase [30]. Phylogenetic relations were obtained from OrthoFinder [32] based on coding genes. *E.coli*, *E.albertii*, and *E.fergusonii* formed distinct clades.
(TIF)

**S4 Fig. Reaction-specific loadings, grouped by subsystem from differential reactions.** The reaction "ALAALAD" (Murein Biosynthesis) was found to be a major driving factor for separation. "PRCOA1" (Cholesterol degradation) was found to be incorrectly assigned to "Histidine metabolism" in the original GSMMs.
(TIF)

**S5 Fig. LPCA scores (a) and t-SNE (b) results using the differential reaction dataset from *Escherichia*-GSMMs, performed 10 times.** While LPCA scores showed reproducible clustering, t-SNE resulted in a more diffuse clustering.
(TIF)

**S6 Fig. Phylogenetic tree of yeast-species taken from [24].** Whole-genome based phylogenetic comparison of strains results in more distinct separation of clades, compared to LPCA scores, t-SNE or hierarchical clustering based on Jaccard similarity.
(TIF)

**S7 Fig. PCA of transcriptomes from (a) whole transcriptomes, (b) "Human1" metabolic genes, and (c) selected genes from context-specific reconstructions.** While clustering between healthy and cancer tissue could be conserved based on whole transcriptomes (a) and metabolic genes (b), it was not possible based on the genes from the context-specific reconstructions (c).
(TIF)

**S1 Table. Top-ranked subsystems driving cluster separation in 222 *Escherichia* strains in LPCA.** Subsystems were ranked either according to the magnitude of the average loadings per subsystem for the first two principal components, $|\text{avg}(U^j)|$, or according to the average magnitude of the loadings per subsystem, $\text{avg}|U^j|$, see Eqs (2) and (3), respectively. † not within the top five in this ranking. ‡ not present in the original set of subsystems.
(XLSX)

**S2 Table. Reaction-centric loadings of the subsystem "Histidine metabolism" † In the original models, "PRCOA1" was associated with "Histidine metabolism".** We reassigned "PRCOA1" to the newly created metabolic subsystem "Cholesterol degradation". See text and S1 Table for details.
(XLSX)

## Author Contributions

**Conceptualization:** Leopold Zehetner, Diana Széliová, Juan A. Hernandez Bort, Jürgen Zanghellini.

**Data curation:** Leopold Zehetner.

**Formal analysis:** Leopold Zehetner.

**Funding acquisition:** Barbara Kraus, Juan A. Hernandez Bort.

**Investigation:** Leopold Zehetner.

**Methodology:** Leopold Zehetner.

**Project administration:** Juan A. Hernandez Bort, Jürgen Zanghellini.

**Resources:** Jürgen Zanghellini.

**Software:** Leopold Zehetner.

**Supervision:** Diana Széliová, Juan A. Hernandez Bort, Jürgen Zanghellini.

**Validation:** Leopold Zehetner.

**Visualization:** Leopold Zehetner, Diana Széliová.

**Writing – original draft:** Leopold Zehetner.

**Writing – review & editing:** Leopold Zehetner, Diana Széliová, Barbara Kraus, Juan A. Hernandez Bort, Jürgen Zanghellini.

## References


1. Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO. Genome-scale metabolic model of Helicobacter pylori 26695; 2002.
2. Smith AC, Robinson AJ. A metabolic model of the mitochondrion and its use in modelling diseases of the tricarboxylic acid cycle. BMC systems biology. 2011; 5(1):1–13. https://doi.org/10.1186/1752-0509-5-102 PMID: 21714867
3. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Molecular systems biology. 2014; 10(3):721. https://doi.org/10.1002/msb.145122 PMID: 24646661
4. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nature communications. 2014; 5(1):3083. https://doi.org/10.1038/ncomms4083 PMID: 24419221
5. Gotsmy M, Strobl F, Weiß F, Gruber P, Kraus B, Mairhofer J, et al. Sulfate limitation increases specific plasmid DNA yield and productivity in E. coli fed-batch processes. Microbial Cell Factories. 2023; 22 (1):242. https://doi.org/10.1186/s12934-023-02248-2 PMID: 38017439
6. Ergün BG, Berrios J, Binay B, Fickers P. Recombinant protein production in Pichia pastoris: from transcriptionally redesigned strains to bioprocess optimization and metabolic modelling. FEMS Yeast Research. 2021; 21(7):foab057. https://doi.org/10.1093/femsyr/foab057 PMID: 34755853
7. Fouladiha H, Marashi SA, Torkashvand F, Mahboudi F, Lewis NE, Vaziri B. A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. Bioprocess and Biosystems Engineering. 2020; 43:1381–1389. https://doi.org/10.1007/s00449-020-02332-6 PMID: 32211960
8. Richelle A, David B, Demaegd D, Dewerchin M, Kinet R, Morreale A, et al. Towards a widespread adoption of metabolic modeling tools in biopharmaceutical industry: a process systems biology engineering perspective. NPJ systems biology and applications. 2020; 6(1):6. https://doi.org/10.1038/s41540-020-0127-y PMID: 32170148
9. Monk JM. Genome-scale metabolic network reconstructions of diverse Escherichia strains reveal strain-specific adaptations. Philosophical Transactions of the Royal Society B. 2022; 377 (1861):20210236. https://doi.org/10.1098/rstb.2021.0236 PMID: 35989599
10. Lu H, Li F, Yuan L, Domenzain I, Yu R, Wang H, et al. Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection. Molecular Systems Biology. 2021; 17(10): e10427. https://doi.org/10.15252/msb.202110427 PMID: 34676984
11. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, et al. Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. Proceedings of the National Academy of Sciences. 2013; 110(50):20338–20343. https://doi.org/10.1073/pnas.1307797110 PMID: 24277855
12. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nature biotechnology. 2010; 28(3):245–248. https://doi.org/10.1038/nbt.1614 PMID: 20212490
13. Cabbia A, Hilbers PA, van Riel NA. A distance-based framework for the characterization of metabolic heterogeneity in large sets of genome-scale metabolic models. Patterns. 2020; 1(6). https://doi.org/10.1016/j.patter.2020.100080
14. Heinken A, Hertel J, Acharya G, Ravcheev DA, Nyga M, Okpala OE, et al. Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. Nature Biotechnology. 2023; p. 1–12. https://doi.org/10.1038/s41587-022-01628-0 PMID: 36658342
15. Lötsch J, Ultsch A. Current projection methods-induced biases at subgroup detection for machine-learning based data-analysis of biomedical data. International journal of molecular sciences. 2019; 21 (1):79. https://doi.org/10.3390/ijms21010079 PMID: 31861946

16. Gove R, Cadalzo L, Leiby N, Singer JM, Zaitzeff A. New guidance for using t-SNE: Alternative defaults, hyperparameter selection automation, and comparative evaluation. Visual Informatics. 2022; 6(2):87–97. https://doi.org/10.1016/j.visinf.2022.04.003

17. Ozgode Yigin B, Saygili G. Effect of distance measures on confidences of t-SNE embeddings and its implications on clustering for scRNA-seq data. Scientific Reports. 2023; 13(1):6567. https://doi.org/10.1038/s41598-023-32966-x PMID: 37085593

18. Song Y, Westerhuis JA, Aben N, Michaut M, Wessels LF, Smilde AK. Principal component analysis of binary genomics data. Briefings in bioinformatics. 2019; 20(1):317–329. https://doi.org/10.1093/bib/bbx119 PMID: 30657888

19. Greenacre M, Groenen PJ, Hastie T, d'Enza AI, Markos A, Tuzhilina E. Principal component analysis. Nature Reviews Methods Primers. 2022; 2(1):100. https://doi.org/10.1038/s43586-022-00184-w

20. Collins M, Dasgupta S, Schapire RE. A generalization of principal components analysis to the exponential family. Advances in neural information processing systems. 2001; 14.

21. Landgraf AJ, Lee Y. Dimensionality reduction for binary data through the projection of natural parameters. Journal of Multivariate Analysis. 2020; 180:104668. https://doi.org/10.1016/j.jmva.2020.104668

22. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: constraints-based reconstruction and analysis for python. BMC systems biology. 2013; 7:1–6. https://doi.org/10.1186/1752-0509-7-74 PMID: 23927696

23. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic acids research. 2016; 44(D1):D515–D522. https://doi.org/10.1093/nar/gkv1049 PMID: 26476456

24. Shen XX, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, et al. Tempo and mode of genome evolution in the budding yeast subphylum. Cell. 2018; 175(6):1533–1545. https://doi.org/10.1016/j.cell.2018.10.023 PMID: 30415838

25. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015; 347(6220):1260419.

26. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science. 2017; 357(6352):eaan2507. https://doi.org/10.1126/science.aan2507 PMID: 28818916

27. Robinson JL, Kocabaş P, Wang H, Cholley PE, Cook D, Nilsson A, et al. An atlas of human metabolism. Science signaling. 2020; 13(624):eaaz1482. https://doi.org/10.1126/scisignal.aaz1482 PMID: 32209698

28. Krijthe J, van der Maaten L, Krijthe MJ. Package 'Rtsne'. R package version 013. 2018;.

29. Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull Soc Vaudoise Sci Nat. 1901; 37:547–579.

30. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M, Brown D, et al. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. Genome research. 2020; 30(1):138–152. https://doi.org/10.1101/gr.251678.119 PMID: 31809257

31. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics. 2010; 11:1–11. https://doi.org/10.1186/1471-2105-11-119 PMID: 20211023

32. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome biology. 2019; 20:1–14. https://doi.org/10.1186/s13059-019-1832-y PMID: 31727128

33. Paradis E, Blomberg S, Bolker B, Brown J, Claude J, Cuong HS, et al. Package 'ape'. Analyses of phylogenetics and evolution, version. 2019; 2(4):47.

34. Sokal RR, Rohlf FJ. The comparison of dendrograms by objective methods. Taxon. 1962; p. 33–40. https://doi.org/10.2307/1217208

35. Lessard IA, Walsh CT. VanX, a bacterial D-alanyl-D-alanine dipeptidase: resistance, immunity, or survival function? Proceedings of the National Academy of Sciences. 1999; 96(20):11028–11032. https://doi.org/10.1073/pnas.96.20.11028 PMID: 10500118

36. Aráoz R, Anhalt E, René L, Badet-Denisot MA, Courvalin P, Badet B. Mechanism-based inactivation of VanX, a D-alanyl-D-alanine dipeptidase necessary for vancomycin resistance. Biochemistry. 2000; 39(51):15971–15979. https://doi.org/10.1021/bi001408b PMID: 11123924

37. Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, et al. BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic acids research. 2020; 48(D1):D402–D406. https://doi.org/10.1093/nar/gkz1054 PMID: 31696234

38. Roth M, Westrick N, Baldwin T. Fungal biotechnology: From yesterday to tomorrow. Front Fungal Biol. 2023; 4; 2023. https://doi.org/10.3389/ffunb.2023.1135263 PMID: 37746125

39. Shaath H, Toor S, Nair VS, Elkord E, Alajez NM. Transcriptomic analyses revealed systemic alterations in gene expression in circulation and tumor microenvironment of colorectal cancer patients. Cancers. 2019; 11(12):1994. https://doi.org/10.3390/cancers11121994 PMID: 31835892

40. Xu L, Wang R, Ziegelbauer J, Wu WW, Shen RF, Juhl H, et al. Transcriptome analysis of human colorectal cancer biopsies reveals extensive expression correlations among genes related to cell proliferation, lipid metabolism, immune response and collagen catabolism. Oncotarget. 2017; 8(43):74703. https://doi.org/10.18632/oncotarget.20345 PMID: 29088818

41. Kamal MV, Damerla RR, Dikhit PS, Kumar NA. Prostaglandin-endoperoxide synthase 2 (PTGS2) gene expression and its association with genes regulating the VEGF signaling pathway in head and neck squamous cell carcinoma. Journal of Oral Biology and Craniofacial Research. 2023; 13(5):567–574. https://doi.org/10.1016/j.jobcr.2023.07.002 PMID: 37559688

42. Xu L, Stevens J, Hilton MB, Seaman S, Conrads TP, Veenstra TD, et al. COX-2 inhibition potentiates antiangiogenic cancer therapy and prevents metastasis in preclinical models. Science translational medicine. 2014; 6(242):242ra84–242ra84. https://doi.org/10.1126/scitranslmed.3008455 PMID: 24964992

43. Vaz CV, Correia S, Cardoso HJ, Figueira MI, Marques R, Maia CJ, et al. The emerging role of regucalcin as a tumor suppressor: Facts and views. Current molecular medicine. 2016; 16(7):607–619. https://doi.org/10.2174/1566524016666160714124550

44. Ghanem NZ, Yamaguchi M. Regucalcin downregulation in human cancer. Life Sciences. 2024; p. 122448. https://doi.org/10.1016/j.lfs.2024.122448 PMID: 38246519

45. Yamaguchi M. Regucalcin Is a Potential Regulator in Human Cancer: Aiming to Expand into Cancer Therapy. Cancers. 2023; 15(22):5489. https://doi.org/10.3390/cancers15225489 PMID: 38001749

46. Li P, Xu Q, Liu K, Ye J. CRYL1 is a Potential Prognostic Biomarker of Clear Cell Renal Cell Carcinoma Correlated with Immune Infiltration and Cuproptosis. Technology in Cancer Research & Treatment. 2024; 23:15330338241237439. https://doi.org/10.1177/15330338241237439 PMID: 38497139

47. Fujii J, Homma T, Miyata S, Takahashi M. Pleiotropic actions of aldehyde reductase (AKR1A). Metabolites. 2021; 11(6):343. https://doi.org/10.3390/metabo11060343 PMID: 34073440

48. Schilling CH, Schuster S, Palsson BO, Heinrich R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnology progress. 1999; 15(3):296–303. https://doi.org/10.1021/bp990048k PMID: 10356246

49. Rezola A, Pey J, Tobalina L, Rubio Á, Beasley JE, Planes FJ. Advances in network-based metabolic pathway analysis and gene expression data integration. Briefings in bioinformatics. 2015; 16(2):265–279. https://doi.org/10.1093/bib/bbu009 PMID: 24626528

50. Li Y, Mansmann U, Du S, Hornung R. Benchmark study of feature selection strategies for multi-omics data. BMC bioinformatics. 2022; 23(1):412. https://doi.org/10.1186/s12859-022-04962-x PMID: 36199022

51. Bhadra T, Mallik S, Hasan N, Zhao Z. Comparison of five supervised feature selection algorithms leading to top features and gene signatures from multi-omics data in cancer. BMC bioinformatics. 2022; 23 (Suppl 3):153. https://doi.org/10.1186/s12859-022-04678-y PMID: 35484501

52. Taguchi Y, Turki T. Novel feature selection method via kernel tensor decomposition for improved multi-omics data analysis. BMC medical genomics. 2022; 15(1):37. https://doi.org/10.1186/s12920-022-01181-4 PMID: 35209912

53. Taguchi YH. Tensor decomposition-based and principal-component-analysis-based unsupervised feature extraction applied to the gene expression and methylation profiles in the brains of social insects with multiple castes. BMC bioinformatics. 2018; 19:1–13. https://doi.org/10.1186/s12859-018-2068-7

Metabolic Engineering

# Improving HEK293-based AAV-production using GSMMs, and a multi-omics approach

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | MBE-D-24-00393R2 |
| **Article Type:** | Research paper |
| **Keywords:** | GSMM; HEK293; AAV; bioprocess; HIF1alpha |
| **Corresponding Author:** | Diana Széliová, Ph.D.<br>University of Vienna<br>AUSTRIA |
| **First Author:** | Leopold Zehetner |
| **Order of Authors:** | Leopold Zehetner |
| | Diana Széliová, Ph.D. |
| | Barbara Kraus |
| | Juan Antonio Hernandez Bort |
| | Juergen Zanghellini |
| **Abstract:** | HEK293 cells are a versatile cell line extensively used in the production of recombinant proteins and viral vectors, notably Adeno-associated virus (AAV). Despite their high transfection efficiency and adaptability to various culture conditions, challenges remain in achieving sufficient yields of active viral particles. This study presents a comprehensive multi-omics analysis of two HEK293 strains under good manufacturing practice conditions, focusing on the metabolic and cellular responses during AAV production. The investigation included lipidomic, exometabolomic, and transcriptomic profiling across different conditions and time points. Genome-scale metabolic models (GSMMs) were reconstructed for these strains to elucidate metabolic shifts and identify potential bottlenecks in AAV production. Notably, the study revealed significant differences between a High-producing (HP) and a Low-producing (LP) HEK293 strains, highlighting pseudohypoxia in the LP strain. Key findings include the identification of hypoxia-inducible factor 1-alpha (HIF1α) as a critical regulator in the LP strain, linking pseudohypoxia to poor AAV productivity. Inhibition of HIF1α resulted in immediate cessation of cell growth and a 2.5-fold increase in viral capsid production, albeit with a decreased number of viral genomes, impacting the full-to-empty particle ratio. This trade-off is significant because it highlights a key challenge in AAV production: achieving a balance between capsid assembly and genome packaging to optimize the yield of functional viral vectors. Overall this suggests that while HIF1α inhibition enhances capsid assembly, it simultaneously hampers nucleotide synthesis via the pentose phosphate pathway (PPP), necessary for nucleotide synthesis, and therefore for AAV genome replication. |
| **Response to Reviewers:** | |

# Improving HEK293-based AAV-production using GSMMs, and a multi-omics approach

L. Zehetner[a,b,c,*,1], D. Széliová[b,*], B. Kraus[a,e], J. A. Hernandez Bort[d,*] and J. Zanghellini[b,*]

[a]Baxalta Innovations GmbH, Part of Takeda Companies, Uferstrasse 15, Orth, 2304, Austria
[b]Department for Analytical Chemistry, University of Vienna, Vienna, 1090, Austria
[c]Doctoral School of Chemistry, University of Vienna, Vienna, 1090, Austria
[d]Department of Applied Life Sciences, Bioengineering, University of Applied Sciences Campus Vienna, Vienna, 1100, Austria
[e]Institute of Molecular Biotechnology, Institut für Molekulare Biotechnologie GmbH, Vienna, 1030, Austria

ARTICLE INFO

*Keywords*:
GSMM
HEK293
AAV
Bioprocess
HIF1$\alpha$

ABSTRACT

HEK293 cells are a versatile cell line extensively used in the production of recombinant proteins and viral vectors, notably Adeno-associated virus (AAV) [72]. Despite their high transfection efficiency and adaptability to various culture conditions, challenges remain in achieving sufficient yields of active viral particles. This study presents a comprehensive multi-omics analysis of two HEK293 strains under good manufacturing practice conditions, focusing on the metabolic and cellular responses during AAV production. The investigation included lipidomic, exometabolomic, and transcriptomic profiling across different conditions and time points. Genome-scale metabolic models (GSMMs) were reconstructed for these strains to elucidate metabolic shifts and identify potential bottlenecks in AAV production. Notably, the study revealed significant differences between a High-producing (HP) and a Low-producing (LP) HEK293 strains, highlighting pseudohypoxia in the LP strain. Key findings include the identification of hypoxia-inducible factor 1-alpha (HIF1alpha) as a critical regulator in the LP strain, linking pseudohypoxia to poor AAV productivity. Inhibition of HIF1alpha resulted in immediate cessation of cell growth and a 2.5-fold increase in viral capsid production, albeit with a decreased number of viral genomes, impacting the full-to-empty particle ratio. This trade-off is significant because it highlights a key challenge in AAV production: achieving a balance between capsid assembly and genome packaging to optimize the yield of functional viral vectors. Overall this suggests that while HIF1alpha inhibition enhances capsid assembly, it simultaneously hampers nucleotide synthesis via the pentose phosphate pathway (PPP), necessary for nucleotide synthesis, and therefore for AAV genome replication.

## Introduction

Human embryonic kidney 293 (HEK293) is a highly adaptable cell line originating from human embryonic kidney cells cultivated in tissue culture. Developed in 1977 [73], HEK293 cells have become pivotal in the generation of recombinant proteins [74, 6, 53, 62] and viral vectors [72, 61, 55, 50], owing to their high transfection efficiency and robust proliferation. At least 564 distinct HEK293 strains are documented [63], each optimized for specific applications. These cells are widely used in the biopharmaceutical sector to produce recombinant proteins not effectively expressed in Chinese Hamster Ovary (CHO) cells [52]. A significant advantage of HEK293 cells is their adaptability to diverse culture conditions; they are capable of growing in both suspension and adherent cultures [53, 54] and are suitable for serum-free media, which facilitates scalable and reproducible production processes.

HEK293 cells are widely used for producing viral vectors essential for cell and gene therapy [59, 4, 45], and vaccine development [99, 98]. These vectors include lentiviruses, adenoviruses, and AAV [72], with AAV especially prominence in *in vivo* gene therapy, leading to six approved drugs in recent years [55]. Despite their therapeutic potential, producing sufficient active viral particles remains challenging due to high associated costs, and the fact that only 10% of the AAV contain the gene of interest in full length [57, 56]. Thus, many studies have aimed at enhancing AAV titers in HEK293 cells [59, 58]. Beyond optimizing process parameters, several studies have examined the cellular behavior of HEK293 cells using single-omics datasets during growth [53, 6, 97] or production [52, 61, 6, 10, 60, 44, 4], as summarized in [62].

The behavior of host cells during AAV production has been extensively studied using transcriptomes [44, 4, 45], proteomes [61, 60], or both [51] (Tab 1), as recently summarized [189]. Despite variations in fermentation conditions, strains, and AAV serotypes, two pathways were consistently enriched during production: (i) endoplasmic reticulum (ER) stress and the unfolded protein response (UPR) [45, 4, 44, 51, 61], and (ii) immune response [60, 45, 44, 51]. ER stress occurs when unfolded or misfolded proteins accumulate in the ER, triggering the UPR to restore ER homeostasis. Key proteins involved in UPR activation include GRP78, ATF6, PERK, and IRE1. PERK phosphorylates *eIF2$\alpha$*, thereby inhibiting CAP-dependent protein translation [82]. ER stress and UPR are common challenges in recombinant protein production [80, 81], so their upregulation in HEK293 during AAV production is expected. Additionally, transcriptomic analyses identified the JAK-STAT signaling pathway as central during

*Corresponding authors, LZ: Leopold.Zehetner@univie.ac.at; DS: diana.szeliova@univie.ac.at; JAHB: juan_antonio.hernandez_bort@fh-campuswien.ac.at; JZ: juergen.zanghellini@univie.ac.at
ORCID(s): 0009-0003-2677-3685 (L. Zehetner); 0000-0002-9885-9758 (D. Széliová); 0000-0003-2348-4864 (J.A. Hernandez Bort); 0000-0002-1964-2455 (J. Zanghellini)

**Table 1**
**Overview of recently performed -omics based studies of HEK293 strains during AAV production**

| HEK293 strain | Medium | Set up | Data | Reference |
|---|---|---|---|---|
| ATCC: CRL-1573 | AMBIC | shake flask | Transcriptome | [44] |
| Mass Biologics, Inc. | BalanCD | shake flask | Transcriptome | [44] |
| unknown origin | FreeStyle F17 | 50L bioreactors | Transcriptome | [79] |
| FreeStyle 293F | BalanCD | shake flask | Proteome | [61] |
| Cell Biolabs, Inc. | DMEM (+10% FBS) | T-flask | Transcriptome, Proteome | [51] |
| Expi293 | HE400AZ | Erlenmeyer flask | Transcriptome | [60] |
| ATCC: CRL-1573 (LP) | FreeStyle F17 | 250mL bioreactors | Transcriptome | [4] |
| HP | FreeStyle F17 | 250mL bioreactors | Transcriptome | [4] |

AAV production in HEK293 strains [44, 45]. This pathway mediates the innate immune response to viral infections and regulates genes involved in antiviral defense, inflammation, cell growth, and differentiation [102, 101, 100]. Inhibiting *STAT1* with Ruxolitinib led to a two-fold titer increase in shake flasks and a 50% improvement in a 3 L bioreactor [45]. However, in certain HEK293 strains [45], the JAK-STAT pathway was upregulated even before transfection, indicating a pre-existing stressed state, while in others, it was either triggered by AAV production [44] or downregulated by ethanol addition [60]. The cause of this immune response, whether due to HEK293 strain adaptation or AAV particle production, remains unclear.

Differences in immune responses prior to transfection could arise from variations between suspension and adherent culturing of HEK293 strains. **(author?)** [46] recently highlighted metabolic differences after adapting an adherent HEK293 strain to suspension culture. The study found increased glucose uptake, lactate secretion, and succinate production in HEK293 cells adapted to suspension culture, suggesting that these cells experience pseudohypoxia [35]. In CHO cells, reducing pseudohypoxia by inhibiting pyruvate dehydrogenase kinase with dichloroacetate has been shown to improve culture performance, and antibody production [49].

Viral vectors may influence host cell metabolism more directly, similar to how viruses reprogram host metabolism to support their replication [111]. This is especially relevant for AAV production, which relies on helper functions from other viruses, like Adenovirus [115, 116], or Herpesvirus [113, 114]. While the impact of viral genes, such as *E4ORF1* from Human adenovirus [112], on host metabolism has been studied, the metabolic effects during AAV replication remain largely unexplored and could be investigated using genome-scale metabolic modeling [50].

GSMMs are mathematical representations of cellular metabolism, first developed over two decades ago [70]. Since then, more than 1000 GSMMs have been reconstructed across all domains of life [71]. In bioprocess engineering, GSMMs have provided deep insight at a cellular level leading to improved growth and production behavior in microbes [196, 197, 198, 85, 86, 25, 88] and mammalian cells [84, 89, 90]. However, currently, there are no GSMM for HEK293 cells during Recombinant protein (RP) or viral particle production. Existing HEK293 models are small-scale, focusing on selected pathways or AAV production via triple transfection [50]. The most detailed HEK293 GSMM includes 357 reactions and was manually curated for RP production [10] using transcriptomic and exometabolomic data [6] based on the Recon2 model [64].

Meanwhile, reconstructions of human metabolism have been continuously refined and updated [12, 65], allowing the use of computational tools to automatically generate cell-specific GSMMs [69, 68, 67, 66, 7]. However, obtaining accurate condition-specific GSMMs, requires integrating additional constraints, such as enzyme limitations [109], thermodynamic constraints [110], and constraints on exchange reactions [9, 106] or biomass composition [9, 107, 108]. Experimental data on exchange rates and biomass' dry weight, are pivotal but can disrupt simulations [9] if not precisely measured [106]. Similarly, macromolecular biomass composition, particularly the lipidome, varies under different environmental conditions and strongly influences predicted flux distributions [107, 108]. Thus, understanding exometabolites and biomass composition–especially the lipidome–is essential for modeling viral vector production. Unfortunately, such data are currently unavailable for AAV producing HEK293 cells.

In this study, we conducted a multi-omics analysis of two HEK293 strains, which were partially examined in previous research for AAV production [4, 45]. Using the latest reconstruction of human metabolism [12], we developed 20 condition-specific GSMMs from the two strains, analyzing mock and transfected conditions across five time points with four biological replicates each. We find that over 99.8% of cellular resources in both stains are used for growth rather than AAV production. To reallocate the resources, we aimed to inhibit growth and therefore increase AAV production in the LP strain. With the multi-omics analysis – encompassing lipidomes, transcriptomes, and fluxomes – we identified HIF-1$\alpha$ as a potential target in the LP strain, suggesting that the LP strain is affected by pseudohypoxia due to recent adaption from adherent to suspension culturing. Inhibiting HIF-1$\alpha$ halted cellular growth and improved the specific productivity of viral capsids by over two-fold. However, the number of viral genomes decreased following HIF-1$\alpha$ inhibition, leading to a less favorable full-empty ratio. Despite this, our study provides a comprehensive multi-omics analysis and

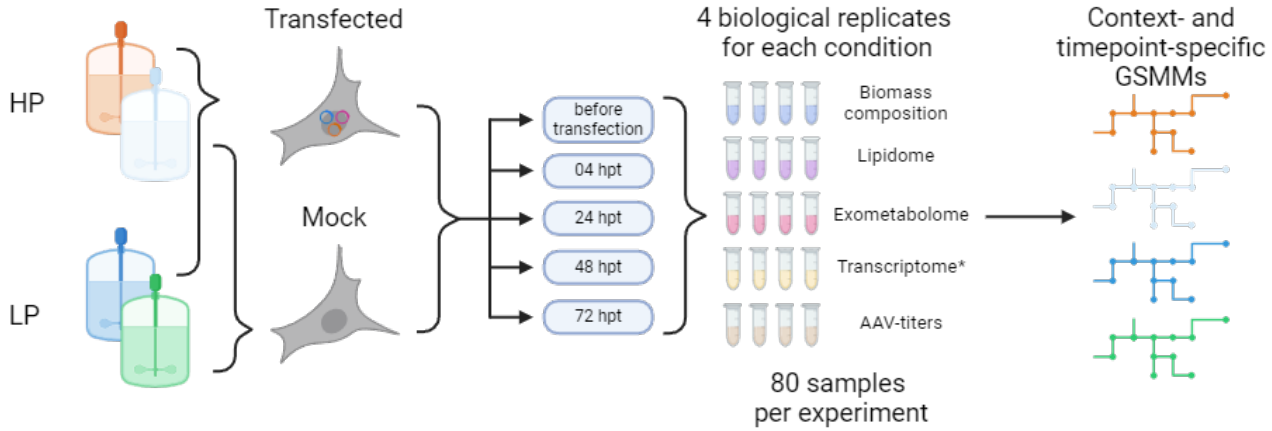


**Figure 1: Graphical overview of the methods section.** All samples were measured as described in the methods. Transcriptomic data were obtained from [4]

represents a significant advancement in data-driven process design. Overall, our study provides valuable insights into the metabolic behavior of different HEK293 strains during AAV production. Rather than employing GSMMs as predictive tools, we utilized them as platforms for data integration and analysis. Despite certain limitations, this work offers a comprehensive multi-omics analysis and constitutes an important advancement toward data-driven process design.

# Materials and Methods

## Bioprocess and Sampling

The bioprocess from cell culture to fermentation was performed as described recently by [4]. Briefly, two distinct HEK293 suspension cell line strain stocks, stored at -130 °C, were cultured at 37 °C in chemically defined, serum-free medium (FreeStyle F17, ThermoFisher, NY, USA) supplemented with 8 mM L-Glutamine and 1.0 g $L^{-1}$ Lutrol. The cultivation was carried out in a HERA Cell 150 (Thermo Fisher Scientific) incubator with 5% $CO_2$. To enhance statistical analysis, each cell line was cultivated in four biological replicates using Single Use Spinner flasks (Corning, Germany), with volumes ranging from 61 mL to 1.600 mL during the cell expansion phase. After 12 days, the cells were transferred to an ambr250 Modular fermentation system (Sartorius, Germany) under controlled conditions (pH, $CO_2$, $O_2$) for transient AAV 8 particle production. Prior to transfection, cells were diluted to a uniform density of $4.0 \cdot 10^6$ cells per mL to facilitate comparison. Each biological replicate was divided into two fermentation vessels: one for plasmid transfection and the other for mock transfection (without plasmids). A triple-plasmid transfection was performed using Polyethylenimine (Merck KGaA, Darmstadt, Germany), containing Adenovirus 5 Helper genes, Rep2Cap8, and a transgene sequence, in accordance with the supplier's instructions. Post-transfection, cells were cultured for 72 hours at 37 °C and sampled for analysis at five different time points: before (0 Hours post transfection (HPT)) and after transfection (4, 24, 48, and 72 HPT), resulting in a total of 80 samples (see Fig 1). Additional samples were taken at 21, 27, 45, and 51 HPT for quantification of lactate, ammonia (BioHT), proline, and glycine.

## Experimental Analyses

### *Biomass and cell count*

As described in [9], 1 mL of cell suspension was centrifuged at 200 g for 10 minutes at 4 °C. The pellet was then washed with 1 mL of pre-cooled phosphate-buffered saline (PBS, 1 M, 4 °C). To protect the cells during subsequent steps, Pluronic™ F-68 Non-ionic Surfactant (Gibco™, MA, USA) was added to the suspension at a final concentration of 2% (v/v). Glass beakers and silica beads were prepared by drying them overnight at 90 °C. They were subsequently cooled in a vacuum desiccator for 20 minutes before being weighed to ensure accuracy. Each sample, resuspended in a 1 M PBS solution, was transferred to the pre-weighed beakers. The samples were then dried at 90 °C until the biomass remained constant, ensuring complete removal of moisture. Before each weighing, the beakers were cooled for 20 minutes in the vacuum desiccator to prevent heat-related measurement inaccuracies. Cell counts were performed using a NucleoCounter202 (Allerod, Denmark), adhering strictly to the manufacturer's protocol.

### *Cellular glucose content*

In alignment with the methodology described by [9], we posited glucose as the predominant carbohydrate component within the cells and thus limited our carbohydrate quantification to glucose. For this purpose, we employed the Total Carbohydrate Quantification Assay Kit (ab155891, Abcam, Cambridge, UK). Beginning with cell pellets containing $3 - 10 \cdot 10^6$ cells, we adjusted the volume of lysis buffer proportionately. Sample preparation was carried out following the manufacturer's instructions. Each sample was analyzed in technical triplicates to enhance the reliability of our measurements.

### *Cellular DNA and RNA content*

DNA and RNA content was extracted from the same sample using the AllPrep DNA/RNA Mini Kit (Qiagen, Venlo, NE). As suggested by the manufacturer, $\beta$-mercaptoethanol (Thermo Fisher Scientific Inc., Waltham, MA, US) was added to inhibit RNases. Samples were quantified on a Nanodrop 2000 (Thermo Fisher Scientific Inc., Waltham, MA, US) in technical triplicates.

### *Cellular protein content*

Cell pellets were lysed by adding lysis buffer (PreOmics, Martinsried, GER), followed by incubation on a preheated thermocycler (95 °C, 10 min, 1000 rpm). Protein content was then quantified using the Qubit protein broad range assay kit (Thermo Fisher Scientific Inc., Waltham, MA, US) following the manfacturer's instructions.

### *Cellular lipidome*

For extraction, one aliquote in 2 mL Eppendorf vessel, each containing $3 - 10 \cdot 10^6$ cells were used. 205 µL of cold Methanole (Thermo Fisher Scientific Inc., Waltham, MA, US) were added to the cells, and they were vortexed for 20 s. As labeled internal lipid standards, 10 µL Splash Mix undiluted, 10 µL Ceramide Mix undiluted, and 10 µL Cardiolipin Mix undiluted were added (Avanti Polar Lipids, Alabaster, AL, US) to each sample. Further, the samples were allowed to thaw at room temperature and were sonicated for 10 min at 4 °C. Then, 750 µL of Methyl-tert-butyl-ether (Thermo Fisher Scientific Inc., Waltham, MA, US) were added, and the mixture was incubated for 1 h at room temperature under agitation. Phase separation was induced by adding 188 µL water with 0.1% ammonium formate (Thermo Fisher Scientific Inc., Waltham, MA, US). The extract was centrifuged at 10,000 g for 5 min, and two 200 µL aliquots of the upper phase were collected in another 2 mL Eppendorf vessel, dried in a vacuum centrifuge, and stored if necessary at -20 °C. For further analyses, samples were dissolved in 200 µL 2-Propanole/Methanole/Chloroform (4:2:1, v/v/v) containing 7.5 mM ammonium formate. LC-MS measurements were performed as described by [1]. A high field Q Exactive HF™ quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) was connected with a robotic nanoflow ion source TriVersa NanoMate® (Advion BioSciences, Ithaca NY, USA) and a nanoelectrospray ionization (nanoESI) chip with spraying nozzles of 5 µm nominal internal diameter. Samples were placed in a 96 twin.tec® well plate (Eppendorf, Hamburg, Germany). Following settings were applied in the Chipsoft 8.3.1 software (Advion BioSciences): ionization voltage 1.25 kV (pos)/ -1.25 kV (neg); backpressure 0.9 psi and in the MS source parameters: capillary temperature 250 °C, S-Lens radio frequency level 50. A 9 min polarity switching method with data independent acquisition (DIA) and MS1 scans was used with the following settings: resolution 240,000 (MS1), 60,000 (MS2), AGC target 1e6 (MS1), 2e5 (MS2), maximum IT 150 ms (MS1), 130 (MS2), scan range in MS1 m/z 350-1,050 (pos)/ m/z 200-1,200 (neg), NCE for MS2 of 21 (pos)/ 26 (neg) and a fixed first mass of m/z 80 (pos)/ m/z 150 (neg). LipidXplorer 1.2.8 settings were the following: mass tolerance 5 ppm, min. occupation of 0. In positive mode: intensity threshold 70,000 (MS1)/ 1,000 (MS2), resolution 230,000 (MS1)/ 60,000 (MS2), resolution gradient -180 (MS1)/ -40 (MS2). In Negative mode: intensity threshold 60,000 (MS1)/ 1,000 (MS2), resolution 260,000 (MS1)/ 80,000 (MS2), resolution gradient -170 (MS1)/ -100 (MS2). Identified lipids were removed if the absolute mass error was above 3, if isobaric compounds were present and if the concentration was below LOQ (limit of quantification). LOQ values were calculated by multiplying the standard deviation of 5 repetitive injections of a low concentrated Splash ISTD with 10. For quality control, a 10 µL plasma sample aliquot, which was placed at RT in a 2 mL Eppi, 290 µL Methanol, 10 µL Splash Mix I (undiluted), 1.000 µL methyl-tert-butyl-ether were added. The mixture was shaken for 1 h at RT. Phase separation was induced by adding 250 µL of MS-grade water. Upon 10 min of incubation at room temperature, the sample was centrifuged at 1,000 g for 10 min. The upper (organic) phase was collected in another 2 mL Eppi and dried in the SpeedVac. The dried lipids were dissolved in 200 µL isopropyl alcohol for LC-MS injection.

### *Estimated metabolite pool from human1 biomass reaction*

To partially account for the undefined part of the biomass composition, we obtained the metabolites of the metabolites_pool from human1 [12], and calculated the amount estimated amount of metabolites of the measured dry weight, as described in [9].

### *Spent media analysis*

**LC-MS based metabolomics** 1 mL suspension was centrifuged (200 g, 8 min, 4 °C), from which 100 µL liquid media was placed into a 2 mL Eppendorf vessel, and kept on ice. 50 µL of yeast internal standard (ISTD) (extract out of 2 billion cells reconstituted in 2 mL water), and 600 µL methanol were added to reach a final volume of 750 µL (80% methanol, v/v), which resulted in a 1:7.5 dilution of the sample and a 1:15 dilution of the ISTD. After thorough vortexing, it was kept on ice for 30 min. Then it was vortexed again and kept at -20 °C overnight. The samples were centrifuged (14,000 g, 4 °C, 15 min) and two 200 µL aliquots were dried in a vacuum centrifuge. They were stored if necessary at -20 °C. Shipping should be done on dry ice. At least $20 \cdot 50$ µL aliquots of the yeast ISTD (extract out of 2 billion cells reconstituted in 2 mL water) should be placed into 2 mL Eppendorf vessels for internal standardization of the calibrators and QC samples. These samples were also dried in a vacuum centrifuge. All samples were dissolved in 200 µL 20% water/80% acetonitrile (v/v). Vigorously vortexed for 3 minutes to ensure complete dissolving of the sample and centrifuged (14,000 g, 4 °C, 10 min) prior to being transferred to an HPLC vial. LC-MS measurements were performed as described in [1]. A Vanquish™ Horizon HPLC (Thermo Fisher Scientific) with an iHILIC®-(P) Classic (2.1 mm × 100 mm, 5 µm, HILICON, Umea, Sweden) was

used for hydrophilic interaction liquid chromatography-high resolution mass spectrometry (HILIC-HRMS). The flow rate was set to 200 µL min-1 and the column temperature to 40 °C. Solvent A was 90% 15 mM ammonium acetate (pH 9.4), 10% acetonitrile, and solvent B was 90% acetonitrile, 10% 15 mM ammonium acetate (pH 9.4). The following gradient was used: 0-12 min ramp from 100% B to 20% B, 12-14 min 20%, 14-17 min ramp to 0% B, 17-19 min 0% B and 20-27 min 100% B as equilibration step. The injection volume was 5 µL and the injector needle was washed with acetonitrile: methanol: water 1:1:1 (v/v/v) for 5 s after each injection. A high field Q Exactive HF™ quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) was used as MS. The following source parameters were applied in both polarities in polarity switching mode: capillary temperature of 280 °C, sheath gas flow rate of 40, an auxiliary flow rate of 3, sweep gas of 0, S-lens RF level of 30 and auxiliary gas heater temperature of 320 °C applying a spray voltage of 3.5 kV in positive mode and 2.8 kV in negative mode. In MS1 mode the mass range in both polarities was set to m/z 60-900 A maximum IT of 100 ms, a resolution of 120,000, and an AGC target of $3\dot{1}0^6$ was applied. Skyline (version 22.2.0.312) was used for peak integration and R/ R studio for final data processing. Identified lipids were removed if the absolute mass error was above 3 and if the concentration was below LOQ. LOQ was defined as the lowest standard concentration in the linear range. For calibration one dried ESTD stock solution is dissolved in 200 µL water for a final conc. of 50 µM. 14 calibrants with concentrations from 25 µM to 0.001 µM were prepared. The internal standard is already dried in the Eppis, therefore the standards are vortexed, centrifuged, and transferred to the well plate.

**Proline and Glycine** Results for proline and glycine showed inconsistent results. Therefore, these amino acids were quantified using Waters AccQ tagging (Milford, MA, USA) at nine different time points. Briefly, spent media samples were filtered by centrifugation (10 min, RT, 12,000 g) using 3K Amicon (Merck KGaA, Darmstadt, GER) filters to remove proteins. The flowthrough was diluted 1:10 (MilliQ Water) and labeled together with calibration standards according to the manufacturer's instructions. Norvaline was used as internal standard for normalization. Waters Acquity (UPLC03, Milford, MA, USA) was used as HPLC system. The mobile phase consisted of Eluent A (1:10 diluted in MilliQ Water) and Eluent B (undiluted). For every sample the runtime was set to 11 minutes with a flowrate of 0.7 mL/min. The gradient of Eluent B was increased from 0.1% at the beginning to 90%.

**Lactate and Ammonia** Lactate and Ammonia were quantified from supernatant for 9 time points. Samples were centrifuged (10 min, 200 g, 4 °C) and measured on a BioHT (Roche, Switzerland), following the manufacturer's instructions. From spent media analyses, limiting compounds were determined, that were fully consumed first. These limiting compounds were then added as supplements to improve the bioprocess.

### *Quantification of AAV*

**Droplet digital PCR** As described by [4], for quantifying vector genomes, a droplet digital PCR method from Bio-Rad was employed using the fully automated QX One System (Hercules, CA, USA), which allows for absolute quantification of vector genomes without the need for standard curves. The sample is partitioned into oil droplets, with each droplet serving as an independent PCR reaction compartment. During the initial phase of PCR in the thermal cycler, capsids are degraded, making the DNA accessible for amplification. The vector genome titer is then determined using a droplet reader. To remove extraneous DNA sequences, samples were treated with DNase I (NEB, Ipswich, MA, USA). Prior to this treatment, samples were pre-diluted to enhance the efficiency of DNase I activity. Following droplet generation, PCR was conducted with Bio-Rad ddPCR Supermix (no dUTP) and transgene-specific primers and probe.

**ELISA** As described by [4], an AAV8 titration ELISA kit (Progen, Heidelberg, GER) was used for quantifying AAV8 capsids, utilizing a monoclonal antibody (ADK8) that recognizes a specific conformational epitope on the assembled AAV8 capsids. The captured AAV 8 particles were identified by the binding of biotinylated anti-AAV 8 ADK8, as this epitope is consistently presented on the AAV 8 capsid structure. Streptavidin peroxidase and a peroxidase substrate were subsequently applied to measure the amount of bound anti-AAV 8, thus indicating the concentration of AAV 8 capsids. The resulting color change was measured at 450 nm using a photometer. The kit also includes an AAV 2/8 particle preparation with a known concentration of labeled AAV 8 particles for calibration purposes.

## Computational Analyses

### *Lipidome analysis*

Principal component analysis (PCA) was performed on the lipidomic dataset was performed using the prcomp function in RStudio with the parameters center, and scale set to TRUE. Prior to that, every sample was normalized by dividing each lipid by the total amount of lipid in this sample. Loadings, obtained from PCA, were grouped by lipid class. The five longest loading vectors (lipid classes) were displayed in the PCA plot as well to identify the drivers behind the separation of the samples.

### *Calculation of experimental growth rates during non-exponential growth and production phase*

The (specific) growth rate was calculated by:

$$\mu_t = \frac{1}{x_t}\frac{\mathrm{d}x_t}{\mathrm{d}t} \tag{1}$$

where $\mu_t$ is the growth rate at time point $t$, $x_t$ is the biomass value derived from the fitted function $x$, and $\dot{x}_t$ is the derivative of $x_t$. A logistic function

$$x = \frac{a}{1 + \exp\left[-b(t-c)\right]} \tag{2}$$

was used for fitting the biomass.

The error of the growth rate was calculated by:

$$\mu_t^{err} = \sqrt{\left(\frac{\dot{x}_t^{err}}{x_t}\right)^2 + \left(-\mu_t \frac{x_t^{err}}{x_t}\right)^2} \tag{3}$$

where $\mu_t^{err}$ is the error of the growth rate at time point $t$, $\dot{x}_t^{err}$ is the error of the derivative value, and $x_t^{err}$ is the error of the fitted biomass value. $x_t^{err}$ and $\dot{x}_t^{err}$ were obtained by:

$$x_t^{err} = \sqrt{\left(\sum_{j=1}^{3} \frac{\partial x_t}{\partial \theta_j} \cdot \theta_j^{err}\right)^2} \tag{4}$$

and

$$\dot{x}_t^{err} = \sqrt{\left(\sum_{j=1}^{n} \frac{\partial \dot{x}_t}{\partial \theta_j} \cdot \theta_j^{err}\right)^2} \tag{5}$$

where $\theta_j$ is a parameter in the logistic fitting function $x$, and $\theta_j^{err}$ is its corresponding error value from the fitting function $h$.

Derivatives were calculated in R, using the package Deriv (v4.1) [91]

### *Calculation of exchange rates*

From the analysis of spent media, we identified 32 metabolites based on their concentration profiles, from which exchange rates could be calculated. These metabolites were categorized into three groups according to the fitting function applied. Specifically, 27 metabolites were fitted using an exponential function, 4 were fitted using a quadratic function, and 1 metabolite was fitted using a cubic function, employing robust fitting techniques (R function nlrob). To determine the exchange rates at each timepoint post-transfection, the derivative of the fitted function was calculated. Error propagation was then performed following the appropriate equation to ensure accurate quantification of exchange rates.

The exchange rate $q_{i,t}$ of metabolite $i$ was calculated as

$$q_{i,t} = \frac{\dot{f}_{i,t}}{x_t} \tag{6}$$

where $\dot{f}_{i,t}$ is the derivative value of the fitted function $f_i$ for metabolite $i$ at timepoint $t$. The error of $q_{i,t}$ is calculated by error propagation following:

$$q_{i,t}^{err} = \sqrt{\left(\frac{1}{x_t} \cdot \dot{f}_{i,t}^{err}\right)^2 + \left(-\frac{\dot{f}_{i,t}}{x_t^2} \cdot x_t^{err}\right)^2} \tag{7}$$

where $q_{i,t}^{err}$ is the error of $q_{i,t}$ for metabolite $i$ at timepoint $t$, $\dot{f}_{i,t}^{err}$ is the error obtained from the derivative $\dot{f}_i$ of the fitted function $f_i$. The derivative $\dot{f}_i$ of the fitted function $f_i$ was obtained by

$$\dot{f}_i = \frac{df_i}{dt} \tag{8}$$

The corresponding error $\dot{f}_{i,t}^{err}$ was calculated by:

$$\dot{f}_{i,t}^{err} = \sqrt{\left(\sum_{j=1}^{n} \frac{\partial \dot{f}_{i,t}}{\partial \theta_j} \cdot \theta_j^{err}\right)^2} \tag{9}$$

where $\partial \dot{f}_{i,t}$ is the partial derivative of the derivative $\dot{f}_{i,t}$ for metabolite $i$ at timepoint $t$ with respect to the parameter $\theta_j$. $\theta_j^{err}$ is the obtained error of every parameter $\theta_j$ in the fitted function $f_i$ with $n$ parameters. To obtain fitted functions $f_i$, the nlrob function in R was used to fit one of the three different functions:

$$f_{1,i} = a \cdot b^t \tag{10}$$

for metabolites, showing an exponential concentration profile,

$$f_{2,i} = a \cdot t^2 + b \cdot t + c \tag{11}$$

for metabolites, showing a quadratic concentration profile (lactate, proline, and glycine), or,

$$f_{3,i} = a \cdot t^3 + b \cdot t^2 + c \cdot t + d \tag{12}$$

for ammonium, showing a cubic concentration profile.

To consider natural degradation rate of glutamine, media was sampled daily. Glutamine and ammonia were quantified on the BioHT (Roche, Switzerland). The natural degradation rate was calculated using an exponential fit $f_{1,i}$. The obtained depletion rate was subtracted from the derivative value $\dot{f}_{i,t}$ for glutamine or added to the derivative value $\dot{f}_{i,t}$ for ammonia.

### *Reconstruction of context-specific models*

To reconstruct context-specific GSMMs, the most recent GSMM of the human metabolism, human1 [12], was used. Initially, the uptakes of 42 compounds were identified to enable feasible solutions of the GSMM. These compounds were selected based on a chemically defined medium (22400 - RPMI 1640, HEPES, Thermo Fisher Scientific Inc., Waltham, MA, USA), since the exact composition of the used F17 expression media is currently not available. Additionally, all compounds identified by [5] that are not essential for growth were included as potential secretion reactions, totaling 23 compounds. Moreover, metabolites for which exchange rates could be calculated (as detailed in the subsequent section) were incorporated into the model as constraints. The compounds in the growth medium were assigned exchange reaction boundaries of [-1,000, 0], while the boundaries for all other exchange reactions were set to [0, 1,000].

To obtain context-specific GSMMs, transcriptomic data from the exponential growth phase were utilized as reported in [6]. This transcriptomic dataset, derived from a non-producing strain, encompasses four time points during exponential growth, with transcriptomes sequenced in triplicates.

For each sample, GSMMs were reconstructed using the CORDA algorithm (v 0.4.2) [7]. We generated five subsets of expressed genes for each sample, corresponding to the top 10%, 25%, 50%, 75%, and 90% of the highest expressed genes. Using these subsets, reactions were identified based on gene-reaction rules. A reaction with assigned isoenzymes ("OR" in the gene-reaction rule) was identified to be active, if one of the associated genes were in the selected list of genes. A reaction with assigned enzyme complexes ("AND" in the gene-reaction rule) was included, if all of the associated genes were in the selected list of genes. Reactions associated with genes in each subset were assigned a confidence score of 3, indicating the highest certainty in the CORDA framework.Additionally, reactions essential for biomass production, and exchange reactions of medium compounds and potential secretory metabolites [5] were also assigned a confidence score of 3. Conversely, reactions not associated with the selected expressed genes were assigned a confidence score of -1, reflecting low confidence due to presumed lack of expression. After reconstructing draft GSMMs from all five thresholds for each sample, the models from all timepoints were merged to create a comprehensive, context-specific GSMM representative of HEK293 cells. This "adapted" CORDA approach was compared to the default method [7], which involved assigning confidence scores ranging from -1 to 3 to reactions based on gene expression levels. Validation was conducted by integrating provided exchange rates calculated from the same samples [6] and predicting growth rates with flux balance analysis (FBA), using Cobrapy (v 0.22.1). The predicted growth rates were compared to experimental growth rates [6] and to predicted growth rates from the same process using a manually curated GSMM [10], based on Recon2 [64].

Next, we analyzed two strains (LP and HP) under two distinct experimental conditions (mock and transfected) across five timepoints (0, 4, 24, 48, and 72 HPT). Each combination of strain, condition, and timepoint was evaluated in four biological replicates, resulting in a total of 80 transcriptomic samples.Using these transcriptomic datasets [4], we applied five distinct thresholds—representing the top 10%, 25%, 50%, 75%, and 90% of expressed genes for draft-GSMM construction. This resulted in 320 individual GSMMs. To facilitate downstream comparative analyses and enhance biological interpretability, we integrated the models generated for each combination of timepoint, condition, and strain. This process resulted in 16 comprehensive, context-specific GSMMs for two strains (LP and HP), two conditions (mock and transfected), at five timepoints (0, 4, 24, 48, 72 HPT).During the subsequent optimization of production reactions, it was noted that the transport reactions of three nucleotides from the cytoplasm into the nucleus were not included in the models and were therefore added manually. To validate the reconstruction, we employed two approaches: first, we applied logisitic principal component analysis (LPCA) [8] to the differential reaction set [95] of all models and compared the results to those obtained from PCA applied to transcriptomes as shown in [4]. For all simulation related analyses, we used only models from four timepoints (4, 24, 48, 72 HPT), since a medium exchange was performed before transfection, which did not allow us to include metabolite concentrations from timepoint 0. Secondly, we integrated the calculated exchange rates and obtained simulated growth rates using FBA, where the lower bound ($lb_i$), and upper bound ($ub_i$) for each metabolite $i$ was calculated by:

$$lb_i = q_i - q_{err,i} \tag{13}$$

$$ub_i = q_i + q_{err,i} \tag{14}$$

These simulated growth rates were then compared to experimentally determined growth rates. Production reactions for AAV were incorporated into the models using Biopython (v1.79) and AAV genomes (NCBI UID: 5075992). After integrating the production reactions into the models, we analyzed the production envelope at each timepoint and performed comparative analyses between the cell lines.

### *Calculation of specific production rate from experimental titers*

First, production reactions for AAV production were added. The viral genome from AAV 8 was downloaded and the sequence for viral proteins VP1, VP2, and VP3 was extracted. The ratio of 1:1:10 (VP1:VP2:VP3) was considered in the production reaction [96]. Additionally, the replicated genome was considered as 4.8 kb long and consisting of 25% Adenine, 25% Guanine, 25% Thymine, and 25% Cytosine. Moreover, we added transport reactions for all amino acids into the nucleus, since AAV production occurs in the nucleus [43]

- Production of VP1 in the nucleus
- Production of VP2 in the nucleus
- Production of VP3 in the nucleus
- Assembly of a full capsid with ratio 1:1:10
- Production of transgene
- Assembly of full capsids
- Transport of full capsids from nucleus into the cytoplasm
- Transport of full capsid from cytoplasm outside of the cell
- Exchange reaction for full capsid

See supplemental material for a detailed explanation of the added reactions for AAV production. Additionally, we added transport reactions of amino acids to the nucleus, since AAV replication happens in the nucleus [43].

We calculated specific production rates from experimental data. Experimentally, titers could only be determined at

24, 48, and 72 HPT, as the productivity at 4 HPT was too low to measure accurately. Consequently, we calculated the percentage of biomass allocated to production at 24, 48, and 72 HPT to minimize discrepancies with the experimentally determined titers. This was done by setting the lower bound ($lb$) of the biomass reaction to the maximum growth rate, multiplied by a specific value $\alpha_t$ for each strain and time point.

$$lb_t = \alpha_t \cdot \mu_{max,t} \tag{15}$$

After that, $\alpha$ was determined by optimizing the production reaction, so that the squared difference between the simulated production rate and the experimental production rate is minimal. For optimization, we used a `differential_evolution` algorithm from scipy (v 1.7.3) with bounds between 0, and 1, since the $\alpha$ values should describe a percentage of the maximal biomass, that is used for production of AAV. We used the `best1bin` strategy with a maximum of 10 iterations, a population size of 15, and a tolerance of $10 \cdot 10^{-7}$. Using these factors ($\alpha$), and assuming that at time point 0 HPT, all resources are directed towards growth giving an $\alpha$ value of 1, we predicted an $\alpha$ value at 4 HPT using a weighted mean, following:

$$\alpha_4 = \frac{\sum_i w_i \alpha_i}{\sum_i w_i}, \tag{16}$$

with

$$w_i = \frac{1}{|t_i - 4|}, \tag{17}$$

where $w_i$ are the factors to weight the impact on the $alpha_t$ values from timepoints $t_i$, considering 0, 24, 48, and 72 HPT. This predicted $\alpha$ value was then used to calculate the specific production rate in the GSMMs at 4 HPT. After determining all production rates, we generated a realistic production envelope to compare the two cell lines by fixing simulated growth rates and maximizing production rates. Additionally, we determined ideal production rates, by decreasing the simulated growth rate percentually from 1 to 0.7 and optimized the production rate at every timepoint, and for both strains.

***Loopless flux sampling for differential reaction analysis***

For analysis of GSMMs between cell lines, we conducted loopless flux sampling. First, we fixed the calculated growth and production rates to the obtained values and used the function `loopless_solution` to generate 1000 flux samples per GSMM using the OptGPSampler with default settings in cobrapy [120]. In case a GSMM could not be sampled due to numerical issues, we relaxed the biomass constraints by 5%. Next, we used the function `calculate_loopless_fluxes` in cobrapy to correct all flux samples for loops [121]. The loopless flux samples were then used to train a random forest classifier [205]. Loopless flux samples from all conditions at one time point were first merged into one dataset, followed by converting them into Dask DataFrames [204]. Each dataset was subsequently used to build a random forest classifier by splitting it into training and test subsets (80/20 ratio) using Dask's train-test split function, ensuring random shuffling. Subsequently, reaction (feature) scaling was performed using a StandardScaler to normalize the data. Hyperparameter optimization was conducted via RandomizedSearchCV, leveraging a wide parameter grid and ROC-AUC as the scoring metric, to identify the optimal model configuration. The final model was validated using stratified k-fold cross-validation (five splits) to account for class imbalance, with performance assessed through ROC-AUC scores. The optimized model was then evaluated on the test set, and classification performance metrics were recorded.

To analyze the underlying features for GSMM separation, SHAP (SHapley Additive exPlanations) analysis was employed to interpret feature importance [188]. SHAP values were computed for the test set to assess contributions of individual features across various conditions. SHAP values for each feature were aggregated and compared across conditions within each time point using the Kruskal-Wallis test (scipy v1.14.1 [122]). Multiple testing correction was applied using the Benjamini-Hochberg procedure using the statsmodels package (v0.14.4) [123], identifying significant features with adjusted p-values $< 0.05$. We discussed those reactions, that were identified to be significantly different across all four time points or across the first three time points (4, 24, 48 HPT).

### 0.1. Differential gene expression analysis of selected genes

Genes assigned to reactions from the above analysis were extracted, and analysed using a t-test (scipy, v1.14.1) [122] with Benjamini-Hochberg correction for multiple testing (statsmodels v0.14.4) [123]. We compared expression of HP and LP samples [4] combining all timepoints and conditions. Genes with differential expression (adjusted p-value $< 0.05$) were extracted and further discussed.

## Fermentation for validation

Six validation experiments were conducted: two vessels with the original AAV production process, two vessels with added limiting amino acids, and two vessels with a HIF-1$\alpha$ inhibitor. We used 1 mg of the HIF-1$\alpha$ inhibitor PX-478 dihydrochloride (Merck KGaA, Darmstadt, Germany), dissolved in 1 mL F17 media per 250 mL of fermentation suspension. For the vessels with added limiting amino acids, aspartate, and glutamate (Merck KGaA, Darmstadt, Germany) were added at 1.5 times the initial amount (0 HPT). Control settings were used as described before (see section Bioprocess and Sampling).

## Code and data availability

All generated, and used data, as well as analyses are provided on GitHub: *github.com/LeoZ93/hek293aav*.

## Results and Discussion

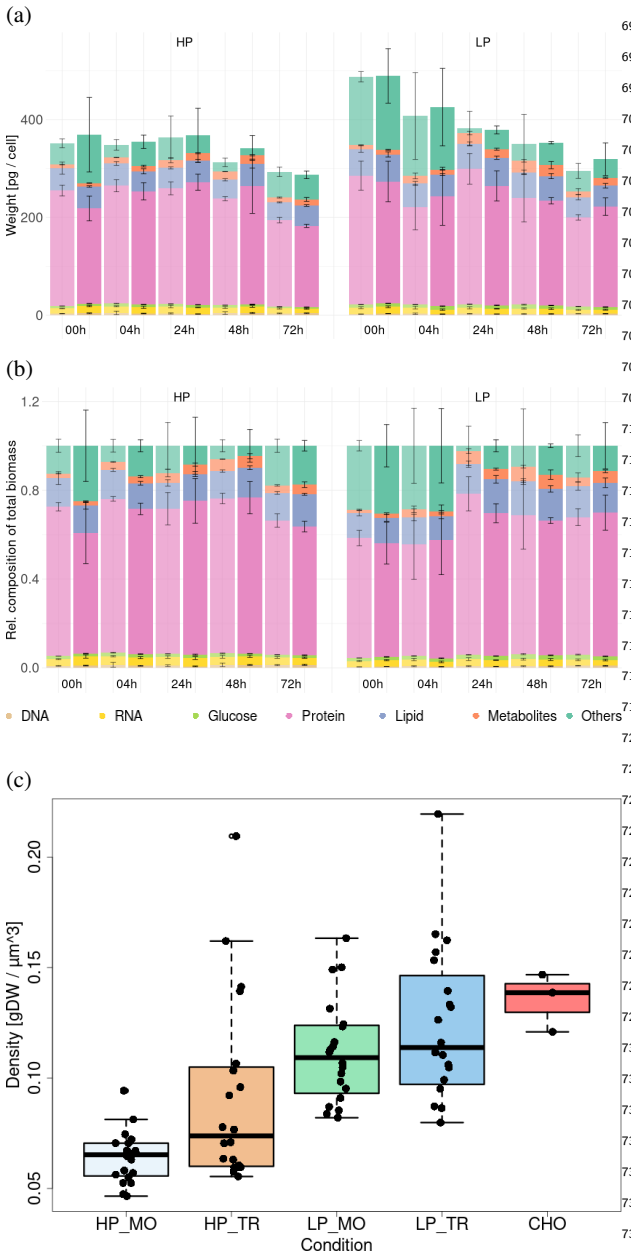


**Figure 2: Absolute (a) and relative (b) biomass composition, and volumetric mass density (c) for HP and LP strains.** The LP had a higher dry mass (adjusted $p = 6.515 \times 10^{-4}$, t-test, Bonferroni correction) and density (adjusted $p = 6.805 \times 10^{-7}$, t-test, Bonferroni correction), but a smaller volume (adjusted $p = 8.637 \times 10^{-7}$, t-test, Bonferroni correction) compared to the HP. In the first 4 h, the HP had a higher relative protein content, while other components were similar between strains. Transfected states are displayed fully colored, while mock states are transparent. We added default metabolite data from the human1 GSMM [12].

### HP cells are larger but lighter than LP cells

The dry cell mass and the composition of biomass–including total proteins, lipids, DNA, RNA, and carbohydrates–were quantified at five time points during fermentation for both the HP and LP strains. In both strains, transfection had no influence on total mass or composition. However, the dry mass decreased over time, with the LP strain showing a steeper decline.

As shown in Fig 2a, the dry mass of the HP strain ranged from 295 pg/cell to 380 pg/cell, with an average of 339.2 pg/cell, demonstrating more consistency over time compared to the LP strain, which ranged from 298 pg/cell to 447 pg/cell, averaging 388.9 pg/cell (adj. p-value $6.515 \cdot 10^{-4}$, t-test, Bonferroni correction).

For both strains, this was lower than the previously reported 514 pg/cell [6, 10]. However, this difference may be attributed to variations in cell composition, as the data in [10] were derived from multiple human cell lines based on Recon2 [11]. In comparison, the average biomass for CHO cells is 264 pg/cell [9], making the dry mass in our study generally higher.

The relative composition analysis showed that the contributions of lipids, DNA, and RNA remained almost constant over time and were independent of the strain (Fig 2b). In contrast, the HP strain had a higher protein content compared to the LP strain, particularly during the first 4 HPT. Since the amounts of lipids, DNA, and RNA were similar in both strains, the higher protein content in the HP strain suggests that other components contributing to dry mass must be higher in the LP strain. In fact, 35 % of the dry mass in the LP strain, especially early post-transfection, remains unaccounted for. A similar high proportion of unknown dry mass has also been observed in some CHO cells [9]. These authors suggested that ions, polyamines, and other small molecules likely make up much of the unaccounted biomass. Differences in the unexplained fraction between the HP and LP strains may reflect strain-specific metabolic adaptations, which could also influence experimental outcomes. For example, variations in lipid composition might alter cell membrane fluidity [118, 117], affecting cellular responses to shear stress and potentially impacting procedures such as sampling, cell counting, or lysis.

The HP strain displayed a higher volume regardless of the production state of the strain, despite having a lower dry mass compared to the LP strain (Fig 2c).

### HP stores energy as triacylglyceride; LP prefers glycogen

Next, we investigate the lipid profile for both strains that on average represented approximately 10 % to 18 % of the total biomass irrespective of the strain. We observed that phosphatidylcholine (PC) and phosphatidylethanolamine (PE) dominated across all time points (Fig 3b) accounting for about 65 % of the total lipids and reflecting their essential role in mammalian cell membranes [13, 14]. Our distribution aligns with previous analyses of HEK293 strains during

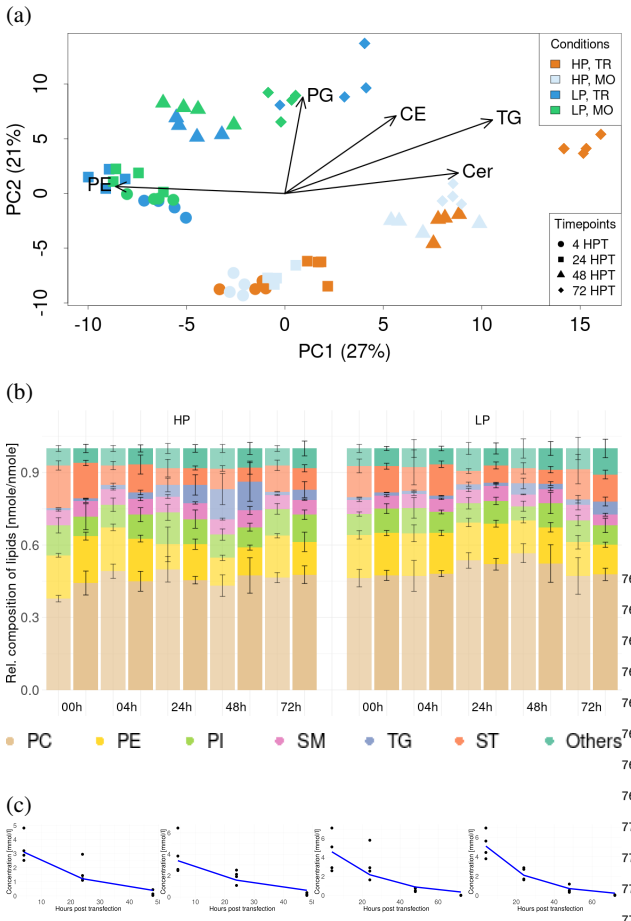


**Figure 3: PCA of lipid profiles (a), relative lipid (b), and glucose (c) composition of HEK293 strains during AAV production and mock states.** PCA resulted in a clear separation of HP and LP (a), with the five most influential lipid groups highlighted by their loadings. The predominant drivers for the separation of the strains were TG, and Cer, pinpointing in the direction of the HP cluster, while PR and PG loading vectors directed in the LP cluster. PE, PI, and PC constitute the majority of lipids in both, HP and LP strains. TG amounts changed most significantly from 1% to 12% in the HP strain until 48 HPT, indicating that the HP stored glucose as TG (b), while the LP stored glucose predominantly as glucogen (c).

exponential and stationary phases [15] and mirrors typical mammalian cell lipid profiles [16, 17].

PCA analysis clearly distinguishes between HP and LP, regardless of whether the cells were mock-transfected or transfected (Fig 3a). This separation is mainly driven by differences in triacylglycerides (TG), ceramide (Cer), phosphatidylethanolamine (PE), and phosphatidylglycerols (PG). HP cells show higher levels of TG and Cer but lower levels of PE and PG compared to LP. These lipid concentrations also change over time, with TG and Cer increasing as fermentation progresses, while PE decreases (Fig S1). In the HP strain, phospholipid levels ranged from 5 % to 12 % (Cer), 7 % to 13 % (PG), and 10 % to 19 % (PE), whereas in the LP strain, they ranged from 4 % to 10 % (Cer), 7 % to 18 % (PG), and 12 % to 18 % (PE).

**Table 2**
**Comparison of the number of reactions in HEK293F-specific GSMMs and their predicted growth rates ($\mu$) with the measured exponential growth rate ($\mu_{ref}$) as reported in [6].** 'Default' and 'adapted' CORDA [7] refer to the models developed in this study (see methods); **(author?)** [10] represents a previously established model. The relative error $\Delta\mu$ is defined as $\mu/\mu_{ref} - 1$.

| Experimental growth rate [6] $\mu_{ref} = 0.0281\,h^{-1}$ | | | |
|---|---|---|---|
| | Quek et al. [10] | CORDA | |
| | | default | adapted |
| Reactions | 357 | 5318 | 7740 |
| Genes | – | 2659 | 3035 |
| Metabolites | – | 3950 | 5015 |
| $\mu$ [$h^{-1}$] | 0.0213 | 0.0121 | 0.0301 |
| $\Delta\mu$ [%] | -24.2 | -56.9 | 7.1 |

The greatest difference was observed in TG levels, which rose from 1 % to 12 % in HP during the first 48 h, compared to only 1 % to 5 % in LP. This suggests distinct lipid storage strategies between the two strains.

An increase in lipid storage and cell volume was also observed in antibody-producing CHO cells [153]. This was linked to fatty acid synthesis exceeding the demand for membrane lipid production, potentially due to the lower surface-to-volume ratio in larger cells.

Interestingly, the LP appeared to store glucose mainly as glycogen, as indicated by its higher glucose content relative to HP (Fig 3c). This suggests a potential metabolic bottleneck in LP, possibly due to oxygen limitation (see below), since glycogen is a less efficient form of energy storage compared to fatty acids [18]. Additionally, HP cells displayed more dynamic changes in storage capabilities over time compared to LP, further highlighting differences in energy storage strategies between the strains.

Cholesterol esters (CE) increased similarly in both strains, correlating with the progression of fermentation rather than distinguishing between them.

Finally, we speculate that lipid remodeling occurs following transfection, characterized by an initial increase in short-chain fatty acids shortly after transfection, with the lipid profile gradually returning to its original distribution by the end of fermentation (Fig 3b). However, this effect is less pronounced and appears less evident in the LP strain.

## GSMMs capture transcriptomic and growth trends

To further investigate the metabolic differences between HP and LP, we reconstructed strain-specific GSMMs using CORDA [7]; see Methods for details. These models are based on the integration of the biomass composition (reported above), together with transcriptome and exometabolome measurements collected throughout the fermentation process.

To validate our reconstruction pipeline, we used transcriptomic data and exchange rates from independent studies

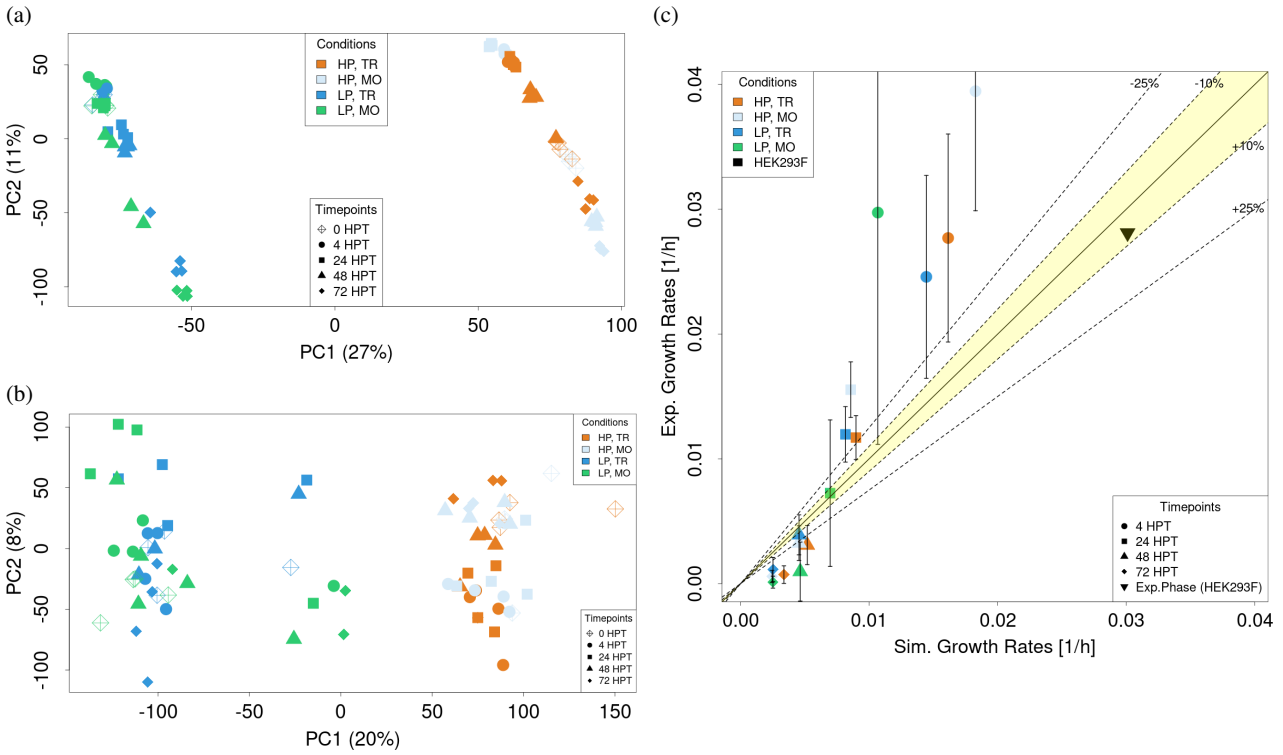


**Figure 4: PCA from transcriptomic data (a), LPCA of binary reactions matrix derived from GSMMs (b), and simulated vs. experimental growth rates (c).** When comparing LPCA (b) with a PCA plot performed on whole transcriptomes (a) we observed a similar clustering, indicating that GSMMs capture differences between strains in a similar way. Growth rates were calculated from dry mass measurements. Error bars were obtained after fitting the regression function (see methods). Simulated growth rates were obtained after merging biological replicates. Aside from the first time points 4 HPT, the simulated and experimental growth rates are comparable, indicating that the GSMMs can capture realistic growth rates during AAV production and in mock state (c).

[6, 10] to build a strain-specific GSMM for a non-producing HEK293F strain. The model predicts a specific growth rate of 0.0301 $h^{-1}$, which closely matches the experimentally measured rate of 0.0281 $h^{-1}$, resulting in a relative error of 7.1 % (Tab 2). This solid agreement indicates a good accuracy of our model and validates our reconstruction approach.

Next, we reconstructed 80 context-specific GSMMs for each of the four replicates of our two producer strains at five time points, using transcriptomic data from **(author?)** [4], and the corresponding media composition. To assess whether these models captured the differential information present in the transcriptome (Fig 4c), we collected model metrices (Tab 3), and applied LPCA [95], a method we previously demonstrated to replicate clustering patterns observed in transcriptome-based PCA using only GSMMs. As expected, LPCA of the reconstructed metabolic models revealed clustering patterns similar to those from transcriptome-based PCA (compare Fig 4c with Fig 4b). These findings suggest that even at the metabolic level strain differences were more pronounced than between conditions or time points.

Finally, to assess the quality of our context-specific GSMMs for both producer strains, we used FBA to simulate growth rates over 4 HPT to 72 HPT, using the measured exchange spectrum as input, and compared the results to a logistic fit of the experimentally measured growth curve. We observed reasonable agreement (±25 %) at intermediate growth rates, while lower rates were overestimated (Fig 4a). For higher growth rates, the large deviations were possibly due to reduced reliability in the experimental data, as indicated by substantial error bars. Nevertheless, the error bars of most experimental data points at least intersect with the ±25 % region around our predictions (Fig 4a), supporting the overall robustness of our model predictions.

**Table 3**
**Comparison of Reactions, Genes, and Metabolites in HEK293-specific GSMMs.** Numbers were extracted at four time points (4 HPT to 72 HPT) for MO and TR. Values represent minima and maxima across all time points and conditions.

| | HP | LP |
|---|---|---|
| Reactions | 9074–9188 | 8993–9155 |
| Genes | 3109–3117 | 3109–3116 |
| Metabolites | 6196–6245 | 6189–6243 |

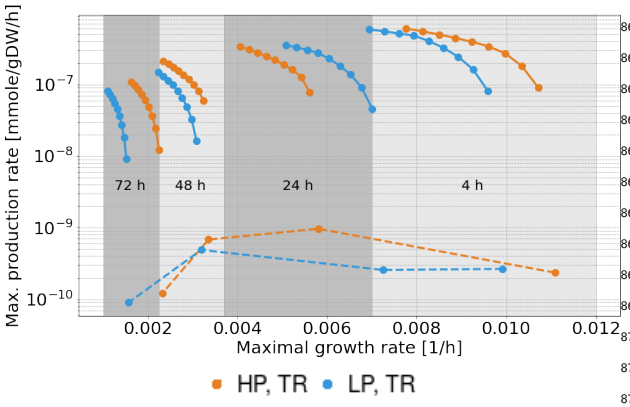


**Figure 5: Production envelope of HEK293 strains (HP: orange, LP: blue) 4 HPT, 24 HPT, 48 HPT, 72 HPT (solid lines), as well as the real production envelope (dashed line).** For every time point and cell line, the maximum growth rate was determined by FBA. These maximum growth rates were then fractionally reduced to 70% and fixed to the $lb_{bm}$, before optimizing for AAV production. The resulting solid lines displayed at the top of the plot represent the maximum production capacity. The production envelope based on experimental data (dashed line) was obtained after calculating the fraction of biomass, used for production. While the experimental production rate of the HP strain is higher compared to the LP strain, both strains are two orders of magnitude lower than their theoretical maximum production rate.

## HEK293 strains differ in limiting compounds in spent media

In addition to biomass analyses, we performed a spent media analysis as described in the methods. We identified the amino acids aspartic acid, and glutamic acid in the LP strain and serine in the HP strain as limiting, regardless of the state (see Fig S2 d, Fig S3 a, and Fig S4 d).

## HEK293 strains have similar AAV production potential, but drastically lower actual rates

After integrating AAV particle production reactions into the GSMMs of the HP and LP strains (see Supplementals), we applied FBA to compute production envelopes at four time points ranging from 4 to 72 HPT (Fig 5). These envelopes represent the maximum production capacity of each strain as a function of growth rate. We found that, for the experimentally measured uptake rates, both strains exhibit nearly identical production capabilities. This means that despite the differences in their GSMMs, the overall production potential of HP and LP is stoichiometrically almost the same. However, while the actual production rate in HP is up to six times higher than in LP, both strains produce at rates that are two orders of magnitude lower than their theoretical maximum production capacity.

## Keto-body secretion differentiates HP from LP strains

To understand why the LP strain underperformed compared to the HP strain, we performed flux sampling for both TR and MO conditions. Significant differences in individual reaction fluxes were identified using a random forest classifier with SHAP value analysis and Kruskal-Wallis tests applied to loopless flux samples. At each time point, the random forest classifier achieved an ROC-AUC score of 1.0, along with perfect precision and recall scores. This demonstrates the classifier's reliability in distinguishing conditions and its robustness in identifying reactions of interest.

We identified eleven reactions with significant flux differences over at least the first three time points (4, 24, and 48 HPT). Of these, four were associated with specific genes, and five were exchange reactions. Subsystems (derived from human1 [12]) and cellular locations were assigned where possible to provide insights into potential enrichments and functional roles (Tab 4).

Six of the eleven reactions exhibited higher flux in LP, while five showed higher flux in HP. Notably, four of the five higher-flux reactions in HP were exchange reactions. Among these, two—glutamine (MAR09063) and glutamate (MAR09071)—were experimentally measured, while pantothenate uptake (MAR09145), an essential minimal media component, was fully coupled to the biomass reaction. The remaining exchange reactions were beta-hydroxybutyrate (BHB) (MAR09134) and acetoacetate (MAR09132).

## Absence of BHB secretion links to HIF-1$\alpha$ activity activity in LP strain

Acetoacetate secretion was consistently upregulated in the LP strain across the first three time points, whereas BHB secretion was not only higher in the HP strain but entirely absent in the LP strain reconstructions. Of the 320 draft GSMMs (two strains, two conditions, four time points, four biological replicates, five gene expression thresholds), which resulted in 16 GSMMs per condition used for simulation, none of the LP models included the exchange reaction for BHB, while all HP models enabled BHB secretion. The complete absence of BHB secretion in LP models highlights a critical metabolic divergence between both strains.

BHB exhibits protective effects against oxidative stress [170] and possesses anti-inflammatory properties [166]. This corresponds with previous transcriptomic analyses of our cell lines, which revealed heightened inflammatory activity in the LP strain compared to the HP strain, irrespective of transfection status [4, 45]. Moreover, BHB has been shown to inhibit a key transcription factor HIF-1$\alpha$ [171]. BHB treatment reduces HIF-1$\alpha$ activity under hypoxic conditions, particularly at high concentrations. This effect occurs post-translationally, as BHB does not alter HIF-1$\alpha$ mRNA levels [171].

## HIF-1$\alpha$ activation significantly impairs cellular metabolism

HIF-1$\alpha$ upregulates glycolytic enzymes and glucose transporters, enhancing glucose uptake and catabolism while suppressing tricarboxylic acid (TCA) cycle activity [163, 165]. A key mechanism in the suppression of the TCA is the HIF-1$\alpha$-mediated upregulation of pyruvate dehydrogenase kinase

**Table 4**
Comparison of reactions with consistently higher (red) or lower (blue) fluxes in the GSMMs of LP vs. HP over the first three time points (4 HPT to 48 HPT). Each reaction is listed with its gene-protein-reaction (GPR) mapping, where overexpressed genes are in red and underexpressed genes in blue. Biochemical functions and metabolic subsystems are also included. Bold genes indicate regulatory targets of HIF-1$\alpha$ (see main text).

| Reaction ID | Gene-Reaction Rule | Subsystem | Cellular localization |
|---|---|---|---|
| MAR09063 | – | Exchange reactions | Extracellular |
| MAR09071 | – | Exchange reactions | Extracellular |
| MAR09132 | – | Exchange reactions | Extracellular |
| MAR09134 | – | Exchange reactions | Extracellular |
| MAR09145 | – | Exchange reactions | Extracellular |
| MAR02421 | *TECR* or *TECRL* | Omega-6 fatty acid metabolism | Cytosol |
| MAR02307 | *HSD17B12* | Fatty acid biosynthesis (unsaturated) | Cytosol |
| MAR02343 | *HSD17B12* | Fatty acid biosynthesis (unsaturated) | Cytosol |
| MAR02229 | *FASN* | Fatty acid biosynthesis (odd-chain) | Cytosol |
| MAR10365 | – | Transport reactions | Cytosol |
| MAR06902 | – | Transport reactions | Mitochondria |

(PDK), which inhibits pyruvate dehydrogenase, reducing pyruvate entry into the TCA cycle [206]. Consequently, cells may utilize glutamine or glutamate as alternative carbon sources [195]. This metabolic shift leads to the accumulation of TCA cycle intermediates, notably succinate, and fumarate, which further activate HIF-1$\alpha$ by inhibiting its regulator, prolyl hydroxylase (PHD) [173].

Indeed, transcriptomic data showed significant upregulation of HIF-1$\alpha$ in the LP strain (adj. p-value $3.239\,047 \times 10^{-17}$, t-test, Bonferroni correction), as well as HIF-1$\alpha$-regulated *PDK* (adj. p-values: *PDK1*: not signficant, *PDK2*: $4.802\,597\times10^{-16}$, *PDK3*: $1.343\,645\times10^{-12}$, *PDK4*: $9.975\,931\times 10^{-16}$, t-test, Bonferroni correction). Given the substantial post-translational regulation of HIF-1$\alpha$ by PHD, we sought additional evidence of its activation within the GSMMs.

### *Results from GSMMs support hypothesis of HIF-1α to be activated in the LP strain*

While both strains exhibited glutamine uptake, the LP strain displayed a significantly higher uptake rate across all time points (4, 24, 48 HPT). The HP strain secreted glutamate, while the LP strain consistently took it up at all time points (see spent media profiles Fig S3 a, Fig S2 g). Both amino acids are involved in glutaminolysis, a pathway upregulated in cells with activated HIF-1$\alpha$ to generate ATP [174]. Glutaminolysis converts glutamine to glutamate, which is then transformed into alpha-Ketoglutarate (AKG) to enter the TCA cycle. Gene expression analysis revealed significant upregulation of all glutaminolysis-related genes in the LP strain (adj. p-values: GLUD1: $1.232318 \cdot 10^{-05}$, GLUD2: $2.74860 \cdot 10^{-05}$, GLUL: $6.402998 \cdot 10^{-05}$, t-test, Bonferroni correction). However, it remains unclear whether increased glutaminolysis drives HIF-1$\alpha$ activation or whether HIF-1$\alpha$ activation induces elevated expression of glutaminolysis-related genes in the LP strain.

To investigate further, we examined which downstream TCA metabolites of AKG might accumulate in cells by analyzing sampled fluxes of secretion reactions in the GSMMs at the first two time points. Notably, fumarate secretion was significantly higher in the LP GSMMs at 4 and 24 HPT. Indeed, excessive fumarate accumulation inhibits succinate dehydrogenase (SDH), a key component of the mitochondrial electron transport chain (ETC). SDH inhibition blocks electron transfer from succinate to ubiquinone, disrupting electron flow through the ETC [199]. This reduces NADH oxidation at complex I, causing NADH accumulation as it cannot efficiently convert back to NAD [200]. Impaired electron flow decreases the proton gradient across the inner mitochondrial membrane, reducing ATP production via oxidative phosphorylation [201]. This may impair viral DNA replication, since the viral Rep68 is an ATP-dependent endonuclease with helicase activity [203].

Further analysis revealed that the gene encoding fumarate hydratase, which converts fumarate to malate, was significantly upregulated in the HP strain (adj. p-value $5.363976 \cdot 10^{-19}$, t-test, Bonferroni correction). This could potentially explain the accumulation of fumarate in the LP strain. Elevated fumarate levels have been identified as a potent factor for HIF-1$\alpha$ stabilization and activation, creating a positive feedback loop [177, 178, 179].

HIF-1$\alpha$ activation as well as fumarate accumulation have been associated with increased glucose uptake [175] and aerobic glycolysis [195]. On the other hand, high fumarate levels have also been linked to reduced glycolysis in macrophages [176]. While both LP and HP exhibited aerobic glycolysis, we found similar glucose uptake rates between the LP and HP strains. However, the LP strain stores higher amounts of glucose as glycogen (Fig 3c), while the HP strain generates higher levels of lactate (Fig S3 e), and stores excessive glucose as triacylglycerides (Fig 3b). The higher lactate secretion aligns with the amount of base required to maintain stable pH in the bioreactor: the HP strain necessitated greater base addition during the culture period (18.39–21.21 mL over 72 HPT) compared to the LP strain (11.14–13.26 mL over 72 HPT), despite no significant differential expression of *LDH*.

HIF-1$\alpha$ is typically activated under oxygen deprivation [172]. However, in bioreactors with constant and monitored oxygen supply, HIF-1$\alpha$ activation may result from elevated reactive oxygen species (ROS) levels or mitochondrial dysfunction [194, 193, 192]. When analyzing gene expression values of ROS detoxifying enzymes like *SOD*-family genes, we observed a higher expression in the LP strain, compared to the HP strain (adj. p-values: *SOD1*: not significant, *SOD2*: $1.974203 \cdot 10^{-20}$, *SOD3*: $1.938036 \cdot 10^{-05}$). High expression of *SOD* genes could indicate an adaption to increased oxidative stress, and therefore elevated ROS levels [207]. These high ROS concentrations could be a potential reason for HIF-1$\alpha$ activation in the LP strain. Despite these findings, the precise cause of HIF-1$\alpha$ stabilization remains unknown. However, we hypothesize that HIF-1$\alpha$ is activated in the LP strain but is either too low expressed or inactive in the HP strain, possibly due to extensive adaptation to suspension growth.

Beyond exchange reactions, we identified four reactions with significant differential flux associated with lipid metabolism as displayed in Tab 4. Two of these reactions are catalyzed by the gene *HSD17B12* (adj. p-value $3.441411 \cdot 10^{-16}$, t-test, Bonferroni correction), one by *FASN* (adj. p-value $4.304141 \cdot 10^{-12}$, t-test, Bonferroni correction), and the fourth by *TECR* (adj. p-value $2.147250 \cdot 10^{-07}$, t-test, Bonferroni correction), which showed higher expression in the HP strain. Evidence suggests that under hypoxic conditions, HIF-1$\alpha$ activation stimulates the expression and activity of *FASN*, enhancing fatty acid synthesis. This metabolic shift is a critical cellular adaptation to low-oxygen environments, supporting survival and proliferation under stress [191]. However, as two of the four reactions demonstrated higher flux despite contrasting gene expression, the significance of these findings remains inconclusive.

### *Enriched unfolded protein response (UPR), ER stress, and inflammatory pathways in the LP strain could be consequenses of HIF-1α activation*

Our findings provide a new perspective on prior results [45, 4]. While gene set enrichment analysis (GSEA) identified upregulation of ER stress and the unfolded protein response (UPR) in the LP strain without pinpointing the underlying cause, the accumulation of fumarate, potentially following HIF-1$\alpha$ activation, offers a plausible explanation. Elevated fumarate levels can contribute to oxidative stress through the production of succinicGSH, resulting in sustained oxidative stress and, consequently, ER stress [180]. This effect can also be exacerbated by high glucose exposure [181]. Additionally, fumarate can induce protein succination at cysteine residues, which directly triggers UPR, and ER stress [182]. Moreover, previous analyses [4] identified immune pathway activation, particularly upregulation of *STAT1*, in the LP strain. Interestingly, *STAT1* was already upregulated prior to transfection, as evidenced in their study, suggesting an alternative driver for its expression [4]. High fumarate levels have been linked to immune activation, including increased tumor necrosis factor expression and suppression of interleukin-10 [183]. Another study demonstrated that fumarate induces mitochondrial DNA (mtDNA) release into the cytosol, triggering inflammatory pathways. This mtDNA release activates the cGAS-STING pathway, leading to the upregulation of interferon-stimulated genes. This pathway also involves phosphorylation of *TBK1* and *IRF3*, which are key components of the type I interferon signaling cascade and strongly associated with *STAT1* activation [184]. This interplay is particularly intriguing because *STAT1* and HIF-1$\alpha$ are known to counteract each other in cancer environments [186, 185]. However, in the LP strain, both are upregulated, presenting an unusual coexistence that warrants further investigation.

## Experimental validation

Based on our hypothesis, we inhibited the transcription factor HIF-1$\alpha$ using PX-478 dihydrochloride, a known HIF-1$\alpha$ inhibitor [36, 37, 38, 39, 40]. As expected, PX-478 completely inhibited cell growth in the LP strain, confirming its strong dependence on HIF-1$\alpha$ (Fig 6a, light blue). In contrast, adding aspartic and glutamic acid (Fig 6a, dark red), identified as limiting compounds in the spent media, did not significantly impact cell growth compared to the control fermentation (Fig 6a, dark blue).

Next, we assessed productivity under these conditions by measuring viral capsid concentration (Fig 6b) and viral genome concentration (Fig 6c). Interestingly, inhibiting HIF-1$\alpha$ nearly doubled per-cell capsid production compared to the control (p-value $< 0.006727$, t-test), but viral genome production decreased. The increased viral capsids might be due to a redirected resource allocation – since cell growth was stalled in the LP strain, available amino acids could be utilized for capsid production. The decline in viral genomes might be due to reduced activity in the PPP, which is tied to HIF-1$\alpha$ and vital for nucleotide synthesis [150, 151]. Since AAV genome replication depends on human DNA replication and repair proteins (e.g., MCM, Polymerase delta) [41], inhibiting HIF-1$\alpha$, and subsequent decreased synthesis of nucleotides in the PPP, possibly lead to the reduced amount of viral genomes. Future experiments could consider supplying HEK293 cells with nucleosides during production phase to overcome the decreased PPP due to inhibition of HIF-1$\alpha$.

## Conclusion

HEK293 strains are currently the most prominent platform for the production of AAV particles in gene therapy [72, 187]. Despite numerous advances in the past decades [58, 61, 59], production titers remain insufficient to meet necessary demands [43]. In this study, we present a comprehensive multi-omics analysis, encompassing transcriptomic, lipidomic, metabolomic, and fluxomic data, to investigate and compare two different HEK293 strains in both mock and AAV production states.

When comparing the HP HEK293 strain with the LP strain, we found that the most significant differences were between the strains themselves rather than between their production states, corroborating findings by [4]. Recent

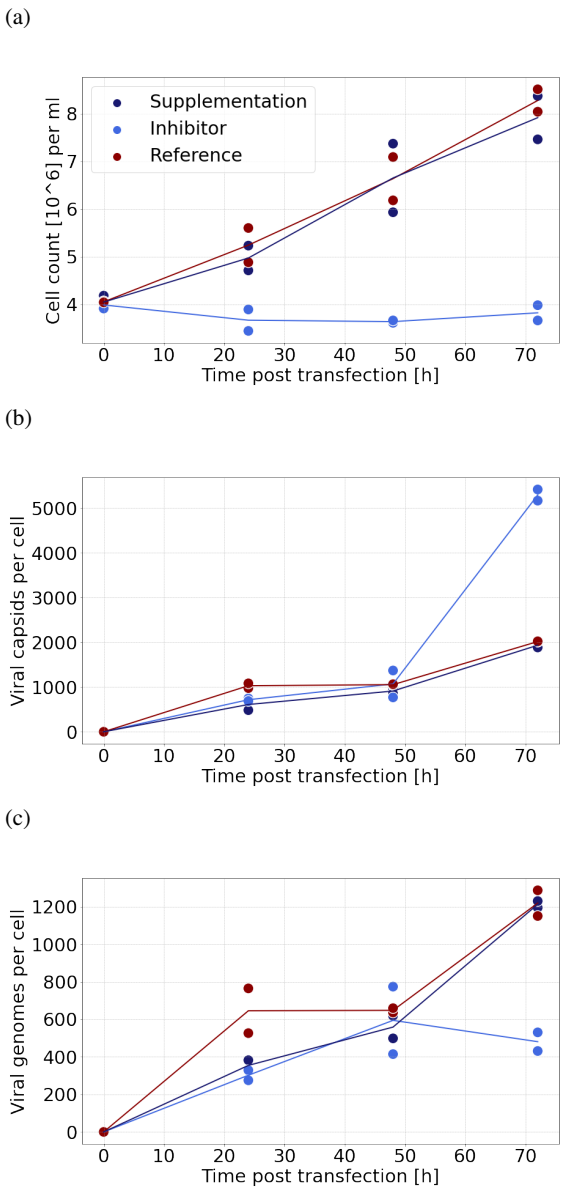


**Figure 6: Validation experiment, when inhibiting HIF-1$\alpha$ in LP HEK293 strain during AAV production.** HIF-1$\alpha$ inhibition (light blue), showed halted growth (a), higher viral capsid production (b), but reduced viral genome production (c), while the production under original conditions (dark blue), and with aspartic acid, and glutamic acid supplemented (dark red), resulted in similar growth rates (a), and production rates (b, c).

studies on HEK293 cells have revealed high glucose consumption and elevated production of lactate and succinate [46], indicative of pseudohypoxia in HEK293 cells recently adapted to suspension culture. These observations were reproduced in subsequent studies during AAV production, which also showed upregulated immune responses [4] and unfolded protein responses [51, 45], without identifying the underlying cause. By employing a multi-omics approach and integrating GSMMs, we identified that the LP strain experiences activated HIF-1$\alpha$. Lipidomic analysis showed higher concentrations of ceramides in HP strains, indicating a tumor-like state in the LP strain. Exometabolomic analysis revealed that the LP strain preferentially consumes aspartate, glutamine, and glutamate, a common behavior of solid tumor cells to maintain energy balance [78]. After identifying HIF-1$\alpha$ as a potential target, we observed that it was significantly upregulated in the LP strain compared to the HP strain. This identification would have been more difficult in whole transcriptome analysis and Gene set enrichment analysis (GSEA) as performed by [45, 4]. Inhibiting HIF-1$\alpha$ immediately halted cellular growth in the LP strain and doubled AAV capsid productivity. However, the number of viral genomes per cell decreased compared to standard production in the LP strain, likely due to insufficient nucleotide synthesis through the PPP. The PPP is central to synthesizing precursors for *de novo* nucleotide synthesis and is closely tied to HIF-1$\alpha$ activity [163, 164, 165] in hypoxic cells by directly regulating the essential PPP genes *TKT*, and *G6PD*, which catalyzes the rate limiting reaction in the PPP [161, 162]. Inhibiting HIF-1$\alpha$ likely downregulated *de novo* nucleotide synthesis, leading to fewer viral genomes. To mitigate this bottleneck, additional nucleotides could be supplemented in the media to maintain viral genome production. The overexpression of HIF-1$\alpha$ has been exploited in bioprocesses involving CHO cells to enhance RP production. Specifically, the target gene was placed under the control of a HIF-1$\alpha$-responsive sequence, thereby ensuring that the RP is expressed upon HIF-1$\alpha$ activation in CHO cells. This activation was achieved by inducing hypoxia through reduced oxygen flow [42]. These findings suggest that hypoxia and the activation of HIF-1$\alpha$ may serve as critical regulators for bioprocess control and AAV production. Given that certain HEK293 strains may exhibit elevated fumarate levels, it is crucial to consider their potential impact on recombinant products, such as RP or AAV. Elevated fumarate levels can lead to succination of cysteine residues [182], potentially affecting proper protein folding or the assembly of AAV capsids by altering post-translational modifications. Succination of viral capsids has the potential to compromise the biopotency of AAV vectors. Biopotency of AAV refers to the specific ability of an AAV vector to deliver and express the intended transgene in target cells or tissues [156]. PTMs, such as phosphorylation, and glycosylation are commonly observed on AAV capsids and are critical for maintaining capsid stability, mediating cellular interactions, and facilitating intracellular trafficking [158, 157]. Finally, the optimization of a GMP-conform bioprocess should not include any substances or additives, such as inhibitors, due to potential difficulties in downstream purification, which was not considered in this study. Therefore, it is necessary to create stable knockout strains, to inhibit the function of HIF-1$\alpha$, instead of using inhibitors, such as PX-478.

## Acknowledgements

The authors kindly thank the whole GTPD Upstream Team 1, specifically Manfred Ostermann, Katharina Portner,

Ronald Leitner, Sonja Muellner, Robert Pachlinger, Susi Heider, Carolin Kahlig, and Lucia Micutkova, as well as Peter Andorfer, Peter Eisenhut, Bernd Innthaler, Felix Fuchsberger, Fabian Knofl, Andrea Seybel, Sabine Unterthurner, and Robert Pletzenauer for their support.

## References


[1] Schoeny, H., Rampler, E., Hermann, G., Grienke, U., Rollinger, J. & Koellensperger, G. Preparative supercritical fluid chromatography for lipid class fractionation—a novel strategy in high-resolution mass spectrometry based lipidomics. *Analytical And Bioanalytical Chemistry*. **412** pp. 2365-2374 (2020)

[2] Pavlova, N., Zhu, J. & Thompson, C. The hallmarks of cancer metabolism: Still emerging. *Cell Metabolism*. **34**, 355-377 (2022)

[3] Herbert, J. & Panagiotou, S. Immune response to viruses. *Encyclopedia Of Infection And Immunity*. pp. 429 (2022)

[4] Pistek, M., Kahlig, C., Hackl, M., Unterthurner, S., Kraus, B., Grabherr, R., Grillari, J. & Hernandez Bort, J. Comprehensive mRNA-sequencing-based characterization of three HEK-293 cell lines during an rAAV production process for gene therapy applications. *Biotechnology Journal*. **18**, 2200513 (2023)

[5] Daskalaki, E., Pillon, N., Krook, A., Wheelock, C. & Checa, A. The influence of culture media upon observed cell secretome metabolite profiles: The balance between cell viability and data interpretability. *Analytica Chimica Acta*. **1037** pp. 338-350 (2018)

[6] Dietmair, S., Hodson, M., Quek, L., Timmins, N., Gray, P. & Nielsen, L. A multi-omics analysis of recombinant protein production in Hek293 cells. *PLoS ONE*. (2012)

[7] Schultz, A. & Qutub, A. Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Computational Biology*. **12**, e1004808 (2016)

[8] Landgraf, A. & Lee, Y. Dimensionality reduction for binary data through the projection of natural parameters. *Journal Of Multivariate Analysis*. **180** pp. 104668 (2020)

[9] Széliová, D., Ruckerbauer, D., Galleguillos, S., Petersen, L., Natter, K., Hanscho, M., Troyer, C., Causon, T., Schoeny, H., Christensen, H. & Others What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metabolic Engineering*. **61** pp. 288-300 (2020)

[10] Quek, L., Dietmair, S., Hanscho, M., Martinez, V., Borth, N. & Nielsen, L. Reducing Recon 2 for steady-state flux analysis of HEK cell culture. *Journal Of Biotechnology*. **184** pp. 172-178 (2014)

[11] Thiele, I., Swainston, N., Fleming, R., Hoppe, A., Sahoo, S., Aurich, M., Haraldsdottir, H., Mo, M., Rolfsson, O., Stobbe, M. & Others A community-driven global reconstruction of human metabolism. *Nature Biotechnology*. **31**, 419-425 (2013)

[12] Robinson, J., Kocabaş, P., Wang, H., Cholley, P., Cook, D., Nilsson, A., Anton, M., Ferreira, R., Domenzain, I., Billa, V. & Others An atlas of human metabolism. *Science Signaling*. **13**, eaaz1482 (2020)

[13] Cockcroft, S. Mammalian lipids: structure, synthesis and function. *Essays In Biochemistry*. **65**, 813-845 (2021)

[14] Tsuji, T., Morita, S., Nakamura, Y., Ikeda, Y., Kambe, T. & Terada, T. Alterations in cellular and organellar phospholipid compositions of HepG2 cells during cell growth. *Scientific Reports*. **11**, 2731 (2021)

[15] Zhang, Y., Baycin-Hizal, D., Kumar, A., Priola, J., Bahri, M., Heffner, K., Wang, M., Han, X., Bowen, M. & Betenbaugh, M. High-throughput lipidomic and transcriptomic analysis to compare SP2/0, CHO, and HEK-293 mammalian cell lines. *Analytical Chemistry*. **89**, 1477-1485 (2017)

[16] Vance, J. Phosphatidylserine and phosphatidylethanolamine in mammalian cells: two metabolically related aminophospholipids. *J Lipid Res*. **49**, 1377-1387 (2008)

[17] Vance, J. Phospholipid synthesis and transport in mammalian cells. *Traffic*. **16**, 1-18 (2015)

[18] Batchuluun, B., Pinkosky, S. & Steinberg, G. Lipogenesis inhibitors: therapeutic opportunities and challenges. *Nature Reviews Drug Discovery*. **21**, 283-305 (2022)

[19] Pizzagalli, M., Bensimon, A. & Superti-Furga, G. A guide to plasma membrane solute carrier proteins. *The FEBS Journal*. **288**, 2784-2835 (2021)

[20] Srikanth, K. & Orrick, J. Biochemistry, polyol or sorbitol pathways. *StatPearls [Internet]*. (2022)

[21] Nishimura, K., Katsuyama, H., Ohishi, M., Hirabayashi, A., Matsuda, K. & Nakagawa, H. Effects of sorbitol and lactate on erythropoietin production in HepG2 cells. *Biochemical And Biophysical Research Communications*. **523**, 54-59 (2020)

[22] Sinton, M., Wernig-Zorc, S., Thomson, J., Walker, P., Thakker, A., Lavery, G., Ludwig, C., Tennant, D., Hay, D., Drake, A. & Others Macrovesicular steatosis in nonalcoholic fatty liver disease is a consequence of purine nucleotide cycle driven fumarate accumulation. (2020)

[23] De Bortoli, M., Taverna, E., Maffioli, E., Casalini, P., Crisafi, F., Kumar, V., Caccia, C., Polli, D., Tedeschi, G. & Bongarzone, I. Lipid accumulation in human breast cancer cells injured by iron depletors. *Journal Of Experimental & Clinical Cancer Research*. **37** pp. 1-17 (2018)

[24] Qiao, R., Wei, M., Chen, H., Zhang, X., Zhang, J., Gao, L., Ma, H., Wang, Y. & Li, L. Critical role of 6-phosphogluconate dehydrogenase in TAp73-mediated cancer cell proliferation. *Molecular Cancer Research*. **21**, 825-835 (2023)

[25] Xu, X., Zur Hausen, A., Coy, J. & Löchelt, M. Transketolase-like protein 1 (TKTL1) is required for rapid cell growth and full viability of human tumor cells. *International Journal Of Cancer*. **124**, 1330-1337 (2009)

[26] Bentz, S., Cee, A., Endlicher, E., Wojtal, K., Naami, A., Pesch, T., Lang, S., Schubert, P., Fried, M., Weber, A. & Others Hypoxia induces the expression of transketolase-like 1 in human colorectal cancer. *Digestion*. **88**, 182-192 (2013)

[27] Wu, Y., Lee, Y., Li, W., Hsu, W., Lin, H., Chang, L., Huang, A., Jhan, J., Wu, W., Li, C. & Others High Transaldolase 1 expression predicts poor survival of patients with upper tract urothelial carcinoma. *Pathology International*. **71**, 463-470 (2021)

[28] Farsijani, N., Liu, Q., Kobayashi, H., Davidoff, O., Sha, F., Fandrey, J., Ikizler, T., O'Connor, P., Haase, V. & Others Renal epithelium regulates erythropoiesis via HIF-dependent suppression of erythropoietin. *The Journal Of Clinical Investigation*. **126**, 1425-1437 (2016)

[29] Huang, Y., Chang, K., Li, C., Lieu, A. & Lin, C. AKR1B1 represses glioma cell proliferation through p38 MAPK-mediated Bcl-2/BAX/Caspase-3 apoptotic signaling pathways. *Current Issues In Molecular Biology*. **45**, 3391-3405 (2023)

[30] Doigneaux, C., Pedley, A., Mistry, I., Papayova, M., Benkovic, S. & Tavassoli, A. Hypoxia drives the assembly of the multienzyme purinosome complex. *Journal Of Biological Chemistry*. **295**, 9551-9566 (2020)

[31] Benita, Y., Kikuchi, H., Smith, A., Zhang, M., Chung, D. & Xavier, R. An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Research*. **37**, 4587-4602 (2009)

[32] Allen, S., Seehra, R., Heath, P., Hall, B., Bates, J., Garwood, C., Matuszyk, M., Wharton, S. & Simpson, J. Transcriptomic analysis of human astrocytes in vitro reveals hypoxia-induced mitochondrial dysfunction, modulation of metabolism, and dysregulation of the immune response. *International Journal Of Molecular Sciences*. **21**, 8028 (2020)

[33] Huss, J., Levy, F. & Kelly, D. Hypoxia inhibits the peroxisome proliferator-activated receptor alpha-retinoid X receptor gene regulatory pathway in cardiac myocytes: a mechanism for O2-dependent modulation of mitochondrial fatty acid oxidation. *Journal Of Biological Chemistry*. **276**, 27605-27612 (2001)

[34] Yu, J., Perri, M., Jones, J., Pierzchalski, K., Ceaicovscaia, N., Cione, E. & Kane, M. Altered RBP1 gene expression impacts epithelial cell retinoic acid, proliferation, and microenvironment. *Cells*. **11**, 792 (2022)

[35] Hayashi, Y., Yokota, A., Harada, H. & Huang, G. Hypoxia/pseudohypoxia-mediated activation of hypoxia-inducible factor-1alpha in cancer. *Cancer Science*. **110**, 1510-1517 (2019)

[36] Welsh, S., Williams, R., Kirkpatrick, L., Paine-Murrieta, G. & Powis, G. Antitumor activity and pharmacodynamic properties of PX-478, an inhibitor of hypoxia-inducible factor-1alpha. *Molecular Cancer Therapeutics*. **3**, 233-244 (2004)

[37] Koh, M., Spivak-Kroizman, T., Venturini, S., Welsh, S., Williams, R., Kirkpatrick, D. & Powis, G. Molecular mechanisms for the activity of PX-478, an antitumor inhibitor of the hypoxia-inducible factor-1alpha. *Molecular Cancer Therapeutics*. **7**, 90-100 (2008)

[38] Zhu, Y., Zang, Y., Zhao, F., Li, Z., Zhang, J., Fang, L., Li, M., Xing, L., Xu, Z. & Yu, J. Inhibition of HIF-1alpha by PX-478 suppresses tumor growth of esophageal squamous cell cancer in vitro and in vivo. *American Journal Of Cancer Research*. **7**, 1198 (2017)

[39] Schwartz, D., Powis, G., Thitai-Kumar, A., He, Y., Bankson, J., Williams, R., Lemos, R., Oh, J., Volgin, A., Soghomonyan, S. & Others The selective hypoxia inducible factor-1 inhibitor PX-478 provides in vivo radiosensitization through tumor stromal effects. *Molecular Cancer Therapeutics*. **8**, 947-958 (2009)

[40] Xu, L., Ishikawa, H., Zhou, Y., Kobayashi, T. & Shozu, M. Antitumor effect of the selective hypoxia-inducible factor-1 inhibitors echinomycin and PX-478 on uterine fibroids. *F&S Science*. **3**, 187-196 (2022)

[41] Nash, K., Chen, W. & Muzyczka, N. Complete in vitro reconstitution of adeno-associated virus DNA replication requires the minichromosome maintenance complex proteins. *Journal Of Virology*. **82**, 1458-1464 (2008)

[42] Zeh, N., Schlossbauer, P., Raab, N., Klingler, F., Handrick, R. & Otte, K. Cell line development for continuous high cell density biomanufacturing: Exploiting hypoxia for improved productivity. *Metabolic Engineering Communications*. **13** pp. e00181 (2021)

[43] Nguyen, T., Sha, S., Hong, M., Maloney, A., Barone, P., Neufeld, C., Wolfrum, J., Springs, S., Sinskey, A. & Braatz, R. Mechanistic model for production of recombinant adeno-associated virus via triple transfection of HEK293 cells. *Molecular Therapy Methods & Clinical Development*. **21** pp. 642-655 (2021)

[44] Wang, Y., Fu, Q., Lee, Y., Sha, S. & Yoon, S. Transcriptomic features reveal molecular signatures associated with recombinant adeno-associated virus production in HEK293 cells. *Biotechnology Progress*. **39**, e3346 (2023)

[45] Kahlig, C., Moser, S., Micutkova, L., Grillari, J., Kraus, B. & Hernandez Bort, J. Enhancement of rAAV titers via inhibition of the interferon signaling cascade in transfected HEK293 suspension cultures. *Biotechnology Journal*. **19**, 2300672 (2024)

[46] Jang, M., Pete, E. & Bruheim, P. The impact of serum-free culture on HEK293 cells: From the establishment of suspension and adherent serum-free adaptation cultures to the investigation of growth and metabolic profiles. *Frontiers In Bioengineering And Biotechnology*. **10** pp. 964397 (2022)

[47] Alberghina, L. The Warburg effect explained: integration of enhanced glycolysis with heterogeneous mitochondria to promote cancer cell proliferation. *International Journal Of Molecular Sciences*. **24**, 15787 (2023)

[48] Warburg, O. On the origin of cancer cells. *Science*. **123**, 309-314 (1956)

[49] Buchsteiner, M., Quek, L., Gray, P. & Nielsen, L. Improving culture performance and antibody production in CHO cell culture processes by reducing the Warburg effect. *Biotechnology And Bioengineering*. **115**, 2315-2327 (2018)

[50] Zehetner, L., Széliová, D., Kraus, B., Graninger, M., Zanghellini, J. & Hernandez Bort, J. Optimizing VLP production in gene therapy: Opportunities and challenges for in silico modeling. *Biotechnology Journal*. **18**, 2200636 (2023)

[51] Lu, M., Lee, Z. & Hu, W. Multi-omics kinetic analysis of recombinant adeno-associated virus production by plasmid transfection of HEK293 cells. *Biotechnology Progress*. pp. e3428 (2024)

[52] Malm, M., Kuo, C., Barzadd, M., Mebrahtu, A., Wistbacka, N., Razavi, R., Volk, A., Lundqvist, M., Kotol, D., Tegel, H. & Others Harnessing secretory pathway differences between HEK293 and CHO to rescue production of difficult to express proteins. *Metabolic Engineering*. **72** pp. 171-187 (2022)

[53] Malm, M., Saghaleyni, R., Lundqvist, M., Giudici, M., Chotteau, V., Field, R., Varley, P., Hatton, D., Grassi, L., Svensson, T. & Others Evolution from adherent to suspension: systems biology of HEK293 cell line development. *Scientific Reports*. **10**, 18996 (2020)

[54] Lin, Y., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R. & Others Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications*. **5**, 4767 (2014)

[55] Catalan-Tatjer, D., Tzimou, K., Nielsen, L. & Lavado-Garcia, J. Unravelling the essential elements for recombinant adeno-associated virus (rAAV) production in animal cell-based platforms. *Biotechnology Advances*. pp. 108370 (2024)

[56] Mendes, J., Fernandes, B., Pineda, E., Kudugunti, S., Bransby, M., Gantier, R., Peixoto, C., Alves, P., Roldão, A. & Silva, R. AAV process intensification by perfusion bioreaction and integrated clarification. *Frontiers In Bioengineering And Biotechnology*. **10** pp. 1020174 (2022)

[57] Moço, P., Xu, X., Silva, C. & Kamen, A. Production of adeno-associated viral vector serotype 6 by triple transfection of suspension HEK293 cells at higher cell densities. *Biotechnology Journal*. **18**, 2300051 (2023)

[58] Clement, N. & Grieger, J. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Molecular Therapy Methods & Clinical Development*. **3** (2016)

[59] Blessing, D., Vachey, G., Pythoud, C., Rey, M., Padrun, V., Wurm, F., Schneider, B. & Déglon, N. Scalable production of AAV vectors in orbitally shaken HEK293 cells. *Molecular Therapy-Methods & Clinical Development*. **13** pp. 14-26 (2019)

[60] Shiina, S., Nagao, N., Hasegawa, J., Sato, T., Mori, C., Ohtaki, K., Kubodera, K., Yamashita, Y., Tanabe, K., Kawano, Y. & Others Effect of alcohol on productivity and quality of adeno-associated virus 2 in HEK293 cells. *Journal Of Bioscience And Bioengineering*. **134**, 338-347 (2022)

[61] Strasser, L., Boi, S., Guapo, F., Donohue, N., Barron, N., Rainbow-Fletcher, A. & Bones, J. Proteomic landscape of adeno-associated virus (AAV)-producing HEK293 cells. *International Journal Of Molecular Sciences*. **22**, 11499 (2021)

[62] Abaandou, L., Quan, D. & Shiloach, J. Affecting HEK293 cell growth and production performance by modifying the expression of specific genes. *Cells*. **10**, 1667 (2021)

[63] Bairoch, A. The cellosaurus, a cell-line knowledge resource. *Journal Of Biomolecular Techniques: JBT*. **29**, 25 (2018)

[64] Swainston, N., Smallbone, K., Hefzi, H., Dobson, P., Brewer, J., Hanscho, M., Zielinski, D., Ang, K., Gardiner, N., Gutierrez, J. & Others Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*. **12** pp. 1-7 (2016)

[65] Brunk, E., Sahoo, S., Zielinski, D., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G., Aurich, M. & Others Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nature Biotechnology*. **36**, 272-281 (2018)

[66] Vlassis, N., Pacheco, M. & Sauter, T. Fast reconstruction of compact context-specific metabolic network models. *PLoS Computational Biology*. **10**, e1003424 (2014)

[67] Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I. & Nielsen, J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Computational Biology*. **8**, e1002518 (2012)

[68] Becker, S. & Palsson, B. Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology*. **4**, e1000082 (2008)

[69] Jerby, L., Shlomi, T. & Ruppin, E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular Systems Biology*. **6**, 401 (2010)

[70] Edwards, J. & Palsson, B. Systems properties of the Haemophilus influenzaeRd metabolic genotype. *Journal Of Biological Chemistry*. **274**, 17410-17416 (1999)

[71] Gu, C., Kim, G., Kim, W., Kim, H. & Lee, S. Current status and applications of genome-scale metabolic models. *Genome Biology*. **20** pp. 1-18 (2019)

[72] Bulcha, J., Wang, Y., Ma, H., Tai, P. & Gao, G. Viral vector platforms within the gene therapy landscape. *Signal Transduction And Targeted Therapy*. **6**, 53 (2021)

[73] Graham, F., Smiley, J., Russell, W. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *Journal Of General Virology*. **36**, 59-72 (1977)

[74] Thomas, P. & Smart, T. HEK293 cell line: a vehicle for the expression of recombinant proteins. *Journal Of Pharmacological And Toxicological Methods*. **51**, 187-200 (2005)

[75] Robitaille, A., Huguet, R., Bomgarden, R., Canterbury, J. & Lopez-Ferrer, D. Real-Time Search enables a new gold standard for TMT quantitation accuracy on the Orbitrap Eclipse Tribrid mass spectrometer. *Braz. J. Anal. Chem.* **7**, 91-99 (2020)

[76] Bartoszewska, S., Collawn, J. & Bartoszewski, R. The role of the hypoxia-related unfolded protein response (UPR) in the tumor microenvironment. *Cancers*. **14**, 4870 (2022)

[77] Dinarello, A., Betto, R., Diamante, L., Tesoriere, A., Ghirardo, R., Cioccarelli, C., Meneghetti, G., Peron, M., Laquatra, C., Tiso, N. & Others STAT3 and HIF1alpha cooperatively mediate the transcriptional and physiological responses to hypoxia. *Cell Death Discovery*. **9**, 226 (2023)

[78] Garcia-Bermudez, J., Baudrier, L., La, K., Zhu, X., Fidelin, J., Sviderskiy, V., Papagiannakopoulos, T., Molina, H., Snuderl, M., Lewis, C. & Others Aspartate is a limiting metabolite for cancer cell proliferation under hypoxia and in tumours. *Nature Cell Biology*. **20**, 775-781 (2018)

[79] Chung, C., Murphy, C., Wingate, V., Pavlicek, J., Nakashima, R., Wei, W., McCarty, D., Rabinowitz, J. & Barton, E. Production of rAAV by plasmid transfection induces antiviral and inflammatory responses in suspension HEK293 cells. *Molecular Therapy Methods & Clinical Development*. **28** pp. 272-283 (2023)

[80] Castellano, B., Tang, D., Marsters, S., Lam, C., Liu, P., Rose, C., Sandoval, W., Ashkenazi, A., Snedecor, B. & Misaghi, S. Activation of the PERK branch of the unfolded protein response during production reduces specific productivity in CHO cells via downregulation of PDGFRa and IRE1a signaling. *Biotechnology Progress*. **39**, e3354 (2023)

[81] Prashad, K. & Mehra, S. Dynamics of unfolded protein response in recombinant CHO cells. *Cytotechnology*. **67** pp. 237-254 (2015)

[82] Madden, E., Logue, S., Healy, S., Manie, S. & Samali, A. The role of the unfolded protein response in cancer progression: From oncogenesis to chemoresistance. *Biology Of The Cell*. **111**, 1-17 (2019)

[83] Henry, O., Perrier, M. & Kamen, A. Metabolic flux analysis of HEK-293 cells in perfusion cultures for the production of adenoviral vectors. *Metabolic Engineering*. **7**, 467-476 (2005)

[84] Huang, Z., Xu, J., Yongky, A., Morris, C., Polanco, A., Reily, M., Borys, M., Li, Z. & Yoon, S. CHO cell productivity improvement by genome-scale modeling and pathway analysis: Application to feed supplements. *Biochemical Engineering Journal*. **160** pp. 107638 (2020)

[85] Cotner, M., Zhan, J. & Zhang, Z. A computational metabolic model for engineered production of resveratrol in Escherichia coli. *ACS Synthetic Biology*. **10**, 1992-2001 (2021)

[86] Vailionis, J., Zhao, W., Zhang, K., Rodionov, D., Lipscomb, G., Tanwee, T., O'Quinn, H., Bing, R., Kelly, R., Adams, M. & Others Optimizing Strategies for Bio-Based Ethanol Production Using Genome-Scale Metabolic Modeling of the Hyperthermophilic Archaeon, Pyrococcus furiosus. *Applied And Environmental Microbiology*. **89**, e00563-23 (2023)

[87] Xu, N., Ye, C. & Liu, L. Genome-scale biological models for industrial microbial systems. *Applied Microbiology And Biotechnology*. **102** pp. 3439-3451 (2018)

[88] Aminian-Dehkordi, J., Mousavi, S., Jafari, A., Mijakovic, I. & Marashi, S. Manually curated genome-scale reconstruction of the metabolic network of Bacillus megaterium DSM319. *Scientific Reports*. **9**, 18762 (2019)

[89] Chen, Y., McConnell, B., Gayatri Dhara, V., Mukesh Naik, H., Li, C., Antoniewicz, M. & Betenbaugh, M. An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells. *NPJ Systems Biology And Applications*. **5**, 25 (2019)

[90] Pérez-Fernández, B., Cossio-Diaz, J., Boggiano, T., León, K. & Mulet, R. In-silico media optimization for continuous cultures using genome scale metabolic networks: The case of CHO-K1. *Biotechnology And Bioengineering*. **118**, 1884-1897 (2021)

[91] Clausen, A. & Sokol, S. Deriv: R-based Symbolic Differentiation. (2020), https://CRAN.R-project.org/package=Deriv, Deriv package version 4.1

[92] El Ansari, R., Craze, M., Miligy, I., Diez-Rodriguez, M., Nolan, C., Ellis, I., Rakha, E. & Green, A. The amino acid transporter SLC7A5 confers a poor prognosis in the highly proliferative breast cancer subtypes and is a key therapeutic target in luminal B tumours. *Breast Cancer Research*. **20** pp. 1-17 (2018)

[93] Najumudeen, A., Ceteci, F., Fey, S., Hamm, G., Steven, R., Hall, H., Nikula, C., Dexter, A., Murta, T., Race, A. & Others The amino acid transporter SLC7A5 is required for efficient growth of KRAS-mutant colorectal cancer. *Nature Genetics*. **53**, 16-26 (2021)

[94] Meskers, C., Franczak, M., Smolenski, R., Giovannetti, E. & Peters, G. Are we still on the right path (way)?: the altered expression of the pentose phosphate pathway in solid tumors and the potential of its inhibition in combination therapy. *Expert Opinion On Drug Metabolism & Toxicology*. **18**, 61-83 (2022)

[95] Zehetner, L., Széliová, D., Kraus, B., Hernandez Bort, J. & Zanghellini, J. Logistic PCA explains differences between genome-scale metabolic models in terms of metabolic pathways. *PLOS Computational Biology*. **20**, e1012236 (2024)

[96] Samulski, R. & Muzyczka, N. AAV-mediated gene therapy for research and therapeutic purposes. *Annual Review Of Virology*. **1**, 427-451 (2014)

[97] Martinez-Monge, I., Albiol, J., Lecina, M., Liste-Calleja, L., Miret, J., Sola, C. & Cairo, J. Metabolic flux balance analysis during lactate and glucose concomitant consumption in HEK293 cell cultures. *Biotechnology And Bioengineering*. **116**, 388-404 (2019)

[98] Chen, K., Wu, X., Yu, D., Ou, H., Li, Y., Zhou, Y. & Li, L. Process optimization for the rapid production of adenoviral vectors for clinical trials in a disposable bioreactor system. *Applied Microbiology And Biotechnology*. **102**, 6469-6477 (2018)

[99] Joe, C., Jiang, J., Linke, T., Li, Y., Fedosyuk, S., Gupta, G., Berg, A., Segireddy, R., Mainwaring, D., Joshi, A. & Others Manufacturing a chimpanzee adenovirus-vectored SARS-CoV-2 vaccine to meet global needs. *Biotechnology And Bioengineering*. **119**, 48 (2022)

[100] Ezeonwumelu, I., Garcia-Vidal, E. & Ballana, E. JAK-STAT pathway: A novel target to tackle viral infections. *Viruses*. **13**, 2379 (2021)

[101] O'Shea, J., Schwartz, D., Villarino, A., Gadina, M., McInnes, I. & Laurence, A. The JAK-STAT pathway: impact on human disease and therapeutic intervention. *Annual Review Of Medicine*. **66**, 311-328 (2015)

[102] Hu, X., Li, J., Fu, M., Zhao, X. & Wang, W. The JAK/STAT signaling pathway: from bench to clinic. *Signal Transduction And Targeted Therapy*. **6**, 402 (2021)

[103] Baumeister, R., Murphy, C. & Heimbucher, T. Metabolic adaptation to hypoxia: do worms and cancer cells share common metabolic responses to hypoxic stress?. *Cell Death & Differentiation*. **28**, 1434-1436 (2021)

[104] Nath, B., Levin, I., Csak, T., Petrasek, J., Mueller, C., Kodys, K., Catalano, D., Mandrekar, P. & Szabo, G. Hepatocyte-specific hypoxia-inducible factor-1alpha is a determinant of lipid accumulation and liver injury in alcohol-induced steatosis in mice. *Hepatology*. **53**, 1526-1537 (2011)

[105] Mesarwi, O., Shin, M., Bevans-Fonti, S., Schlesinger, C., Shaw, J. & Polotsky, V. Hepatocyte hypoxia inducible factor-1 mediates the development of liver fibrosis in a mouse model of nonalcoholic fatty liver disease. *PLoS One*. **11**, e0168572 (2016)

[106] Széliová, D., Iurashev, D., Ruckerbauer, D., Koellensperger, G., Borth, N., Melcher, M. & Zanghellini, J. Error propagation in constraint-based

modeling of Chinese hamster ovary cells. *Biotechnology Journal*. **16**, 2000320 (2021)

[107] Choi, Y., Choi, D., Lee, Y., Koduru, L., Lewis, N., Lakshmanan, M. & Lee, D. Mitigating biomass composition uncertainties in flux balance analysis using ensemble representations. *Computational And Structural Biotechnology Journal*. **21** pp. 3736-3745 (2023)

[108] Bernstein, D., Sulheim, S., Almaas, E. & Segrè, D. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biology*. **22** pp. 1-22 (2021)

[109] Chen, Y., Gustafsson, J., Tafur Rangel, A., Anton, M., Domenzain, I., Kittikunapong, C., Li, F., Yuan, L., Nielsen, J. & Kerkhoven, E. Reconstruction, simulation and analysis of enzyme-constrained metabolic models using GECKO Toolbox 3.0. *Nature Protocols*. **19**, 629-667 (2024)

[110] Salvy, P., Fengos, G., Ataman, M., Pathier, T., Soh, K. & Hatzimanikatis, V. pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis. *Bioinformatics*. **35**, 167-169 (2019)

[111] Thaker, S., Ch'ng, J. & Christofk, H. Viral hijacking of cellular metabolism. *BMC Biology*. **17** pp. 1-15 (2019)

[112] Thai, M., Graham, N., Braas, D., Nehil, M., Komisopoulou, E., Kurdistani, S., McCormick, F., Graeber, T. & Christofk, H. Adenovirus E4ORF1-induced MYC activation promotes host cell anabolic glucose metabolism and virus replication. *Cell Metabolism*. **19**, 694-701 (2014)

[113] Feudner, E., De Alwis, M., Thrasher, A., Ali, R. & Fauser, S. Optimization of recombinant adeno-associated virus production using an herpes simplex virus amplicon system. *Journal Of Virological Methods*. **96**, 97-105 (2001)

[114] Alazard-Dany, N., Nicolas, A., Ploquin, A., Strasser, R., Greco, A., Epstein, A., Fraefel, C. & Salvetti, A. Definition of herpes simplex virus type 1 helper activities for adeno-associated virus early replication events. *PLoS Pathogens*. **5**, e1000340 (2009)

[115] Meier, A., Fraefel, C. & Seyffert, M. The interplay between adeno-associated virus and its helper viruses. *Viruses*. **12**, 662 (2020)

[116] Nie, Y., Pan, H., Li, Q., Na, H., Figueroa, B. & Vincent, K. Characterization of the function of Adenovirus L4 gene products and their impact on AAV vector production. *Molecular Therapy Methods & Clinical Development*. **32** (2024)

[117] Meer, G. & Kroon, A. Lipid map of the mammalian cell. *Journal Of Cell Science*. **124**, 5-8 (2011)

[118] Van Meer, G., Voelker, D. & Feigenson, G. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology*. **9**, 112-124 (2008)

[119] Sheikh, K., Förster, J. & Nielsen, L. Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus. *Biotechnology Progress*. **21**, 112-121 (2005)

[120] Megchelenbrink, W., Huynen, M. & Marchiori, E. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PloS One*. **9**, e86587 (2014)

[121] Desouki, A., Jarre, F., Gelius-Dietrich, G. & Lercher, M. CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions. *Bioinformatics*. **31**, 2159-2165 (2015)

[122] Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van der Walt, S., Brett, M., Wilson, J., Millman, K., Mayorov, N., Nelson, A., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C., Archibald, A., Ribeiro, A., Pedregosa, F., Van Mulbregt, P. & SciPy 1.0 Contributors SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*. **17** pp. 261-272 (2020)

[123] Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python. *9th Python In Science Conference*. (2010)

[124] Ramachandran, A., Betts, G., Bhana, S., Helme, G., Blick, C., Moller, Levet, C., Saunders, E., Valentine, H., Pepper, S., Miller, C. & Others An in vivo hypoxia metagene identifies the novel hypoxia inducible factor target gene SLCO1B3. *European Journal Of Cancer*. **49**, 1741-1751 (2013)

[125] Thompson, B., Sanchez-Covarrubias, L., Slosky, L., Zhang, Y., Laracuente, M. & Ronaldson, P. Hypoxia/reoxygenation stress signals an increase in organic anion transporting polypeptide 1a4 (Oatp1a4) at the blood–brain barrier: relevance to CNS drug delivery. *Journal Of Cerebral Blood Flow & Metabolism*. **34**, 699-707 (2014)

[126] Liang, Z., He, H., Zhang, B., Kai, Z. & Zong, L. Hypoxia expedites the progression of papillary thyroid carcinoma by promoting the CPT1A-mediated fatty acid oxidative pathway. *Drug Development Research*. **85**, e22168 (2024)

[127] Rios-Colon, L., Kumar, P., Kim, S., Sharma, M., Su, Y., Kumar, A., Singh, S., Stocks, N., Liu, L., Joshi, M. & Others Carnitine Palmitoyltransferase 1 Regulates Prostate Cancer Growth under Hypoxia., 2021, 13, 6302. *DOI: Https://doi. Org/10.3390/cancers13246302*.

[128] Zhou, Y., Chen, Y., Zhao, P., Xian, T., Gao, Y., Fan, S., Fang, J., Huang, M. & Bi, H. The YY1-CPT1C signaling axis modulates the proliferation and metabolism of pancreatic tumor cells under hypoxia. *Biochemical Pharmacology*. **227** pp. 116422 (2024)

[129] Niu, N., Miao, H. & Ren, H. Transcriptome Analysis of Myocardial Ischemic-Hypoxic Injury in Rats and Hypoxic H9C2 Cells. *ESC Heart Failure*. (2024)

[130] Yuan, P., Mu, J., Wang, Z., Ma, S., Da, X., Song, J., Zhang, H., Yang, L., Li, J. & Yang, J. Down-regulation of SLC25A20 promotes hepatocellular carcinoma growth and metastasis through suppression of fatty-acid oxidation. *Cell Death & Disease*. **12**, 361 (2021)

[131] Zhang, B., Chen, Y., Bao, L. & Luo, W. GPT2 is induced by hypoxia-inducible factor (HIF)-2 and promotes glioblastoma growth. *Cells*. **11**, 2597 (2022)

[132] Nayak, D., Weadick, B., Persaud, A., Raj, R., Shakya, R., Li, J., Campbell, M. & Govindarajan, R. EMT alterations in the solute carrier landscape uncover SLC22A10/A15 imposed vulnerabilities in pancreatic cancer. *Iscience*. **25** (2022)

[133] Yan, W., Xie, C., Sun, S., Zheng, Q., Wang, J., Wang, Z., Man, C., Wang, H., Yang, Y., Wang, T. & Others SUCLG1 restricts POLRMT succinylation to enhance mitochondrial biogenesis and leukemia progression. *The EMBO Journal*. pp. 1-31 (2024)

[134] Hu, Q., Xu, J., Wang, L., Yuan, Y., Luo, R., Gan, M., Wang, K., Zhao, T., Wang, Y., Han, T. & Others SUCLG2 regulates mitochondrial dysfunction through Succinylation in Lung Adenocarcinoma. *Advanced Science*. **10**, 2303535 (2023)

[135] Boese, A., Kang, J., Hwang, J., Kim, J., Eun, K., Malin, C., Magliocca, K., Pan, C., Jin, L. & Kang, S. Succinyl-CoA ligase ADP-forming subunit beta promotes stress granule assembly to regulate redox and drive cancer metastasis. *Proceedings Of The National Academy Of Sciences*. **120**, e2217332120 (2023)

[136] Okui, T., Iwashita, M., Rogers, M., Halu, A., Atkins, S., Kuraoka, S., Abdelhamid, I., Higashi, H., Ramsaroop, A., Aikawa, M. & Others CROT (carnitine O-octanoyltransferase) is a novel contributing factor in vascular calcification via promoting fatty acid metabolism and mitochondrial dysfunction. *Arteriosclerosis, Thrombosis, And Vascular Biology*. **41**, 755-768 (2021)

[137] Wang, D., Wang, Q., Yan, G., Qiao, Y., Zhu, B., Liu, B. & Tang, C. Hypoxia induces lactate secretion and glycolytic efflux by downregulating mitochondrial pyruvate carrier levels in human umbilical vein endothelial cells. *Molecular Medicine Reports*. **18**, 1710-1717 (2018)

[138] Rios-Colon, L., Kumar, P., Kim, S., Sharma, M., Su, Y., Kumar, A., Singh, S., Stocks, N., Liu, L., Joshi, M. & Others Carnitine palmitoyltransferase 1 regulates prostate cancer growth under hypoxia. *Cancers*. **13**, 6302 (2021)

[139] Kerk, S., Lin, L., Myers, A., Sutton, D., Andren, A., Sajjakulnukit, P., Zhang, L., Zhang, Y., Jiménez, J., Nelson, B. & Others Metabolic requirement for GOT2 in pancreatic cancer depends on environmental context. *Elife*. **11** pp. e73245 (2022)

[140] Liang, Q., Liu, S., Yin, F., Liu, M., Wang, L., Guo, E., Lei, L., Wu, L., Yang, Y., Zhang, D. & Others Low expression of GOT2 promotes tumor progress and predicts poor prognosis in hepatocellular carcinoma. *Biomarkers In Medicine*. **17**, 755-765 (2023)

[141] Martano, M., Power, K., Cuccaro, B., Razzuoli, E., Maiolino, P. & Restucci, B. Overexpression of the key metabolic protein Carnitine

Palmitoyl Transferase 1A (CPT1A) in equine sarcoid. *Journal Of Equine Veterinary Science*. **143** pp. 105205 (2024)

[142] Ma, W., Li, C., Tao, R., Wang, X. & Yan, L. Reductive stress-induced mitochondrial dysfunction and cardiomyopathy. *Oxidative Medicine And Cellular Longevity*. **2020**, 5136957 (2020)

[143] Fink, B., Bai, F., Yu, L., Sheldon, R., Sharma, A., Taylor, E. & Sivitz, W. Oxaloacetic acid mediates ADP-dependent inhibition of mitochondrial complex II–driven respiration. *Journal Of Biological Chemistry*. **293**, 19932-19941 (2018)

[144] Tannahill, G., Curtis, A., Adamik, J., Palsson-McDermott, E., McGettrick, A., Goel, G., Frezza, C., Bernard, N., Kelly, B., Foley, N. & Others Succinate is a danger signal that induces IL-1b via HIF-1a. *Nature*. **496**, 238-242 (2013)

[145] Selak, M., Armour, S., MacKenzie, E., Boulahbel, H., Watson, D., Mansfield, K., Pan, Y., Simon, M., Thompson, C. & Gottlieb, E. Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-alpha prolyl hydroxylase. *Cancer Cell*. **7**, 77-85 (2005)

[146] Yu, T., Robotham, J. & Yoon, Y. Increased production of reactive oxygen species in hyperglycemic conditions requires dynamic change of mitochondrial morphology. *Proceedings Of The National Academy Of Sciences*. **103**, 2653-2658 (2006)

[147] Chipurupalli, S., Kannan, E., Tergaonkar, V., D'Andrea, R. & Robinson, N. Hypoxia induced ER stress response as an adaptive mechanism in cancer. *International Journal Of Molecular Sciences*. **20**, 749 (2019)

[148] Bayley, J. & Devilee, P. The Warburg effect in 2012. *Current Opinion In Oncology*. **24**, 62-67 (2012)

[149] Zhang, M., Cheng, Y., Zhai, Y., Cui, Y., Zhang, W., Sun, H., Xin, W., Zhou, L., Gao, X., Li, S. & Others Attenuated succinate accumulation relieves neuronal injury induced by hypoxia in neonatal mice. *Cell Death Discovery*. **8**, 138 (2022)

[150] Jiang, P., Du, W. & Wu, M. Regulation of the pentose phosphate pathway in cancer. *Protein & Cell*. **5**, 592-602 (2014)

[151] Jin, L. & Zhou, Y. Crucial role of the pentose phosphate pathway in malignant tumors. *Oncology Letters*. **17**, 4213-4221 (2019)

[152] Borst, P. The malate–aspartate shuttle (Borst cycle): How it started and developed into a major metabolic pathway. *IUBMB Life*. **72**, 2241-2259 (2020)

[153] Pan, X., Dalm, C., Wijffels, R. & Martens, D. Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology And Biotechnology*. **101** pp. 8101-8113 (2017)

[154] Singh, R. & Manna, P. Reactive oxygen species in cancer progression and its role in therapeutics. *Exploration Of Medicine*. **3**, 43-57 (2022)

[155] Alnajjar, K. & Sweasy, J. A new perspective on oxidation of DNA repair proteins and cancer. *DNA Repair*. **76** pp. 60-69 (2019)

[156] Salmikangas, P., Carlsson, B., Klumb, C., Reimer, T. & Thirstrup, S. Potency testing of cell and gene therapy products. *Frontiers In Medicine*. **10** pp. 1190016 (2023)

[157] Mary, B., Maurya, S., Arumugam, S., Kumar, V. & Jayandharan, G. Post-translational modifications in capsid proteins of recombinant adeno-associated virus (AAV) 1-rh10 serotypes. *The FEBS Journal*. **286**, 4964-4981 (2019)

[158] Mulcrone, P., Lam, A., Frabutt, D., Zhang, J., Chrzanowski, M., Herzog, R. & Xiao, W. Chemical modification of AAV9 capsid with N-ethyl maleimide alters vector tissue tropism. *Scientific Reports*. **13**, 8436 (2023)

[159] Giorgio, M., Dellino, G., Gambino, V., Pelicci, P. & Others On the epigenetic role of guanosine oxidation. *Redox Biology*. **29** pp. 101398 (2020)

[160] Delbrel, E., Soumare, A., Naguez, A., Label, R., Bernard, O., Bruhat, A., Fafournoux, P., Tremblais, G., Marchant, D., Gille, T. & Others HIF-1alpha triggers ER stress and CHOP-mediated apoptosis in alveolar epithelial cells, a key event in pulmonary fibrosis. *Scientific Reports*. **8**, 17939 (2018)

[161] Trejo-Solis, C., Castillo-Rodriguez, R., Serrano-Garcia, N., Silva-Adaya, D., Vargas-Cruz, S., Chavez-Cortez, E., Gallardo-Perez, J., Zavala-Vega, S., Cruz-Salgado, A. & Magana-Maldonado, R. Metabolic Roles of HIF1, c-Myc, and p53 in Glioma Cells. *Metabolites*. **14**, 249 (2024)

[162] Tokuda, K., Baron, B., Yamashiro, C., Kuramitsu, Y., Kitagawa, T., Kobayashi, M., Sonoda, K. & Kimura, K. Up-regulation of the pentose phosphate pathway and HIF-1alpha expression during neural progenitor cell induction following glutamate treatment in rat ex vivo retina. *Cell Biology International*. **44**, 137-144 (2020)

[163] Wang, T., Liu, H., Lian, G., Zhang, S., Wang, X. & Jiang, C. HIF1alpha-induced glycolysis metabolism is essential to the activation of inflammatory macrophages. *Mediators Of Inflammation*. **2017**, 9029327 (2017)

[164] Nagao, A., Kobayashi, M., Koyasu, S., Chow, C. & Harada, H. HIF-1-dependent reprogramming of glucose metabolic pathway of cancer cells and its therapeutic significance. *International Journal Of Molecular Sciences*. **20**, 238 (2019)

[165] Corcoran, S., O'Neill, L. & Others HIF1alpha and metabolic reprogramming in inflammation. *The Journal Of Clinical Investigation*. **126**, 3699-3707 (2016)

[166] Qi, J., Gan, L., Fang, J., Zhang, J., Yu, X., Guo, H., Cai, D., Cui, H., Gou, L., Deng, J. & Others Beta-hydroxybutyrate: a dual function molecular and immunological barrier function regulator. *Frontiers In Immunology*. **13** pp. 805881 (2022)

[167] Cuenoud, B., Hartweg, M., Godin, J., Croteau, E., Maltais, M., Castellano, C., Carpentier, A. & Cunnane, S. Metabolism of exogenous D-beta-hydroxybutyrate, an energy substrate avidly consumed by the heart and kidney. *Frontiers In Nutrition*. **7** pp. 503509 (2020)

[168] Chu, Y., Zhang, C. & Xie, M. Beta-hydroxybutyrate, friend or foe for stressed hearts. *Frontiers In Aging*. **2** pp. 681513 (2021)

[169] Fritzsche, I., Bührdel, P., Melcher, R. & Böhme, H. Stability of ketone bodies in serum in dependence on storage time and storage temperature.. *Clinical Laboratory*. **47**, 399-403 (2001)

[170] Newman, J. & Verdin, E. beta-hydroxybutyrate: much more than a metabolite. *Diabetes Research And Clinical Practice*. **106**, 173-181 (2014)

[171] Ma, X., Dong, Z., Liu, J., Ma, L., Sun, X., Gao, R., Pan, L., Zhang, J., An, J., Hu, K. & Others beta-Hydroxybutyrate exacerbates hypoxic injury by inhibiting HIF-1alpha-dependent glycolysis in cardiomyocytes—adding fuel to the fire?. *Cardiovascular Drugs And Therapy*. pp. 1-15 (2022)

[172] Chun, Y., Kim, M. & Park, J. Oxygen-dependent and-independent regulation of HIF-1alpha.. *Journal Of Korean Medical Science*. **17**, 581 (2002)

[173] Semenza, G. Targeting HIF-1 for cancer therapy. *Nature Reviews Cancer*. **3**, 721-732 (2003)

[174] Kappler, M., Pabst, U., Weinholdt, C., Taubert, H., Rot, S., Kaune, T., Kotrba, J., Porsch, M., Güttler, A., Bache, M. & Others Causes and consequences of a glutamine induced normoxic HIF1 activity for the tumor metabolism. *International Journal Of Molecular Sciences*. **20**, 4742 (2019)

[175] Yang, Y., Lane, A., Ricketts, C., Sourbier, C., Wei, M., Shuch, B., Pike, L., Wu, M., Rouault, T., Boros, L. & Others Metabolic reprogramming for producing energy and reducing power in fumarate hydratase null cells from hereditary leiomyomatosis renal cell carcinoma. *PLoS One*. **8**, e72179 (2013)

[176] Jeridi, A., Kapellos, T. & Yildirim, A. Fumarate hydratase: a new checkpoint of metabolic regulation in inflammatory macrophages. *Signal Transduction And Targeted Therapy*. **8**, 332 (2023)

[177] Shanmugasundaram, K., Nayak, B., Shim, E., Livi, C., Block, K. & Sudarshan, S. The oncometabolite fumarate promotes pseudohypoxia through noncanonical activation of NF-KB signaling. *Journal Of Biological Chemistry*. **289**, 24691-24699 (2014)

[178] Ooi, A. & Furge, K. Fumarate hydratase inactivation in renal tumors: HIF1alpha, NRF2, and "cryptic targets" of transcription factors. *Chinese Journal Of Cancer*. **31**, 413 (2012)

[179] Yang, M., Soga, T., Pollard, P. & Adam, J. The emerging role of fumarate as an oncometabolite. *Frontiers In Oncology*. **2** pp. 85 (2012)

[180] Zheng, L., Cardaci, S., Jerby, L., MacKenzie, E., Sciacovelli, M., Johnson, T., Gaude, E., King, A., Leach, J., Edrada-Ebel, R. & Others Fumarate induces redox-dependent senescence by modifying

glutathione metabolism. *Nature Communications*. **6**, 6001 (2015)

[181] Manuel, A., Walla, M., Dorn, M., Tanis, R., Piroli, G. & Frizzell, N. Fumarate and oxidative stress synergize to promote stability of C/EBP homologous protein in the adipocyte. *Free Radical Biology And Medicine*. **148** pp. 70-82 (2020)

[182] Giallongo, S., Costa, F., Longhitano, L., Giallongo, C., Ferrigno, J., Tropea, E., Vicario, N., Li Volti, G., Parenti, R., Barbagallo, I. & Others The Pleiotropic Effects of Fumarate: From Mitochondrial Respiration to Epigenetic Rewiring and DNA Repair Mechanisms. *Metabolites*. **13**, 880 (2023)

[183] Hooftman, A., Peace, C., Ryan, D., Day, E., Yang, M., McGettrick, A., Yin, M., Montano, E., Huo, L., Toller-Kawahisa, J. & Others Macrophage fumarate hydratase restrains mtRNA-mediated interferon production. *Nature*. **615**, 490-498 (2023)

[184] Zecchini, V., Paupe, V., Herranz-Montoya, I., Janssen, J., Wortel, I., Morris, J., Ferguson, A., Chowdury, S., Segarra-Mondejar, M., Costa, A. & Others Fumarate induces vesicular release of mtDNA to drive innate immunity. *Nature*. **615**, 499-506 (2023)

[185] Zhang, Y., Wang, S., Hu, H. & Li, X. A systematic study of HIF1A cofactors in hypoxic cancer cells. *Scientific Reports*. **12**, 18962 (2022)

[186] Hiroi, M., Mori, K., Sakaeda, Y., Shimada, J. & Ohmori, Y. STAT1 represses hypoxia-inducible factor-1-mediated transcription. *Biochemical And Biophysical Research Communications*. **387**, 806-810 (2009)

[187] Wang, J., Gessler, D., Zhan, W., Gallagher, T. & Gao, G. Adeno-associated virus as a delivery vector for gene therapy of human diseases. *Signal Transduction And Targeted Therapy*. **9**, 78 (2024)

[188] Lundberg, S. A unified approach to interpreting model predictions. *ArXiv Preprint ArXiv:1705.07874*. (2017)

[189] Gurazada, S., Kennedy, H., Braatz, R., Mehrman, S., Polson, S. & Rombel, I. HEK-omics: The promise of omics to optimize HEK293 for recombinant adeno-associated virus (rAAV) gene therapy manufacturing. *Biotechnology Advances*. pp. 108506 (2024)

[190] Bordel, S., Agren, R. & Nielsen, J. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Computational Biology*. **6**, e1000859 (2010)

[191] Ezzeddini, R., Taghikhani, M., Somi, M., Samadi, N. & Rasaee, M. Clinical importance of FASN in relation to HIF-1alpha and SREBP-1c in gastric adenocarcinoma. *Life Sciences*. **224** pp. 169-176 (2019)

[192] Chen, R., Lai, U., Zhu, L., Singh, A., Ahmed, M. & Forsyth, N. Reactive oxygen species formation in the brain at different oxygen levels: the role of hypoxia inducible factors. *Frontiers In Cell And Developmental Biology*. **6** pp. 132 (2018)

[193] Liang, S., Dong, S., Liu, W., Wang, M., Tian, S., Ai, Y. & Wang, H. Accumulated ROS activates HIF-1alpha-induced glycolysis and exerts a protective effect on sensory hair cells against noise-induced damage. *Frontiers In Molecular Biosciences*. **8** pp. 806650 (2022)

[194] Zeidan, E., Hossain, M., El-Daly, M., Abourehab, M., Khalifa, M. & Taye, A. Mitochondrial regulation of the hypoxia-inducible factor in the development of pulmonary hypertension. *Journal Of Clinical Medicine*. **11**, 5219 (2022)

[195] Frezza, C., Zheng, L., Folger, O., Rajagopalan, K., MacKenzie, E., Jerby, L., Micaroni, M., Chaneton, B., Adam, J., Hedley, A. & Others Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature*. **477**, 225-228 (2011)

[196] Schlögel, G., Lück, R., Kittler, S., Spadiut, O., Kopp, J., Zanghellini, J. & Gotsmy, M. Optimizing bioprocessing efficiency with OptFed: Dynamic nonlinear modeling improves product-to-biomass yield. *Computational And Structural Biotechnology Journal*. **23** pp. 3651-3661 (2024)

[197] Gotsmy, M., Erian, A., Marx, H., Pflügl, S. & Zanghellini, J. Predictive dynamic control accurately maps the design space for 2, 3-butanediol production. *Computational And Structural Biotechnology Journal*. **23** pp. 3850-3858 (2024)

[198] Gotsmy, M., Strobl, F., Weiß, F., Gruber, P., Kraus, B., Mairhofer, J. & Zanghellini, J. Sulfate limitation increases specific plasmid DNA yield and productivity in E. coli fed-batch processes. *Microbial Cell Factories*. **22**, 242 (2023)

[199] Tyrakis, P., Yurkovich, M., Sciacovelli, M., Papachristou, E., Bridges, H., Gaude, E., Schreiner, A., D'Santos, C., Hirst, J., Hernandez-Fernaud, J. & Others Fumarate hydratase loss causes combined respiratory chain defects. *Cell Reports*. **21**, 1036-1047 (2017)

[200] Rustin, P., Munnich, A. & Rötig, A. Succinate dehydrogenase and human diseases: new insights into a well-known enzyme. *European Journal Of Human Genetics*. **10**, 289-291 (2002)

[201] Mills, E., Kelly, B., Logan, A., Costa, A., Varma, M., Bryant, C., Tourlomousis, P., Däbritz, J., Gottlieb, E., Latorre, I. & Others Succinate dehydrogenase supports metabolic repurposing of mitochondria to drive inflammatory macrophages. *Cell*. **167**, 457-470 (2016)

[202] Munk, S., Merchut-Maya, J., Adelantado Rubio, A., Hall, A., Pappas, G., Milletti, G., Lee, M., Johnsen, L., Guldberg, P., Bartek, J. & Others NAD+ regulates nucleotide metabolism and genomic DNA replication. *Nature Cell Biology*. **25**, 1774-1786 (2023)

[203] Im, D. & Muzyczka, N. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell*. **61**, 447-457 (1990)

[204] Rocklin, M. Dask: Parallel computation with blocked algorithms and task scheduling. *Proceedings Of The 14th Python In Science Conference*. (2015)

[205] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal Of Machine Learning Research*. **12** pp. 2825-2830 (2011)

[206] Kim, J., Tchernyshyov, I., Semenza, G. & Dang, C. HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metabolism*. **3**, 177-185 (2006)

[207] Leone, A., Roca, M., Ciardiello, C., Costantini, S. & Budillon, A. Oxidative stress gene expression profile correlates with cancer patient poor prognosis: identification of crucial pathways might select novel therapeutic approaches. *Oxidative Medicine And Cellular Longevity*. **2017**, 2597581 (2017)

# Supplementary Information

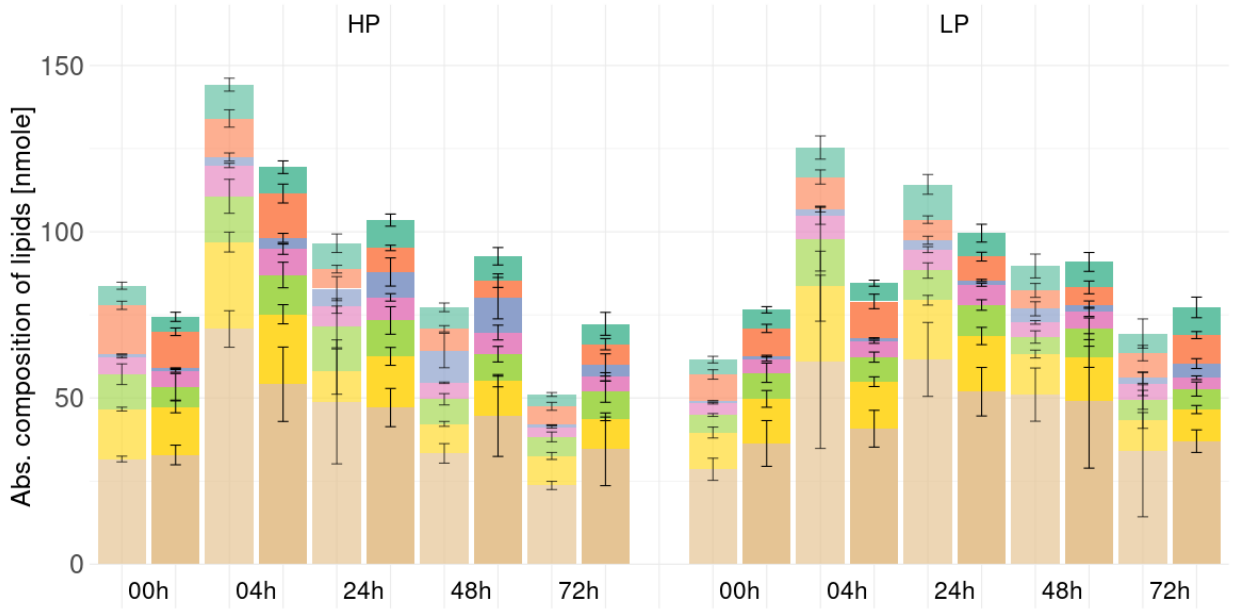


**Figure S1 : Absolute lipid composition of HP and LP strains during AAV 8 production and during growth.** Lipid profiles were quantified in four replicates for every timepoint, and state. TR is in full colors, MO is transparent.

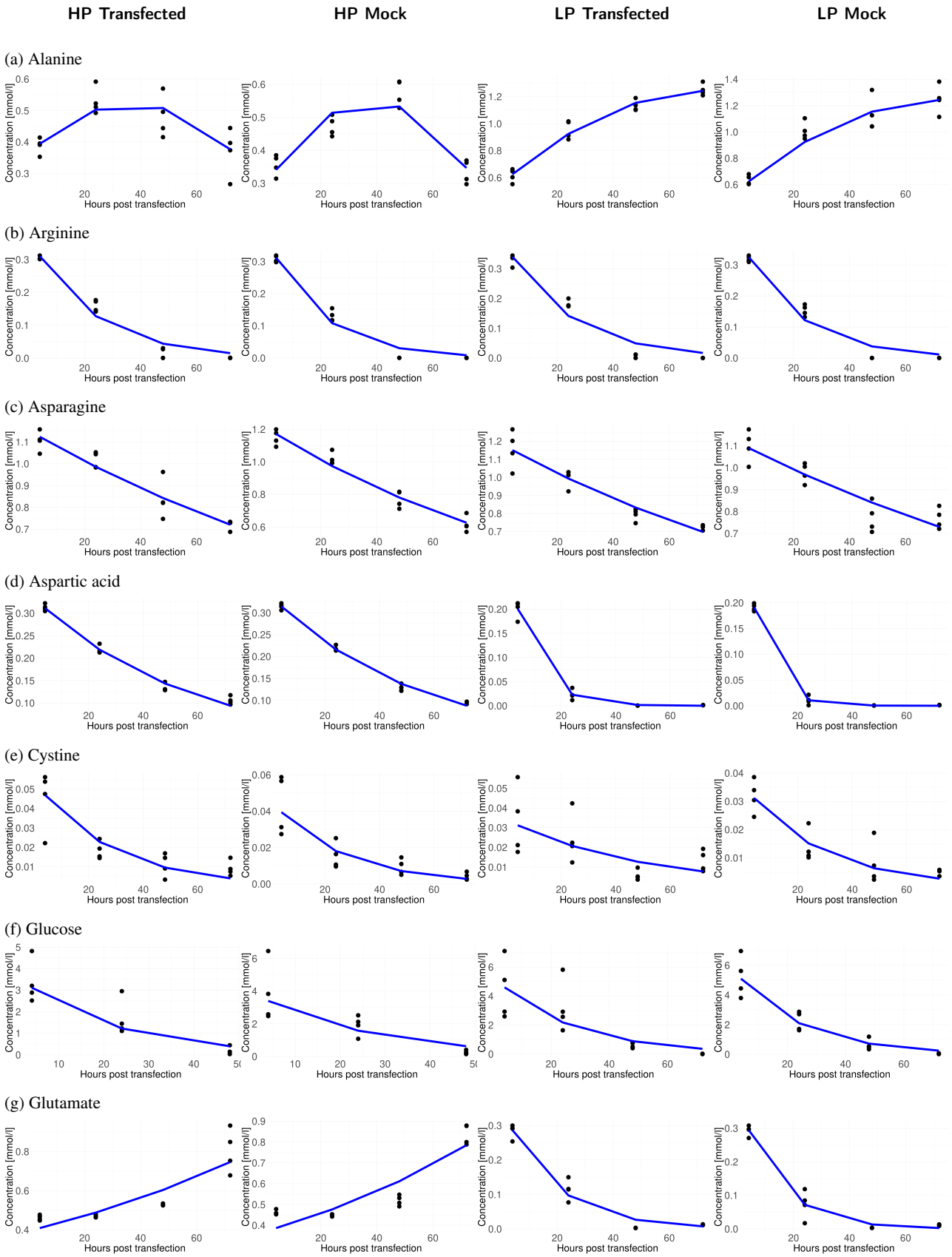


**Figure S2 : Metabolic profiles from spent media analysis for metabolites including 4, 24, 48 and 72 HPT for HP and LP when producing AAV 8 (transfected) or not (mock).** Dots indicate biological repliactes, curves were fitted using the nlrob function in R. See methods for details.

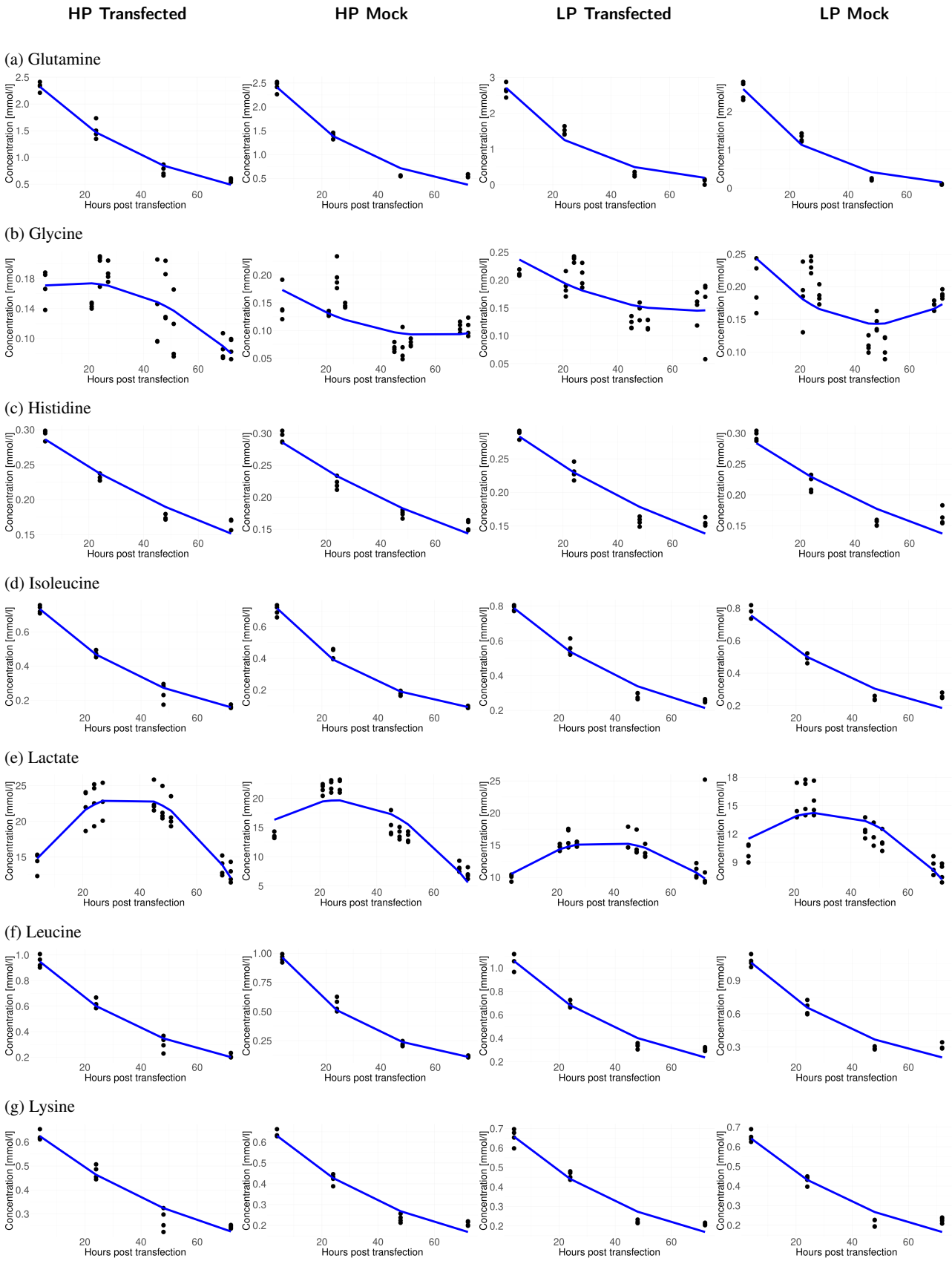


**Figure S3 : Metabolic profiles from spent media analysis for metabolites including 4, 24, 48 and 72 HPT for HP and LP when producing AAV 8 (transfected) or not (mock).** Dots indicate biological repliactes, curves were fitted using the nlrob function in R. See methods for details.

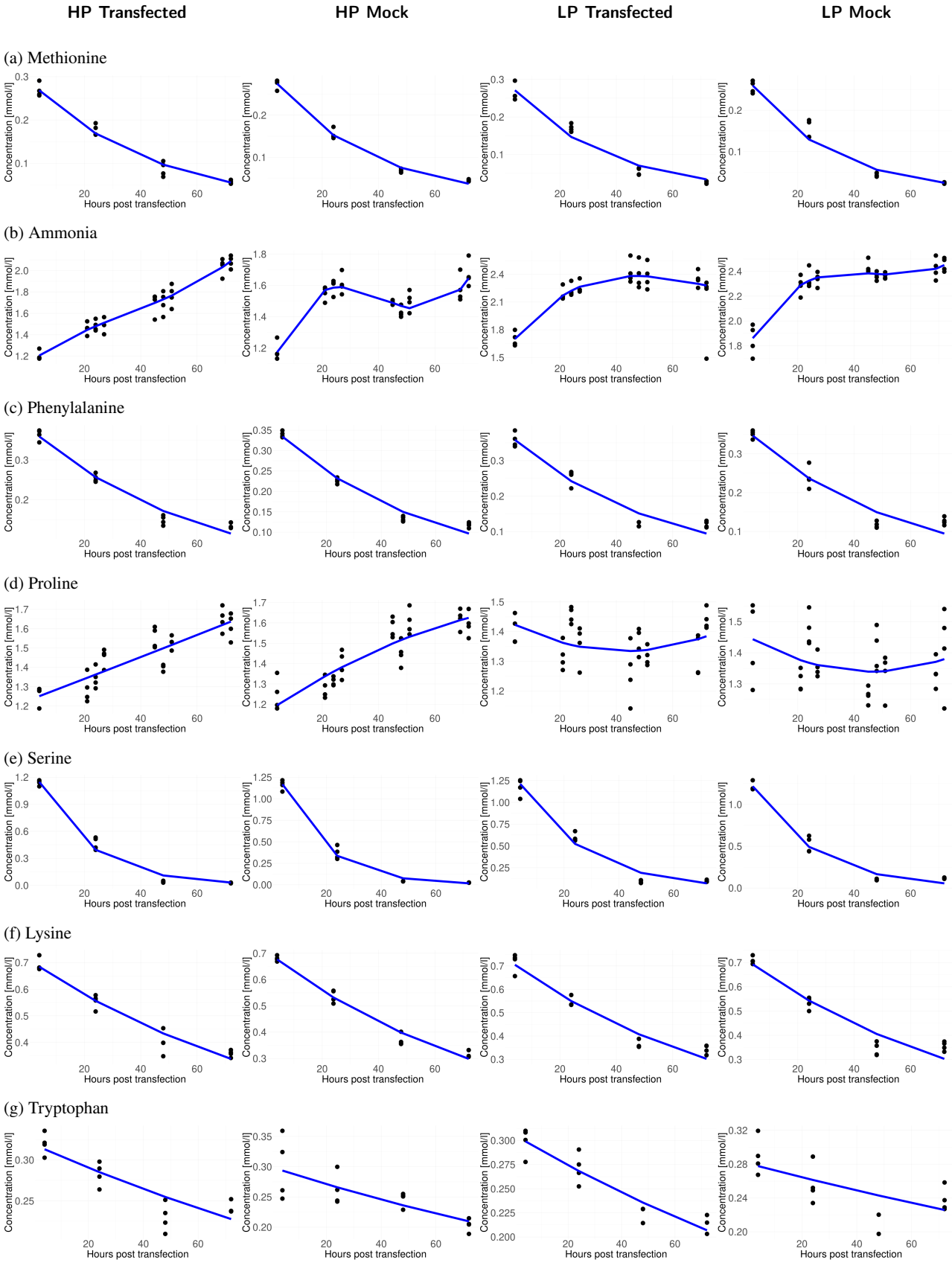


**Figure S4 : Metabolic profiles from spent media analysis for metabolites including 4, 24, 48 and 72 HPT for HP and LP when producing AAV 8 (transfected) or not (mock).** Dots indicate biological repliactes, curves were fitted using the nlrob function in R. See methods for details.

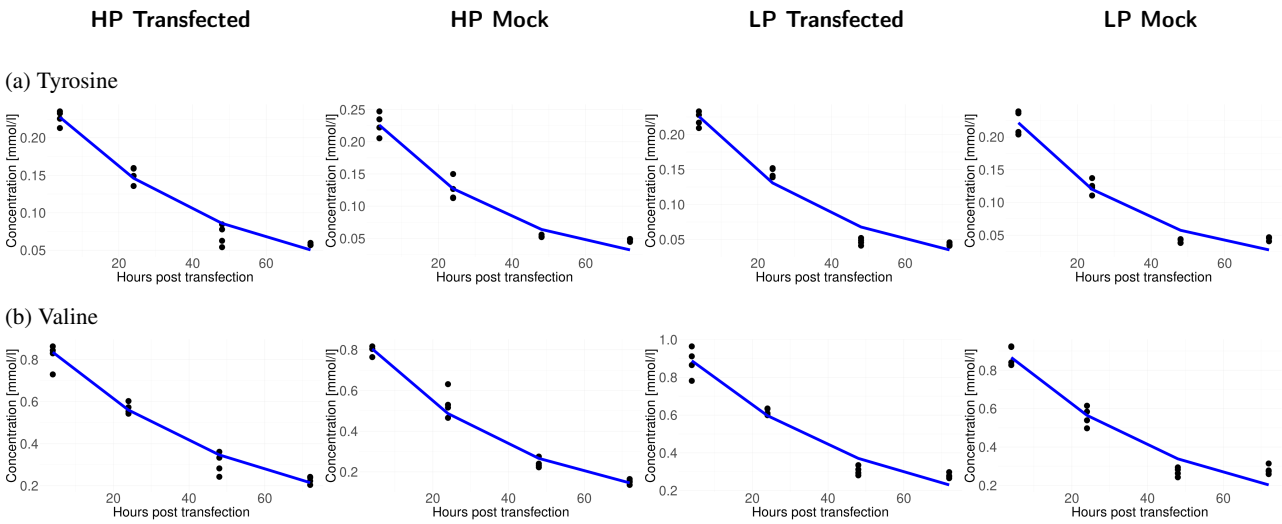


**Figure S5 : Metabolic profiles from spent media analysis for metabolites including 4, 24, 48 and 72 HPT for HP and LP when producing AAV 8 (transfected) or not (mock).** Dots indicate biological repliactes, curves were fitted using the nlrob function in R. See methods for details.

## AAV production reactions

$$51\,\text{ala} + 31\,\text{arg} + 58\,\text{asn} + 37\,\text{asp} + 5\,\text{cys} + 48\,\text{gln} + 34\,\text{glu} + 69\,\text{gly} + 14\,\text{his} + 24\,\text{ile} + 53\,\text{leu} + 32\,\text{lys} + 11\,\text{met} + 32\,\text{phe} + 57\,\text{pro} + 49\,\text{ser} + 55\,\text{thr} + 15\,\text{trp} + 33\,\text{tyr} + 30\,\text{val} + 2215\,H_2O + 2214\,\text{GTP} \longrightarrow 2214\,\text{GDP} + 2214\,\text{Pi} + 1660\,H^+ + 1\,VP_1$$

$$34\,\text{ala} + 25\,\text{arg} + 51\,\text{asn} + 25\,\text{asp} + 5\,\text{cys} + 39\,\text{gln} + 23\,\text{glu} + 54\,\text{gly} + 12\,\text{his} + 23\,\text{ile} + 37\,\text{leu} + 22\,\text{lys} + 10\,\text{met} + 28\,\text{phe} + 48\,\text{pro} + 47\,\text{ser} + 54\,\text{thr} + 12\,\text{trp} + 27\,\text{tyr} + 25\,\text{val} + 1804\,H_2O + 1803\,\text{GTP} \longrightarrow 1803\,\text{GDP} + 1803\,\text{Pi} + 1352\,H^+ + 1\,VP_2$$

$$30\,\text{ala} + 21\,\text{arg} + 49\,\text{asn} + 22\,\text{asp} + 5\,\text{cys} + 34\,\text{gln} + 20\,\text{glu} + 45\,\text{gly} + 12\,\text{his} + 22\,\text{ile} + 35\,\text{leu} + 17\,\text{lys} + 10\,\text{met} + 27\,\text{phe} + 35\,\text{pro} + 40\,\text{ser} + 50\,\text{thr} + 12\,\text{trp} + 27\,\text{tyr} + 22\,\text{val} + 1605\,H_2O + 1605\,\text{GTP} \longrightarrow 1605\,\text{GDP} + 1605\,\text{Pi} + 1204\,H^+ + 1\,VP_3$$

$$5\,VP_1 + 5\,VP_2 + 50\,VP_3 \longrightarrow 1\,\text{capsid}$$

$$1175\,\text{ATP} + 1175\,\text{GTP} + 1175\,\text{CTP} + 1175\,\text{TTP} \longrightarrow \text{transgene}$$

$$1\,\text{transgene} + 1\,\text{capsid} \longrightarrow 1\,\text{full capsid (nucleus)}$$

$$1\,\text{full capsid (nucleus)} \longrightarrow 1\,\text{full capsid (cytoplasm)}$$

$$1\,\text{full capsid (cytoplasm)} \longrightarrow 1\,\text{full capsid (extracellular space)}$$

$$1\,\text{full capsid (extracellular space)} \longrightarrow$$

Betreff **Fwd: Decision on submission to Metabolic Engineering**
Von Diana Széliová <diana.szeliova@univie.ac.at>
An Jürgen Zanghellini <juergen.zanghellini@univie.ac.at>, <Leopold.Zehetner@gmx.at>, <barbara.kraus.72@gmail.com>, <juan_antonio.hernandez_bort@fh-campuswien.ac.at>, <Leopold.Zehetner@univie.ac.at>
Datum 19.03.2025 20:06

Congratulations everyone!

-------- Originalnachricht --------
Betreff: Decision on submission to Metabolic Engineering
Datum: 19.03.2025 20:00
Von: Metabolic Engineering <em@editorialmanager.com>
An: Diana Széliová <diana.szeliova@univie.ac.at>
Antwort an: Metabolic Engineering <support@elsevier.com>

Manuscript Number: MBE-D-24-00393R2

Improving HEK293-based AAV-production using GSMMs, and a multi-omics approach

Dear Dr Széliová,

Thank you for submitting your manuscript to Metabolic Engineering.

I am pleased to inform you that your manuscript has been accepted for publication.

My comments, and any reviewer comments, are below.
Your accepted manuscript will now be transferred to our production department. We will create a proof which you will be asked to check, and you will also be asked to complete a number of online forms required for publication. If we need additional information from you during the production process, we will contact you directly.

We appreciate you submitting your manuscript to Metabolic Engineering and hope you will consider us again for future submissions.

We encourage authors of original research papers to share the research objects – including raw data, methods, protocols, software, hardware and other outputs – associated with their paper. More information on how our open access Research Elements journals can help you do this is available at https://www.elsevier.com/authors/tools-and-resources/research-elements-journals?dgcid=ec_em_research_elements_email.

Kind regards,

Editorial Board

Metabolic Engineering

Editor and Reviewer comments:

Reviewer #1: The authors have addressed the reviewer's comments /concerns.

More information and support

FAQ: When and how will I receive the proofs of my article?
https://service.elsevier.com/app/answers/detail/a_id/6007/p/10592/supporthub/publishing/related/

You will find information relevant for you as an author on Elsevier’s Author Hub:
https://www.elsevier.com/authors

FAQ: How can I reset a forgotten password?
https://service.elsevier.com/app/answers/detail/a_id/28452/supporthub/publishing/
For further assistance, please visit our customer service site:
https://service.elsevier.com/app/home/supporthub/publishing/
Here you can search for solutions on a range of topics, find answers to frequently asked questions, and

learn more about Editorial Manager via interactive tutorials. You can also talk 24/7 to our customer support team by phone and 24/7 by live chat and email

At Elsevier, we want to help all our authors to stay safe when publishing. Please be aware of fraudulent messages requesting money in return for the publication of your paper. If you are publishing open access with Elsevier, bear in mind that we will never request payment before the paper has been accepted. We have prepared some guidelines (https://www.elsevier.com/connect/authors-update/seven-top-tips-on-stopping-apc-scams ) that you may find helpful, including a short video on Identifying fake acceptance letters (https://www.youtube.com/watch?v=o5l8thD9XtE ). Please remember that you can contact Elsevier s Researcher Support team (https://service.elsevier.com/app/home/supporthub/publishing/) at any time if you have questions about your manuscript, and you can log into Editorial Manager to check the status of your manuscript (https://service.elsevier.com/app/answers/detail/a_id/29155/c/10530/supporthub/publishing/kw/status/).
#AU_MBE#

To ensure this email reaches the intended recipient, please do not delete the above code

---