GENOMICS OF VERO CELLS: UNDERSTANDING THIS CELL LINE AND ITS VIRUS-HOST INTERACTIONS FOR IMPROVED VACCINE PRODUCTION

Marie-Angélique Sène

Department of Bioengineering

McGill University, Montreal

March 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Abstract

Many vaccines were traditionally produced by growing viruses in eggs. This process is labor intensive and requires significant supply of eggs. Furthermore, during virus growth, egg-adapted changes might arise with significant implications for the body's immune response to vaccination.

Therefore, a cell-based vaccine production method has been developed as an alternative. Indeed, using cells instead of eggs to grow viruses has the potential to better protect populations by avoiding egg-adapted changes and by leading to a faster and cost-effective delivery of vaccine candidates in the event of a pandemic. Moreover, cells have been successfully used to produce various licensed vaccines such as vaccines for rotavirus, polio, smallpox, hepatitis, rubella, Ebola, and chickenpox.

The Vero cell line is the most used continuous cell line for viral vaccine manufacturing with more than 40 years of accumulated experience in the vaccine industry, emerging as an important discovery and screening tool to support the global research and development efforts including during this COVID-19 pandemic. However, the lack of a reference genome for the Vero cell line has limited our understanding of the cells' adaptation to suspension, host-virus interactions underlying the previously reported affinity of the Vero cell towards key emerging pathogens, and more importantly our ability to re-design high-yield vaccine production processes using Vero cell genome editing.

Hence, through the first aim of this project, we determined the genomic sequence of the Vero cells and identified the genes that it contains. In the second aim, we further characterized Vero cells adaptation to suspension and viral reproduction after infection by first detecting in the Vero genome, genes and pathways that are involved in adaptation to suspension and virus replication (i.e., host factors). Finally, the third aim focused on genetically modifying the cells targeting those previously identified host factors to create super producers Vero cells that will increase vaccine production quantity, quality and eventually speed.

Widely used as a platform for various studies ranging from virus vaccine manufacturing to

virus culture for other applications, the Vero cell line has the potential to become a cost-effective, high-throughput platform globally accessible through gene editing to increase virus production and to achieve high yield production and robust scalability of bioprocesses. Thus, by providing a reference genome for Vero cells and through deep quantitative profiling of Vero cells adapted to suspension and virus replication, this study will pave the way for widespread applications of genome analysis and editing tools for the Vero cell line. Considering the use of Vero cells in vaccine manufacturing processes and in particular the acceptance of this cell line by regulatory authorities, successful applications of genome editing can significantly improve virus production and ultimately lower the cost of vaccine manufacturing, thus opening new possibilities to quickly generate smarter, more efficient, globally accessible, and pandemic-ready vaccine platforms to protect global populations from current and emerging diseases.

Résumé

Plusieurs vaccins étaient traditionnellement produits en cultivant des virus dans des œufs, ce qui demande beaucoup de travail et nécessite un approvisionnement important en œufs. En outre, au cours de la croissance du virus, des modifications (c'est-à-dire des modifications adaptées à l'œuf) peuvent survenir et entraîner des répercussions importantes sur la réponse immunitaire de l'organisme à la vaccination.

Par conséquent, une méthode de production de vaccins à base de cellules a été développée comme alternative. En effet, l'utilisation de cellules au lieu d'œufs pour cultiver des virus a le potentiel de mieux protéger les populations en évitant les modifications adaptées à l'œuf et en permettant un démarrage plus rapide et à moindre coût du processus de préparation du candidat vaccin en cas de pandémie. En outre, les cellules Vero ont été utilisées pour produire divers vaccins homologués tels que les vaccins contre le rotavirus, la polio, la variole, l'hépatite, la rubéole, le virus Ebola et la varicelle.

La lignée cellulaire Vero est la lignée cellulaire continue la plus utilisée pour la fabrication de vaccins viraux, avec plus de 40 ans d'expérience accumulée dans l'industrie du vaccin. Elle est en train de devenir un outil important de découverte et de criblage pour soutenir les efforts mondiaux de recherche et de développement, y compris dans le cadre de la pandémie de COVID-19. Cependant, l'absence d'un génome de référence pour la lignée cellulaire Vero a limité notre compréhension de l'adaptation des cellules à la suspension, des interactions hôte-virus qui sous-tendent l'affinité précédemment signalée de la cellule Vero pour les principaux pathogènes émergents et, plus important encore, notre capacité à reconcevoir des processus de production de vaccins à haut rendement en utilisant l'édition du génome des cellules Vero.

Ainsi, dans le premier objectif de ce projet, nous avons déterminé l'identité génomique des cellules Vero et identifié les gènes qu'elles contiennent. Dans le deuxième objectif, nous avons caractérisé davantage l'adaptation des cellules Vero à la suspension et la croissance virale à l'intérieur de ces cellules en détectant d'abord dans le génome Vero, les gènes et les voies qui sont impliqués dans l'adaptation à la suspension et la reproduction du virus (c'est-à-dire les

facteurs de l'hôte). Ensuite, l'objectif 3 s'est concentré sur la modification génétique de ces facteurs hôtes précédemment identifiés pour créer des cellules Vero super productrices qui augmenteront la quantité, la qualité et la vitesse de production des vaccins.

La lignée cellulaire Vero étant utilisée comme plateforme pour diverses études allant de la fabrication de vaccins viraux à la culture de virus pour d'autres applications, elle a le potentiel de devenir une plateforme rentable, à haut débit, accessible à l'échelle mondiale par modification génétique pour augmenter la production de virus et obtenir une production à haut rendement et une évolutivité robuste des processus. Ainsi, en fournissant un génome de référence pour les cellules Vero et grâce à un profilage quantitatif profond des cellules adaptées à la suspension et des cellules infectées, cette étude ouvrira la voie à une application généralisée des outils d'analyse et d'édition du génome pour la lignée cellulaire Vero. Compte tenu de l'utilisation des cellules Vero dans les processus de fabrication des vaccins et, en particulier, de l'acceptation de cette lignée cellulaire par les autorités réglementaires, les applications réussies de l'édition du génome peuvent améliorer de manière significative la production de virus et, en fin de compte, réduire le coût de la fabrication des vaccins, ouvrant ainsi de nouvelles possibilités pour générer rapidement des plateformes vaccinales plus « intelligentes », plus efficaces, accessibles dans le monde entier et prêtes à être utilisées en cas de pandémie, afin de protéger les populations mondiales contre les maladies actuelles et émergentes.

Acknowledgements

بسم الله الرحمن الرحيم

Je tiens tout d'abord à dédier cette thèse à ma mère et mes sœurs :

Sama Yaye boye, ma petite maman adorée, merci pour ton altruisme inconditionnel, ton soutien indéfectible, tes innombrables sacrifices qui m'ont ouvert de nombreuses portes pour grandir et réussir, et toutes les petites (et grandes) choses que tu m'as apprises en cours de route. Je suis extrêmement fière d'être ta fille !

À mes sœurs Khady, Awa et Astou, merci de m'avoir toujours soutenue et d'avoir été là pour moi au fil des années. Cela m'a donné confiance et m'a poussé à travailler dur et à poursuivre mes rêves sans limites. J'ai beaucoup de chance de vous avoir !

Un grand merci aussi à Elhadji, merci pour tes conseils, ta disponibilité et pour m'avoir permis de me sentir chez moi au Canada.

I would also like to express my sincere gratitude and warm appreciation to the following persons:

To my thesis supervisor, Dr. Amine Kamen, thank you for introducing me to the field of Bioprocess and Vaccine manufacturing and to the possibilities to not only contribute to the overall advancement of knowledge but also to provide concrete solutions for the improvement of global health accessible to developing countries like my own. Thank you for allowing me to join the Viral Vectors and Vaccines Bioprocessing Group giving me the opportunity to train alongside and learn from highly qualified colleagues in such an excellent environment. Thank you for being my mentor, for guiding me throughout these years, for teaching me to look at the big picture, for always respecting my ideas and for giving me freedom of research while steering me in the right direction when needed.

To my thesis co-supervisor, Dr. Brandon Xia, thank you for accompanying me through this PhD. Thank you for always taking the time to help me navigate the field of Computational Biology and

for your critical timely reviews. Thank you also for teaching me how to handle the tumultuous and sometimes emotionally draining manuscript review process.

To both Dr. Amine Kamen and Dr. Brandon Xia, thank you for your unwavering support and for all your valuable advice that helped grow not only as a scientist but also as a person.

I would like to express my profound gratitude to my committee members Prof. Christine Tardif, Prof. Ioannis Ragoussis and Prof. Amin Emad for their time, support, and critical evaluation of my research progress. Your tough but fair feedbacks challenged me, helped me grow and acquire high-quality research skills.

To all my co-authors and others who contributed to the success of this Vero genomics project, thank you for helping me shape and conduct this project.

I would also like to thank the Viral Vectors and Vaccines Bioprocessing Group members:

To all the past and current senior members: Alice, Omar, Adeline, Cecile, Jean-Francois, Alina, Hanan, Shini, Mounia, Kumar and Zeyu, thank you for your valuable guidance and support. A special thanks to Omar, Alina, and Shini for patiently training me and walking me through cell culture, gene editing and analytical processes and for their great tips!

Also thank you Alina and Shini for the great and endless conversations both science and non-science related. I really enjoyed and value those moments!

To all the other past and current lab members: Pranav, Sascha, David, Aline, Michelle, Pablo, Sean,Cindy, Xingge, Barbara, Christina, Patrick, Yichen, Julia, Milica, and Divya for being such wonderful colleagues. My warmest welcome to our new members Rim and Faizul, I am excited to see what you will accomplish in the upcoming years.

A special thanks to Sascha and Julia for their valuable assistance with Vero cells culture and their great advice.

My words cannot express how grateful I am to count all of you as my colleagues and to have spent such valuable moments with you over the past three years.

I would also like to thank all faculty and staff from the Department of Bioengineering and the Department of Biomedical Engineering, especially Xavier Elisseeff, Pina Sorrini, Sabrina Teoli, Antonella Fratino, Aimée Jabour and Luisa Seidl, for all their help.

Special thanks to Pina who was always available to help me through administrative processes. Thank you for allowing me to use the “Pina card” every time I was confused and lost !

To my wonderful “ crazy ajummas” Elena and Sofia, one of the best things that happened to me was meeting you. Not only did you become my family in Korea but whenever where I am your support, encouragement and your advice never fail to reach me. I am extremely grateful to have you as my soul sisters!

To my Riima, thank you for your unwavering love and support all these years. Thank you for believing in me even when I didn’t.

I am also truly grateful for the financial support provided by the McGill Faculty of Engineering through the MEDA scholarship.

Contribution of Authors

The present thesis consists of three original research manuscripts - one published (Chapter 3), one under review (Chapter 4) and one to be submitted (Chapter 5). I am the first author of all three manuscripts.The contributions of all authors to each manuscript are listed below.

Chapter 3:	**Haplotype-Resolved de novo Assembly of the Vero Cell Line Genome**
Published in	*NPJ Vaccines*
Authors:	**Marie-Angélique Sène**, Sascha Kiesslich, Haig Djambazian, Ioannis Ragoussis, Yu Xia, Amine A. Kamen
Contributions:	**M.A-S.** designed all experiments, executed most experiments, and prepared the manuscript. S.K. cultured the cells and prepared the samples for sequencing at the McGill Genome Centre. H.D. performed the raw reads filtering and Supernova first draft assembly. I.R., Y.X., and A.A.K. provided guidance, supervision, and critical reading of the manuscript. A.A.K. provided project original design and supervision.

Chapter 4:	**Comparative transcriptomics analyses of a Vero cell line in suspension versus adherent culture conditions**
Under review in	*Scientific Reports*
Authors:	**Marie-Angélique Sène**, Yu Xia, Amine A. Kamen
Contributions:	**M.-A.S.** designed, executed all experiments, and prepared the manuscript. Y.X., and A.A.K. pro-vided guidance, supervision, and critical reading of the manuscript.

Chapter 5:	**From Functional Genomics of Vero cells to CRISPR-based Genomic deletion: Understanding Host-Virus Interactions for improved viral production rates**
Status:	Submission on hold due to IP filing
Authors:	**Marie-Angélique Sène**, Yu Xia, Amine A. Kamen
Contributions:	**M.-A.S.** designed, executed all experiments, and prepared the manuscript. Y.X., and A.A.K. provided guidance, supervision, and critical reading of the manuscript.

Contribution to the Original Knowledge

The research project presented in this thesis was structured around 3 main aims covered in Chapters 3 to 5 (Chapter 2 is a literature review of recent advances in genomics studies of Vero cells):

Generation of a Vero cell reference genome assembly:

In Chapter 3, we present a *de novo* assembly and annotation of two pseudo haplotypes for the Vero cell line, thus generating an annotated Vero genome that can be used as a baseline for screening approaches such as more rigorous gene expression profiling, pathway enrichment analysis and even the generation of a CRISPR screening library for Vero cells. As such, our study provides a genomic tool for a better understanding of the Vero cell line and its interactions with viruses, as well as for the design of more efficient cell engineering strategies such as CRISPR screenings and gene editing. Given the identification of Vero cells as an aneuploid cell line, it is necessary to fully characterize the heterogeneity of the Vero cell population using tools such as single cell analysis to detect cellular heterogeneity and more precisely identify major subpopulations emerging from that heterogeneity. Due to the lack of a fully annotated reference genome to assist in single cell analysis, we took a first preliminary step proposing this haplotype-resolved draft assembly genome giving a first glance into the heterogeneity of this continuous cell line by comparing the two haplotypes.

Moreover, comparing this assembly with the African Green Monkey genome helped unravel genomic events explaining Vero cell characteristics as a cell culture platform.

Notably, the interchromosomal translocations, previously discovered via karyotyping, were identified alongside other large-scale and small-scale structural variants through variant calling and can be used to fully characterize the interchromosomal events in Vero cells. Furthermore, the effects of those genomic rearrangements on the functionality of gene products further explained the continuous nature of the cell line and its relative high susceptibility to infection.

In addition to the Simian retroviral sequence insertions previously identified, the analysis of this Vero cell genome further showed the insertion of several complete viral genomes including the human endogenous retroviral sequence, proviral sequences, sarcoma viruses and adeno-

associated viruses which might provide insights on developing efficient downstream processing steps and quality control tools for manufacturing biologics. Furthermore, the availability of an annotated Vero cell line genome provides new possibilities for viral sequence clearance through gene editing. Also, the proviral genes identified as having lost their function are involved in key stages of the reproduction cycle of retroviruses, leukemia viruses, influenza virus and adenoviruses, despite the wide use of Vero cells as a highly susceptible cell line. Given the success of the Vero cell line as a virus production platform, the predicted loss of function of proviral genes might appear as counter intuitive. However, given that these proviral genes are highly specific to their associated type of virus, while the loss of function might affect the production rate for that specific virus, the production rates for other viruses might not be affected. For example, in the case of influenza virus strains, while the Vero cell line was one of the first cell lines considered as a vaccine production platform for influenza vaccines, its relatively low production yield of viral particles limited the use of Vero cell line as a manufacturing platform. Our analysis of the Vero genome demonstrated that Influenza virus NS1A binding protein (IVNS1ABP) also lost its function, hindering the production of M2 protein and the overall Influenza virus production yield. Thus, with an annotated genome now available, new strategies can be designed to re-engineer high-yield Vero cell line for influenza vaccine production. On the other hand, the predicted loss of function might hint towards an alternative cell-virus mode of interaction used by Vero cells to bypass this predicted loss of function and still produce the virus at high rates. This might be the case for ACE2 where despite the predicted loss of function and the confirmation of partial loss of function for the ACE2 gene (receptor for SARS-CoV and SARS-CoV-2 host cell entry), Vero cell line has a high susceptibility to SARS-CoV and SARS-CoV-2 infection and is consequently used to produce inactivated and attenuated COVID-19 vaccines.

Furthermore, the insertion of the complete Adeno-associated viral sequences into the Vero genome might provide new alternatives for the study of Vero cell-based production of Adeno-associated virus and eventually design of alternative production platforms for AAV serotypes.

In addition, gene editing tools have been used to create Vero suspension cell lines, which are considered to facilitate bioprocess development efforts. Nevertheless, singular genome modifications did not seem to lead to successful and sustainable results. Using the annotated genome presented here, further studies can be conducted to investigate suspension adaptation efforts. For example, large-scale screens could identify if a combination of modifications is needed.

In the long run, the findings of this study are expected to pave the way for widespread application of genome analysis, screening and editing tools for the Vero cell line. Considering the use of Vero cells in vaccine manufacturing processes and in particular the acceptance of this cell line by regulatory authorities, successful applications of genome editing can significantly improve virus production and ultimately lower the cost of vaccine manufacturing.

Characterization of Vero cells adaptation to suspension:

Over the years, significant efforts have been made to adapt Vero cells to suspension, to engineer high-throughput and scalable vaccine production platforms. However, those efforts were limited by hurdles such as low cell viability and long doubling time. To design more efficient strategies for successful adaptation to suspension cultures maintaining acceptable cell viability and doubling time, it is necessary to better understand the genetic and phenotypic changes triggered by the adaptation to the suspension state. Thus, we propose in Chapter 4 a comparative functional genomics analysis of adherent and suspension Vero cells from the gene level to the protein interaction network level.

Indeed, integrating results from differential expression analysis, metabolic pathway enrichment analyses, gene set enrichment analyses and network topology analyses highlights key genes and pathways at play during the adaptation of Vero cells to suspension and their complex checks-and-balances which could assist in the successful adaptation of Vero cells to suspension. Indeed, those key genes, notably associated with cell adhesion or hindering cell viability or doubling time, could be potentially targeted via gene editing as new strategies for the adaptation to suspension. In addition, the observed competitions between the regulation of competing pathways can be studied more in detail via targeted perturbations using gene editing tools such as CRISPR.

For instance, surprisingly, similar to what was observed by Malm et al. [Chapter 4, Reference 20] in HEK293 cells adapted to suspension, cellular component organization pathways associated with cell adhesion (such as actin filament-based process, regulation of cell adhesion, apical part of cell, plasma membrane bounded cell projection morphogenesis and extracellular matrix) are upregulated via pathway enrichment analysis, while Network Topology-based Analysis(NTA) showed an upregulation of the cell adhesion PPI network with CDH18 as its central gene and which can be considered as possible target for engineering in order to improve the cells adaptation to

suspension. This upregulation of cell adhesion related genes could be due to the cells' attempt to restore the attachment to culture surfaces and surrounding cells which could explain the aggregates that are often observed in Vero cells suspension cultures and the cell aggregation rings that form on the suspension culture shake-flasks.

Finally, the epithelial to mesenchymal transition (EMT) pathway, which was previously thought to be involved in the adaptation to suspension, and thus investigated as a potential avenue to explore, is downregulated in Vero cells adapted to suspension, thus highlighting the fact that the adaptation to suspension is not associated with EMT as previously shown with HEK293 cells adapted to suspension.

Increase of Vero cells vaccine production yield:

In Chapter 5, we present a deep analysis of virus-host interactions at play during Influenza A (IVA) and rVSV-GFP infection in Vero cells to better understand the dynamics taking place between the host attempting to minimize the impact of viral infection and the viruses attempting to evade host immune responses. Using an interdisciplinary approach combining functional genomics and cell biology, we propose a new strategy for a more efficient targeted CRISPR gene editing, opening new possibilities for the establishment of Vero cells as a pandemic-ready high-throughput vaccine production platform.

Indeed, previous attempts to generate engineered cell lines for high yield virus production relied mainly on screening data to select gene targets and then knock them out via CRISPR to verify that the previously reported viral production yield in the screening data is obtained. But, in most cases, the knockout did not lead to significant increase in viral production yield due to several factors. These factors include: (i) knock-down and knockout do not lead to the same phenotypic effects; (ii) up to 4 different guide RNAs are used for screening which increase the possibility of off target effects that could lead to an increased viral production yield in unknown ways; and (iii) another genome or cell line is used for screening compared to the actual cell line used for targeted gene editing. With the recent publication of the de novo assembled and annotated Vero genome (Aim 1) and the use of functional genomics, we were able to achieve a deeper understanding of the mechanisms at play during infection and the role of the gene target candidates before selection, thus giving a better control and oversight on the gene editing experiments.

Notably, the deletion of a whole genomic region, here the coding DNA sequence (CDS) region, was preferred over a single guide RNA-based cut to increase the probability of getting a biallelic deletion, to ensure that the deletion will lead to the desired loss of function of the targeted gene product but also to make the validation step easier, thus ensuring a rapid and high-throughput gene-editing protocol.

Notably, our deletion of ISG15 in Vero cells led to the overall increase in total particles production to up to 87-fold, as well as an increase of infectious particles production ratio from 3.16% to 65.3% for IVA and from 5.42% to 67.9% for rVSV-GFP. Our results confirm previous reports that ISG15 modulates the amount of released infectious particles while intracellular viral replication remains intact. In addition, during preliminary study of IVA infection kinetics, it was shown that infectious particles production increases until 24 hours post infection (hpi) before decreasing dramatically, which was also observed during influenza infection in MDCK cells. Thus, this newly engineered cell line could present even more attractive advantages specially to produce live attenuated viruses.

Overall, the Vero cell line was one of the first cell lines considered as a vaccine production platform with various vaccines produced in Vero cells being approved and used to immunize millions of patients including in the current COVID-19 pandemic. With this new interdisciplinary approach, we successfully engineered a cell line capable of increasing virus production yield to up to 87-fold while also increasing infectious viral particles release that was around 3.16-5.42% to up to 65.3-67.9%. Thus, our study opens new avenues for the development of pandemic-ready vaccine production platforms in line with the global preparedness plan.

Table of Contents

List of Figures

List of Tables

List of Acronyms

COVID-19	coronavirus disease 2019
DMEM	Dulbecco′s Modified Eagle′s Medium
ddPCR	digital droplet polymerase chain reaction
EVD	Ebola virus disease
FBS	Fetal Bovine Serum
FDA	US Food and Drug Administration
GFP	green fluorescent protein
HEK	Human Embryonic Kidney cells
HIV	human immunodeficiency viruses
hpi	hours post-infection
MOI	multiplicity of infection
rVSV	recombinant vesicular stomatitis virus
$TCID_{50}$	Median Tissue Culture Infectious Dose
TOI	time of infection
ZEBOV	Zaire Ebola virus

Chapter 1 : Introduction

Background and State of the Art

The CHO cell line, originally derived from the Chinese hamster ovary, has become a staple source of cells over the years with regards to the establishment of the cell line as a go-to monoclonal production platform. The whole genome sequence for CHO-K1 was published in 2011. In 2016, a consensus on the first community-generated genome-scale metabolic model of CHO cells was reached (Hefzi et al., 2016). Not only did the model provide a valuable resource that led to CHO cell line development through engineering tools such as gene editing using the CRISPR/Cas9 technology (Lee, Grav, et al. 2015; Lee, Kallehauge, et al. 2015) but it also provided a blueprint for future cell line genome annotations.

The advances in gene editing have made it possible to – with high-throughput and cost-effective methods – edit the genome of cell lines using available genomic data, thus providing new possibilities for cell line development and vaccine bioprocessing intensification.
In the vaccine manufacturing space, another cell line has been extensively used for virus vaccine manufacturing: the Vero cell line. Originating from *Chlorocebus sabaeus* (African Green Monkey), the Vero cell line represents the most widely used continuous cell line to produce viral vaccines with over 40 years of experience. This includes the development and production of vaccines against dengue fever, influenza, Japanese encephalitis, polio, rabies, rotavirus and smallpox (Barrett et al. 2009; Ammerman, Beier-Sexton, and Azad 2008). Recently, some attempts to develop engineered Vero cell line have been made (van der Sanden et al. 2015).

However, genome editing of the Vero cell line still suffers from the lack of published genomic information. Despite the publication of the Vero genomic landscape (Osada et al. 2014), there are no tools available yet to efficiently select CRISPR/Cas9 target sites. This contrasts with other common vaccine manufacturing cell lines, e. g. derived from human cells and CHO cells.

Rationale

The main limitations to recent efforts to develop Vero cells as high-throughput vaccine manufacturing platforms are not only the lack of a comprehensive and well-annotated reference specific to this cell line, but also, the limited scale-up possibilities during large-scale manufacturing (which can be alleviated by adapting the adherent Vero cells to suspension) and the lack of understanding of virus-host interactions at play during infection and cell-based virus production.

Therefore, this project mainly focuses on generating a reference genome for the Vero cell line through whole genome *de novo* assembly and annotation using Linked Reads and RNA sequencing technologies. Then, the factors at play during adaptation to suspension and virus infection and reproduction in Vero host cells were identified and the best candidate for gene editing was selected using a multi-omics study on virus-infected Vero cells combined with pathway predictions. After selection of that candidate, its genomic sequence was isolated for CRISPR guide RNAs design and off targets predictions to select guide RNAs with the highest knockout efficiency score. After that final target selection, a validation CRISPR editing experiment was run to quantify the hits and virus production rates of modified cells.

Theoretical Approach

Since the Vero cell line genome has been partially sequenced using bulk sequencing methods, we propose to complete this genomic information already available using Linked reads sequencing methods. Indeed, by consolidating multiple assays into a single, high-throughput workflow with low input requirements, the Linked Reads sequencing method has emerged as a tool for obtaining detailed understanding of the whole genomic, epigenomic and transcriptional variations for complex cell lines such as Vero cells.

The Vero cell line is the most used continuous cell line for viral vaccine manufacturing. Its anchorage-dependent use renders scaling-up challenging and vaccine manufacturing operations very labor intensive which affects cost effectiveness. Thus, efforts to adapt Vero cells to suspension cultures have been invested but hurdles such as the long doubling time and low cell viability remain to be addressed. Building on the Vero cell line annotated genome that

was generated in this study, a functional genomics analysis of the Vero cells adapted to suspension is proposed to better understand the genetic and phenotypic switches at play during the adaptation of Vero cells from anchorage-dependent to suspension cultures. In addition, when it comes to deep understanding of virus-host interactions, by combining total RNA sequencing with functional genomics analysis, high-throughput and high-resolution functional genetic screens can be performed in millions of cells simultaneously, thus enabling the visualization of the transcriptomics, the assessment of comprehensive gene expression phenotypes and the specific identification of the cells subpopulations that are responsive to virus infection.

Research questions, Methods and Procedures

In this project, we hypothesise that by studying the genomics aspect of cells adaptation to suspension and host-virus interactions during infection, key pathways triggered during adaptation to suspension and host factors at play during viral infection can be analysed, resulting in the identification of the best gene target for CRISPR editing to enhance virus production in Vero cells.
Thus, this project is divided into three main aims responding to three main research questions:

Can a comprehensive whole genome assembly and annotation be achieved for a complex cell type such as Vero cells? (Aim 1): Due to the lack of a published reference for the Vero cells, the first aim was to generate a comprehensive and well annotated Vero cells reference genome via whole genome assembly and annotation, which served as the foundation for this project.

What are the host factors characterizing cells adaptation to suspension and virus-host interactions during infection? (Aim 2): The second aim focused on the detection of cellular phenotypes triggered in response to adaptation to suspension, as well as in response to virus infection. Therefore, transcriptomics profiling was conducted on Vero cells adapted to suspension and virus infected Vero cells (using the genome generated in aim 1) to uncover the changes that influence the phenotype of the suspension cells and the infected cells, thus providing information on the key antiviral and proviral factors that can be targeted through

gene editing to increase the virus production rates. The viruses of interest selected for this project are Influenza A virus (PR8 strain) and a recombinant VSV for GFP expression.

Can Vero cells be efficiently genetically edited to enhance their virus production rate? (Aim 3): The host factor selected from aim 2 study of infection effects on Vero cells host response was used as target for gene editing and validation experiments were run for maximization of infectivity. Gene editing was performed using the CRISPR/Cas9 technique and the efficiency was quantified through PCR before cell line validation.

Thesis Organization

Following the latest thesis preparation guidelines provided by McGill's Graduate and Postdoctoral Studies (GPS), this thesis is organized in a manuscript-based format. The three original research manuscripts constitute the main body of the thesis.

Chapter 1 is an introduction chapter focusing on the rationale behind this doctoral thesis, main objective, specific aims, and thesis organization.

The following chapter (Chapter 2) is a literature review on Vero cells genome characterisation and all genomic experiments previously conducted for viral production and adaptation to suspension.

Chapter 3 presents the assembly of the reference genome for the Vero cell, to respond to the need highlighted in Chapter 2 to have a highly contiguous reference-grade genome assembly of Vero cells to unlock the significant potential of this cell line to become the go-to cell line for vaccine production.

Chapter 2 also gives insights on the attempts that were made to increase the overall viral particles produced in Vero cells using two key approaches:

1. Achieving a high throughput and cost-effective scale-up of production facilities by adapting the anchorage-dependent Vero cells to suspension culture thus increasing the volumetric amount of Vero cells. But due to a lack of understanding of what is happening during the cells' adaptation to suspension and more precisely of why there are hindrances such as the low cell density and high doubling time observed, there isn't a totally successful adaptation of Vero cells to suspension.
2. Increasing the cell-specific viral production by using gene editing approaches targeting

antiviral host genes to boost virus replication in Vero cells. Again, due to the lack of understanding of the host-virus interactions at play, the cell engineering attempts did not lead to an efficient cell engineering protocol for high-yield Vero cell vaccine manufacturing platform.

Chapter 4 focuses on the first approach by aiming at the understanding of the genomics behind Vero cells adaptation to suspension. The processes identified during transcriptomics analysis of Vero cells adapted to suspension are presented to highlight possible strategies that can be used to achieve a successful adaptation of Vero cells to suspension for high yield viral production.

Chapter 5 focuses on the second approach by providing a better understanding of host-virus interactions during infection and the resulting gene editing of Vero cells for high yield virus production. Based on the findings from aim 2 and aim 3, this chapter covers the whole experiment concerning Vero cells infection with Influenza and VSV-GFP and the resulting CRISPR targeted gene editing.

Chapter 6 provides an overall comprehensive conclusion and future perspectives.

Chapter 2

Recent Advances in Vero Cells Genomic Characterization and Engineering for High-Throughput and High-Yield Vaccine Manufacturing

Authors: **Marie-Angélique Sène[1]**, Yu Xia[1], Amine A. Kamen[1,*]

Author affiliation : [1] Department of Bioengineering, McGill University, Montreal, QC, Canada

* Corresponding author

Journal: Clinical and Translational Discovery

Year: 2022

DOI: pending

Status: Accepted

Abstract

Background

The Vero cell line is the most used continuous cell line for viral vaccine manufacturing with more than 30 years of accumulated experience in the vaccine industry. Nonetheless, virus production yield with Vero cells remains limited. Therefore, given Vero cell line infection susceptibility to a wide range of viruses, alleviating limitations of viral replication and increasing production kinetics in Vero cells could significantly reduce vaccine production time and cost.

Main body of the abstract

Here we review the literature on the development of Vero cells as the most effective manufacturing platform for viral vaccines. Whereas various bioprocess development strategies were proposed to improve the production, this review focuses on the genomic characterization and genetic engineering aspects of Vero cells. Rational design of production cell line emerged as a state-of-the-art approach to significantly enhance Vero cell viral production yield and adaptation to suspension culture which aligns with the global preparedness efforts to accelerate and intensify vaccine production capacity to better respond to pandemic situations and epidemic outbreaks.

Conclusion

Until recently, the lack of a reference genome for the Vero cell line has limited the understanding of Vero cells behavior in defined culture conditions as well as host-virus interactions underlying the affinity of the Vero cell line with emerging and re-emerging pathogens. Importantly this limited our ability to re-design high-yield vaccine production processes using Vero genome editing.

Keywords : Vero, Cell culture, Process development, Virus production, Optimization, Vaccines, Functional Genomics, Suspension culture, Gene editing

1.0 INTRODUCTION

The establishment of Vero cells as an experimental tool for studies ranging from virology to toxicology started in 1962 with the extraction of the cells from the Female African Green Monkey kidney by Yasumura and Kawakita [1]. It was later discovered the Vero cell line is a continuous cell line, meaning that it can be passaged over a significant period of time without losing its growth characteristics. Importantly, this is maintained without acquisition of tumorigenic functions as compared to other primary cell lines with limited passage numbers, allowing for several sublines to be derived and cell banking. Additionally, Emeny and Morgan [2] reported that Vero cells have a deficiency in interferon expression, translating in an impaired antiviral response to infection which make this cell line susceptible to infection by a wide range of viruses. Even infected at high multiplicity of infection, Vero cells do not show an active cell death pathway generally triggered by interferon expression.

Consequently, WHO regulatory authorities recommended the use of Vero cell line as the first continuous cell line for human viral vaccines manufacturing [3] and the creation of a fully documented WHO master cell bank as reported by Barret et al. [4].

Vaccines were traditionally produced by growing viruses in eggs which is labor intensive and requires significant supply of eggs. Furthermore, during virus growth, changes (i.e., egg-adapted changes) that can have significant implications for the body's immune response to vaccination might be introduced. Therefore, cell-based vaccine production methods have been developed as an alternative. Indeed, using cells instead of eggs for viral production has the potential to cost-effectively produce protective vaccines while avoiding egg-adapted changes, as well as providing rapid responses in the event of a pandemic through a faster start-up of the vaccine preparation

process. Moreover, cell cultures have been used to produce various U.S. licensed vaccines such as rotavirus, polio, smallpox, hepatitis, rubella, and chickenpox with a well-defined regulatory path.

In general, viral vaccines manufacturing requires a specific component of the virus or the whole viral particle to be produced in high yields before adding the other components. Therefore, Vero cells susceptibility to a wide range of viruses, its ability to be infected at high multiplicity of infection and its establishment as a continuous cell line make it a valuable candidate for viral vaccine production. Thus, significant efforts have been invested to produce approved Vero cell-based vaccines being used worldwide including vaccines against Ebola virus disease, Influenza, Japanese encephalitis, Poliomyelitis, Rabies, Rotavirus, Smallpox [5-7] and more recently vaccines to control the current COVID-19 pandemic which were approved for use on millions by WHO [8] (Table 1).

Disease	Vaccine	Manufacturer	Cell Culture Method
Ebola virus disease	Ervebo®	Merck	Roller-bottle with microcarriers [9]
Poliomyelitis	IMOVAX Polio®	Sanofi Pasteur	Microcarriers [10]
Rotavirus gastroenteritis	RotaRIX®	GSK	T-flasks and multi-tray units [11]
Rotavirus gastroenteritis	RotaTeq®	Merck	T-flask and Cell factories [12]
Smallpox	ACAM2000®	Acambis	Microcarriers[13]
SARS-CoV-2	CoronaVac®	SinoVac Life Sciences	microcarriers[14]

Table 1. Cell culture methods for approved Vero cell-based vaccines

However, Vero cell-based vaccine manufacturing with adherent cells is hindered by an important viral production yield limitation. Therefore, inspired by the CHO cells journey towards its establishment as the cell line of choice for antibody production notably with the use (by Lee, Grav, et al. [15]; Lee, Kallehauge, et al.[16]) of engineering tools such as gene editing, several attempts have been made to increase virus production yield in Vero cells. These attempts relied not only on bioprocess development (optimization of cell density before infection, multiplicity of infection, time of infection, time of harvest among others), but also on the genomics of the cell line to increase volumetric viral production via the adaptation of the Vero cell line to suspension cultures and unlocking the cell-specific viral production yield using a combination of genetic screening and gene editing strategies.

2.0 GENETIC ENGINEERING FOR A SUCCESSFUL ADAPTATION TO SUSPENSION

To increase cell culture capacities and overall viral production yield, different techniques have been developed including microcarriers [17] and fixed-bed bioreactor[18], but due to the surface area growth limiting factor [19], other avenues were explored and notably the adaptation to culture in suspension of well-established cell lines such as CHO cells or HEK293 cells led to dedicating efforts to also adapt Vero cells to suspension.

Those efforts faced significant challenges for more than 20 years [20-23] until the report of Shen et al.[24] claiming an adaption to suspension of Vero cells using media formulation. A proof of principle was established by producing the recombinant Vesicular Stomatitis Virus (rVSV) using a Vero cell line adapted to suspension in batch and perfusion bioreactor cultures.

A genomic approach was also attempted. Jaluria et al. [25] reported, via DNA microarrays analysis that the upregulation of the human SIAT7E gene, which encodes a type II membrane glycosylating sialyl transferase, can reduce Hela cells adhesion. Following those results, Mehrbod et al. [26] transfected MDCK cells and Vero cells with the human SIAT7E gene and analyzed the cells viability after transfection and their ability to grow in suspension. While MDCK cells did not survive past 150 hours post transfection, the modified anchorage-independent Vero cells remained viable.

Nonetheless, due to reported limitations such as low cell growth rate, aggregation and long doubling times, it is necessary to further investigate the adaption process and the Vero cells phenotypic effects triggered in order to achieve a successful adaption of this valuable cell line to suspension.

3.0 GENETIC ENGINEERING APPROACHES FOR CELL-SPECIFIC VIRUS PRODUCTION YIELD ENHANCEMENT

Given viruses ability to hijack the host machinery for replication purposes, identifying genes involved in antiviral host response to viral infection is paramount. In order to bypass the host hindrance during viral reproduction in cells, those genes are targeted using gene editing tools to mitigate their effects.

Gene identification is usually achieved using screening strategies such as random knock-down (KD) of genes and quantitative analysis of the resulting effects on viral replication.

For instance, Van der Sanden et al. [27] performed a genome-wide RNA interference (RNAi) screen, knocking down genes to characterize those leading to an increase in poliovirus

production yield after infection with Sabin strain and wild-type poliovirus. Reporting significant increase of up to 80-fold, the authors generated thereafter stable Vero multiple knockout (KO) cell lines based on the RNAi data using CRISPR targeted gene editing [28]. The same process was used by Wu et al. for other viruses such as the rotavirus with a reported increase of up to 18-fold [29]. The knockouts were confirmed using Sanger sequencing, but the subsequent testing of virus production yield of the engineered Vero cell lines did not confirm a significant viral yield increase. The possibility that the Sanger sequencing confirmation of knockout genes does not guarantee a homozygous knockout was discussed.

Further studies by Hoeksema et al.[30] combined gene targets identified in both Van der Sanden and Wu studies to generate double knockout stable cell lines by knocking out targeted genes using CRISPR and confirming the gene edition via Next generation sequencing (NGS). Despite rigorous ELISA, $TCID_{50}$ and immunoblotting-based testing of both wild type and confirmed knockout clones, none of the clones showed similar enhancement of viral production to what was previously reported[27,29]. Furthermore, wild-type sub-clones of Vero 10-87 cells showed similar poliovirus production rates which highlighted the possible absence of knockout impact in the originally reported increase in virus production yield.

Among the listed possible factors causing these observed discrepancies, three main possible factors came to light:

For both screening and knockout experiments, the human genome and the African Green Monkey genome were respectively used. Knowing that the Vero cell line is aneuploid with noteworthy genomic rearrangements compared to the African Green Monkey first described by Osada et al.[31], it is expected that the Vero cell genome will be significantly different from both the human genome and the African Green Monkey genome. Thus, despite some genes being

conserved across species and sub-cell lines, using gene targets identified in screening processes with different genomes might lead to significant discrepancies between the primary KO effects observed and the KO experiments in Vero cells.

The difference between the Vero cell line used for screening and the Vero cell line used for knockout might also have played a significant role in the lack of reproducibility observed by Hoeksema et al.[30]. Indeed, it was previously shown that several sub cell lines derived from the same parental CHO cell line showed noteworthy variations over time [32].

Moreover, it is important to note that the phenotypes induced by transcriptional suppression (RNAi-based knockdown) and genetic deletion (CRISPR knockout) differ in such way that the former might increase virus production while the latter does not. Hence, the previously identified genes with RNAi screen and the CRISPR-based KO genes might not yield the same results.

4.0 INITIAL GENOMIC CHARACTERIZATION OF THE VERO CELL LINE

The results discussed in the previous sections further stress the importance of characterizing the Vero cell line genome, in order to better understand the virus-host dynamics during infection, as well as uncover the phenotypic behavior of Vero cells under different culture conditions, including the adaptation to suspension process. This valuable sequence information would be an important asset in the current effort to achieve high-throughput, high yield vaccine production cell platforms via a successful adaptation of Vero cells to suspension culture to enable straightforward scale-up as well as a successful enhancement of cell-specific virus production yield.

Indeed, the efforts to characterize Vero cells genomic structure started with the publication of a draft genome assembly by Osada et al.[31] that shed light on the peculiar nature of Vero cells. The lack of type I interferon response in Vero cells (9 Mb deleted region on chromosome 12), as previously shown with the homozygous deletion of α and β1 interferon genes via DNA hybridization [33], could explain the high susceptibility of Vero cells to infection for a wide range of viruses. Moreover, the continuous nature of Vero cells may also be due to the subsequent deletion of CDKN2A and CDKN2B but increased passage numbers should be avoided to reduce the risk of acquiring tumorgenicity due to unknown mutation accumulating over time. While noteworthy chromosomal rearrangements were identified through karyotyping, the translocation events could not be confirmed by sequencing. This might be due to technological limitations therefore, haplotype resolved sequencing and assembly should be performed to help unravel these translocation events and better understand and characterize the Vero cell genome.

Among the peculiar events observed in the Vero cell genome, it is necessary to highlight the presence of Simian Retroviral sequences (SRV) in the Vero genome. Indeed, a heterogenous presence of defective SRV variant sequences lacking the U3 and R regions of 3'LTR were identified in the Vero genome [31] which stresses the necessity to further characterize the Vero genome in order to ensure a better quality control of Vero-derived pharmaceutical products.

Following those SRV integrations findings, in an effort to ensure the safety of Vero cells used for vaccine production and more particularly live vaccines, Sakuma et al. [34] studied the characteristics of Simian Endogenous type D Retroviruses (SERV) insertions in the Vero genome, through transcriptomics analysis of several sublines, reporting ~80 SERV integrations. They discovered that the epigenetic repression of SERVs was released in Vero cells by DNA hypomethylation. They also confirmed the loss of type I interferon response previously identified

by Osada et al. by reporting the absence of the type I interferon gene cluster. Nonetheless, no retrotransposition of the integrated SERVs was observed following the establishment of the Vero cell line under standard cell culture and number of passage conditions.

More recently, reference-grade genome for Vero cells was published [35]. Indeed, a highly contiguous genome was assembled, thus providing a comprehensive list of the Vero cell genomic, transcriptomic, and proteomic components. Through this assembly, significant characteristics that helped better understand the peculiar nature of Vero cells were unraveled. By comparing the Vero genome to the African green monkey genome, significant genomic variants leading to the loss of function (LOF) of hundreds of genes, notably genes involved in the cell death pathway activation and the cytokine pathway for instance, were reported thus highlighting one of the reasons behind the high susceptibility of Vero cells to various viruses and its ability to be infected at high MOI without triggering the cell death pathway.

Among those genes which lost their function, several proviral genes were identified, explaining the low productivity of certain viruses such as Influenza in Vero cells but also opening new implication for the engineering of Vero cell-based manufacturing platforms.

Indeed, viral production of some viruses such as Influenza can be increased through genetic engineering by targeting those LOF genes that led to the low viral production yields.

5.0 CONCLUSION

The Vero cell line is the most used continuous cell line for viral vaccine manufacturing with more than 40 years of accumulated experience in the vaccine industry, emerging as an important discovery and screening tool to support the global research and development efforts in this COVID-19 pandemic.

The Vero cell line is indeed used as a platform for various studies ranging from virus vaccine manufacturing to virus culture for other applications (such as virus isolation, characterization and banking) and has the potential to become a cost-effective, high throughput platform globally accessible through gene editing to increase virus production and achieve high-yield production and robust scalability of processes. Several attempts have been made to genetically engineer this cell line in order to bypass the viral production yield limitations by successfully silencing host antiviral response to infection and adapting Vero cells to suspension culture to enable more effective scale-up.

However, all those efforts were met with one fact that cannot be overlooked: the lack of complete and accurate genomic information on Vero cells. That lack of a reference genome for the Vero cell line has limited our understanding of not only phenotypic events governing Vero cells adaptation to suspension, but also host-virus interactions underlying such affinity of the Vero cell towards key emerging pathogens, and more importantly our ability to re-design high-yield vaccine production processes using Vero genome editing.

Thus, by providing a reference genome for Vero cells and by exploring the interactions and phenotypes at play during adaptation to suspension and virus infection of Vero cells in order to fully characterize them and to exploit them to develop a tool for improved virus vaccine

production, recent studies [35] will pave the way for widespread application of genome analysis and editing tools for the Vero cell line. Furthermore, considering the use of Vero cells in vaccine manufacturing processes and in particular the acceptance of this cell line by regulatory authorities, successful applications of genome editing can significantly improve virus production and ultimately lower the cost of vaccine manufacturing.

6.0 ACKNOWLEDGEMENTS

This work was supported by funding from the Natural Sciences and Engineering Research Council (NSERC-Grant-CRDPJ 511957). M-A-S is supported by the McGill Engineering Doctoral Award (MEDA) and A.A.K by a Canada Research Chair (CRC-240394).

7.0 REFERENCE

1. Yasumura, Y., Kawakita, Y., 1963. The research for the SV40 by means of tissue culture technique. Nippon rinsho 21, 1201–1219.

2. Emeny JM, Morgan MJ. Regulation of the interferon system: evidence that Vero cells have a genetic defect in interferon production. *J Gen Virol*. 1979;43(1):247-252. doi:10.1099/0022-1317-43-1-247

3. World Health Organization, 1987a. Acceptability of cells substrates for production of biologicals. WHO Tech. Rep. Ser. 747, 5–24.

4. Barrett PN, Mundt W, Kistner O, Howard MK. Vero cell platform in vaccine production: moving towards cell culture-based viral vaccines. *Exp review of vacc*. 2009 May 1;8(5): 607-18.

5. Ammerman NC, Beier-Sexton M, Azad AF. Growth and maintenance of Vero cell lines. *Curr Protoc Microbiol*. 2008;Appendix 4:Appendix 4E.

6. Suder E, Furuyama W, Feldmann H, Marzi A, de Wit E. The vesicular stomatitis virus-based Ebola virus vaccine: From concept to clinical trials. *Hum Vaccin Immunother*. 2018;14(9):2107-2113.

7. Agnandji ST, Huttner A, Zinser ME, et al. Phase 1 Trials of rVSV Ebola Vaccine in Africa and Europe. *N Engl J Med*. 2016;374(17):1647-1660. doi:10.1056/NEJMoa1502924

8. COVID-19 vaccine tracker and landscape : https://www.who.int/teams/blueprint/covid-19/covid-19-vaccine-tracker-and-landscape

9. https://dc.engconfintl.org/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1022&context=vaccine_vi

10. https://pdf.hres.ca/dpd_pm/00012948.PDF

11. https://www.ema.europa.eu/en/documents/scientific-discussion/rotarix-epar-scientific-discussion_en.pdf

12. https://www.ema.europa.eu/en/documents/scientific-discussion/rotateq-epar-scientific-discussion_en.pdf

13. Monath TP, Caldwell JR, Mundt W, et al. ACAM2000 clonal Vero cell culture vaccinia virus (New York City Board of Health strain)--a second-generation smallpox vaccine for biological defense. Int J Infect Dis. 2004;8 Suppl 2:S31-S44. doi:10.1016/j.ijid.2004.09.002

14. Xia S, Zhang Y, Wang Y, et al. Safety and immunogenicity of an inactivated SARS-CoV-2 vaccine, BBIBP-CorV: a randomised, double-blind, placebo-controlled, phase 1/2 trial. Lancet Infect Dis. 2021;21(1):39-51. doi:10.1016/S1473-3099(20)30831-8

15. Lee JS, Grav LM, Lewis NE, Kildegaard HF. CRISPR/Cas9 Mediated Genome Engineering of CHO Cell Factories: Application and Perspectives. Biotechnol J 10 (7): 979-94.

16. Lee JS, Kallehauge TB, Pedersen LE, Kildegaard HF. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. Sci Rep. 2015;5:8572.

17. Arifin MA, Mel M, Abdul Karim MI, Ideris A. Production of Newcastle disease virus by Vero cells grown on cytodex 1 microcarriers in a 2-litre stirred tank bioreactor. *J Biomed Biotechnol*. 2010;2010:586363. doi:10.1155/2010/586363

18. Knop DR, Harrell H. Bioreactor production of recombinant herpes simplex virus vectors. *Biotechnol Prog*. 2007;23(3):715-721. doi:10.1021/bp060373p

19. Genzel Y. Designing cell lines for viral vaccine production: Where do we stand?. *Biotechnol J*. 2015;10(5):728-740. doi:10.1002/biot.201400388

20. Litwin J. The growth of Vero cells in suspension as cell-aggregates in serum-free media. *Cytotechnology*. 1992;10(2):169-174. doi:10.1007/BF00570893

21. Dee KU, Shuler ML. Optimization of an assay for baculovirus titer and design of regimens for the synchronous infection of insect cells. *Biotechnol Prog*. 1997;13(1):14-24. doi:10.1021/bp960086t

22. Paillet C, Forno G, Kratje R, Etcheverrigaray M. Suspension-Vero cell cultures as a platform for viral vaccine production. *Vaccine*. 2009;27(46):6464-6467. doi:10.1016/j.vaccine.2009.06.020

23. Rourou S, Ben Ayed Y, Trabelsi K, Majoul S, Kallel H. An animal component free medium that promotes the growth of various animal cell lines for the production of viral vaccines. *Vaccine*. 2014;32:2767–2769.

24. Shen CF, Guilbault C, Li X, et al. Development of suspension adapted Vero cell culture process technology for production of viral vaccines. *Vaccine*. 2019;37:6996–7002.

25. Jaluria P, Betenbaugh M, Konstantopoulos K, Frank B, Shiloach J. Application of microarrays to identify and characterize genes involved in attachment dependence in HeLa cells. *Metab Eng*. 2007;9(3):241-251. doi:10.1016/j.ymben.2006.12.001

26. Mehrbod, P., Fotouhi, F., G, I.M., Mohammadpour, A., Farahmand, B., 2015. Comparison of Vero and MDCK cell lines transfected with human siat7e gene for conversion to suspension culture. Vaccine Res. 2, 69–73. 10.18869/acadpub.vacres.2.4.69.

27. Van der Sanden SM, Wu W, Dybdahl-Sissoko N, et al. Engineering Enhanced Vaccine Cell Lines To Eradicate Vaccine-Preventable Diseases: the Polio End Game. J Virol. 2016;90(4):1694-704.

28. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9,** 1911 (2018).

29. Wu, W., Orr-Burks, N., Karpilow, J. *et al.* Development of improved vaccine cell lines against rotavirus. *Sci Data* **4,** 170021 (2017). https://doi.org/10.1038/sdata.2017.21

30. Hoeksema F, Karpilow J, Luitjens A, et al. Enhancing viral vaccine production using engineered knockout vero cell lines - A second look. Vaccine. 2018;36(16):2093-2103. doi:10.1016/j.vaccine.2018.03.010

31. Osada N, Kohara A, Yamaji T, et al. The genome landscape of the african green monkey kidney-derived vero cell line. *DNA Res*. 2014;21(6):673-683. doi:10.1093/dnares/dsu029

32. Davies SL, Lovelady CS, Grainger RK, Racher AJ, Young RJ, James DC. Functional heterogeneity and heritability in CHO cell populations. Biotechnol Bioeng 2013;110:260–74.

33.

34. Diaz, M.O., Ziemin, S., Le Beau, M.M., et al. 1988,Homozygous deletion of the a- and b interferon genes in human leukemia and derived cell lines, *Proc.Natl. Acad. Sci*. USA, 85, 5259–63.

35. Sakuma, C., Sekizuka, T., Kuroda, M. *et al.* Novel endogenous simian retroviral integrations in Vero cells: implications for quality control of a human vaccine cell substrate. *Sci Rep* **8,** 644 (2018).

36. Sène MA, Kiesslich S, Djambazian H, Ragoussis J, Xia Y, Kamen AA. Haplotype-resolved de novo assembly of the Vero cell line genome. *NPJ Vaccines*. 2021 Aug 20;6(1):106. doi: 10.1038/s41541-021-00358-9. PMID: 34417462.

AUTHOR CONTRIBUTION STATEMENT

MAS prepared the manuscript. YX, and AAK provided guidance, supervision, and critical reading of the manuscript.

Preface to Chapter 3

Vero cells are derived from a female African green monkey kidney, and they are a widely used cell line for vaccine manufacturing. With over 40 years of experience, they are frequently used for notably, newly discovered virus isolation and reproduction, but also, they are susceptible to various range of viruses, and they can be infected at high multiplicity of infection without triggering the cell death pathway. This cell line is also used for the manufacturing of various vaccines including vaccines against Covid19. But despite this wide use, there is still an issue with Vero cells which is their low viral yield. But aside from that, the cell line is still considered as an attractive candidate for virus-based therapeutics. Indeed, there is one key question that drives research in the vaccine manufacturing field for decades now: how to achieve a cost effective and time efficient approach to produce high number of viruses per dose or viral particles per dose.

To answer that question, two main directions were taken, which are the adaptation to suspension by having the cell multiply and grow suspended. Vero cells are normally adherent cells, they grow attached to the cell culture dish. Usually, to increase the production, it is necessary to scale up. Thus, efforts have been made to adapt Vero cells to suspension to achieve higher sale density while having a less labor intensive manufacturing setting with a smaller footprint.

But there are still some limitations after attempts over the years which are the high doubling time that is still observed, the relatively low cell density and the formation of aggregates.
The other approach taken is increasing the cell specific viral production, meaning the number of

viruses produced per cell by knocking out some antiviral genes. But so far there was no substantial increase in the virus yield. The issues raised being notably the use of gene targets from screen from the human genome and the design guide RNAs for CRISPR gene editing using the African green genome.

Therefore, significant limitations are observed with both approaches and some unanswered questions remain. On one side, why do we observe this kind of limitation when adapting Vero cells to suspension? On the other, why, after doing CRISPR-based knockout of antiviral genes, there was no significant increase in virus yield. To answer those questions, it is necessary to connect the genomics and the phenotype observed by understanding the phenotype using functional genomics. Indeed, by studying the different aspects of genomics from the genome to transcriptome to determine how those individual components work together to produce a particular phenotype. In our case, to determine the reason behind the low cell density, high doubling time and cell aggregates when adapting Vero cells to suspension, and the reason behind the low viral yield after CRISPR editing.

Prior to answering those questions, one major element is still needed which is building a reference grade genome for Vero cells.

Chapter 3

Haplotype-Resolved de novo Assembly of the Vero Cell Line Genome

Authors: **Marie-Angélique Sène[1],** Sascha Kiesslich[1], Haig Djambazian[2], Ioannis Ragoussis[2], Yu Xia[1], Amine A. Kamen[1,*]

Author affiliation : [1] Department of Bioengineering, McGill University, Montreal, QC, Canada
[2] McGill Genome Center, Montréal, QC, Canada
* Corresponding author

Journal: NPJ Vaccines
Year: 2021
DOI: https://doi.org/10.1038/s41541-021-00358-9
Status: Published

ABSTRACT

The Vero cell line is the most used continuous cell line for viral vaccine manufacturing with more than 40 years of accumulated experience in the vaccine industry. Additionally, the Vero cell line has shown high affinity for infection by MERS-CoV, SARS-CoV and recently SARS-CoV-2, emerging as an important discovery and screening tool to support the global research and development efforts in this COVID-19 pandemic. However, the lack of a reference genome for the Vero cell line has limited our understanding of host-virus interactions underlying such affinity of the Vero cell towards key emerging pathogens, and more importantly our ability to re-design high-yield vaccine production processes using Vero genome editing. In this paper, we present an annotated highly contiguous 2.9 Gb assembly of the Vero cell genome. In addition, several viral genome insertions, including Adeno-associated virus serotypes 3, 4, 7 and 8, have been identified, giving valuable insights into quality control considerations for cell-based vaccine production systems. Variant calling revealed that, in addition to interferon, chemokines and caspases related genes lost their functions. Surprisingly, the ACE2 gene, which was previously identified as the host cell entry receptor for SARS-CoV and SARS-CoV-2, also lost function in the Vero genome due to structural variations.

1.0 INTRODUCTION

Originated from a female *Chlorocebus sabaeus* (African Green Monkey) kidney, the Vero cell line represents the most widely used continuous cell line for the production of viral vaccines with over 40 years of experience[1]. This includes the development and production of vaccines against dengue fever, influenza, Japanese encephalitis, polio, rabies, rotavirus, smallpox and more recently, Ebola (using a recombinant Vesicular Stomatitis Virus)[2-4].

The advances in gene editing have made it possible to edit the genome of cell lines with high-throughput and cost-effective methods using available genomic data, thus providing new possibilities for cell line development and vaccine bioprocessing intensification. Some attempts to develop engineered Vero cell line have been made[5] but genome editing of this cell line still suffers from the lack of annotated reference-grade genomic information. Despite the publication of the Vero genomic landscape[6], there are no tools available yet to efficiently select CRISPR/Cas9 target sites with sufficient accuracy such as a CRISPR screening library for Vero cells.

Furthermore, Vero cells have been identified as the cell line with the highest susceptibility to MERS-CoV[7], SARS-CoV and recently SARS-CoV-2[8]. Consequently, Vero cells have been extensively used in the current response to COVID-19 as a platform for SARS-CoV-2 isolation and replication, viral vaccine production and identification of potential drug targets[9]. Currently, several COVID-19 attenuated or inactivated vaccine candidates in pre-clinical and clinical trials use Vero cells as a production platform. We thus propose a haplotype resolved annotated assembly of the WHO-Vero genome which will provide a valuable resource for quality control, enable the generation of high-throughput engineered sub-cell lines and accelerate the

development of vaccine manufacturing platforms contributing to the global preparedness plan to counteract emerging and reemerging infectious diseases.

2.0 RESULTS

2.1 De novo assembly of the Vero genome and annotation

Using sequencing reads with a mean coverage per base pair of 100.2 (Figure 1) of the African Green Monkey genome[10] , we present here a principal pseudohaplotype and an alternate pseudohaplotype of the Vero genome consisting respectively of 6872 and 6876 scaffolds, with a total length of 2.9Gb, a L50 count of 12 and NG50 length of 82Mb and 70Mb (Figure 2) with 39449 predicted genes (29824 genes were predicted for the African Green Monkey genome using the same default parameters), 35004 genes and pseudo genes annotated (including 21620 protein-coding genes).

The completeness of the genome assembly was assessed by BUSCO[11] and CEGMA[12] via the gVolante portal[13]. 94.85% complete, 2.15% partial genes and 94.85% complete, 2.57% partial genes were identified in the principal pseudohaplotype and the alternate pseudohaplotype respectively. CEGMA[12] revealed that, 98.71% of the 233 core vertebrate genes were evolutionarily conserved genes identified in both pseudohaplotypes of the Vero genome.

Comparing both pseudohaplotypes output before downstream processing, we find that from the initial 55,755 scaffolds 723 scaffolds have differing sequences. These 723 scaffolds account for 2,512,305,804 bases from a total of 2,848,013,978 assembled bases. In these differences there are 5,027,642 mismatch (SNP) bases with 4,830,436 of 1bp, 94,229 of 2-5bp and 217 of 6-25bp. Comparing the pseudohaplotypes we also count indel of different sizes: 436,916 indels of 1bp, 327,037 indels of 2-5bp and 103,216 indels of 6-25bp.

Our assembly sequence quality was further confirmed by the 96.5% alignment of randomly generated illumine short reads downloaded from the SRA database. Furthermore, using those short reads, the QV (quality value) was calculated using Merqury[14] pipelined with Meryl[15] and pseudohaplotype 1 and 2 reached, respectively, a QV of 28.7531 and 28.3972 which correspond to an accuracy of 99.87% and 99.86%. Furthermore, a new 30X sequencing round was done and short reads were generated and aligned to the Vero genome for SNP call. Considering calls with AF of 1, an error rate of 0.015523% or one error every 6441 bases was observed.

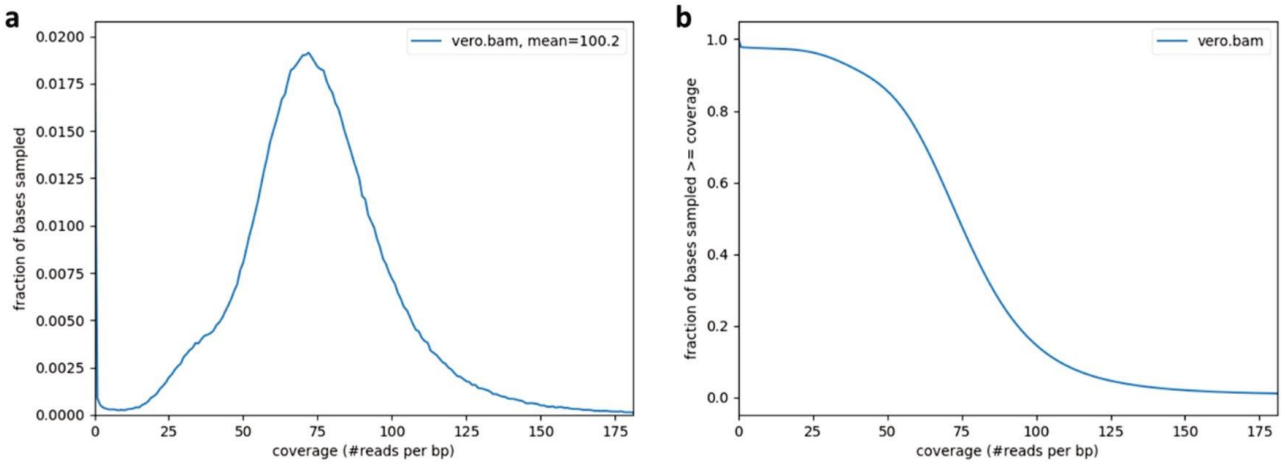


Figure 1 Vero genome sequencing depth. a Distribution of read coverages. b Genome fractions depth of sequencing.

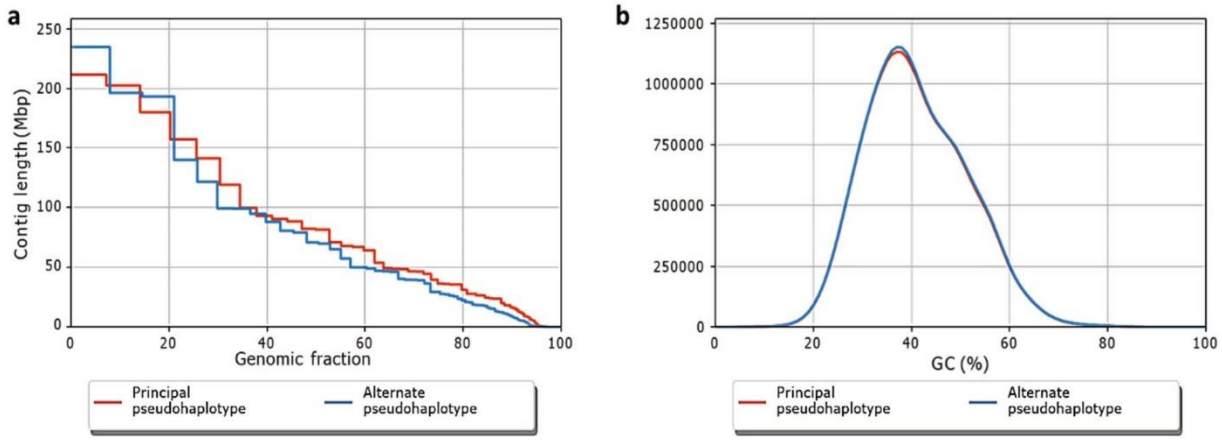


Figure 2 Quality control of the Vero genome de novo assembly. a Nx plot: Length ordered contigs. b Distribution of GC content in the contigs.

Additionally, Vero RNAseq data was deposited under the Vero Bioproject alongside the two Vero genome pseudohaplotype assemblies in order to be used with vervet RNAseq data as evidence for annotation by NCBI. Comparing the resulting annotation (Annotation release 102) with the previous African Green Monkey annotation(Annotation release 100), only 1% of the genes in the Vero annotation are identical (i.e Genes with perfect match in exon boundaries) to those of the African green monkey annotation, 46% of the genes had minor changes (i.e. Highly similar genes, with support scores of 0.66 or more (on a scale of 0 to 1) on both sides of the comparison, the support score is derived from a combination of matching exon boundaries and sequence overlap), 23% of the genes have major changes (i.e. Genes with support scores lower than 0.66 (on a scale of 0 to 1) on one or both sides of the comparison, and genes with changed locus, biotype or changed completeness, and split or moved genes), 30 % of the genes are new (i.e. Novel genes or genes without a match in the African Green Monkey annotation). In addition, 68 viral proteins (36 viral genes) were also annotated.

2.2 Detection of genomic rearrangements in the Vero cell line

Using the African Green Monkey genome[10] as a reference, the Vero cell line sequenced reads covered 91.3% of the African Green Monkey genome while ~12 million small indels and SNPs were called with SNVSniffer[16] (Figure 3a, 3b) and 7354 large-scale structural variants (including interchromosomal translocations) were called using Manta[17] (Figure 3c, 3d, Figure 4).

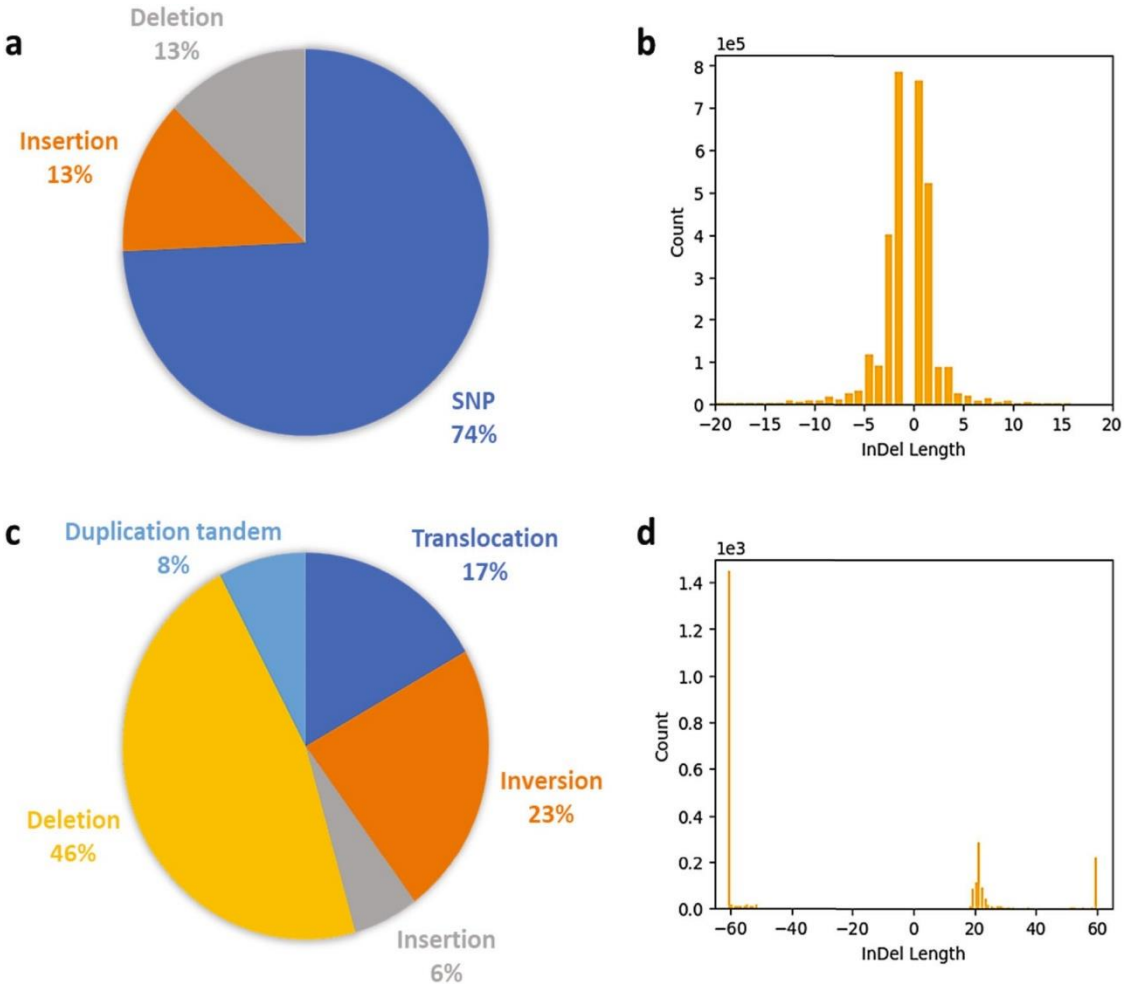


Figure 3 Characterization of called structural variants. a Different SNVSniffer called small-variant types distributions b SNVSniffer called small- scale indels distribution. c Different Manta called small-variant types distributions. d Manta called large-scale indels distribution.

Following the annotation of those variants, among 7585 genes predicted as having lost their function, a total of 551 genes were identified as having lost their function as a result of feature ablation(notably due to chromosome copy number variations), while 12 genes were identified as having lost their function due to transcript ablation (Table 2,Supplementary Datas1-3).

Variant type-based distribution of genes with a predicted loss of function.

Variant caller	Variant type	Number of genes with predicted loss of function
Manta	Feature ablation	551
	Exon loss	38
	Frameshift	27
	Transcript ablation	12
	Gene fusion	3
SNVSniffer	Frameshift	8894
	Stop gained	592
	Splice donor	390
	Start lost	130

Table 2Variant type-based distribution of genes with a predicted loss of function

Besides these ablation variants, given the nature of the Loss of Function predictions, additional analysis such as proteomics might be needed in order to provide a detailed insight on the effect of those variants-caused gene loss of function on gene products functionality for instance in cases such as ACE2 or IVNS1AB predicted loss of function. After filtering, overall, 7585 genes with predicted loss of function are involved in cellular organization, including pro tumorigenic genes, as well as cytopathic pathways, immune response mechanisms, response to viral infection and protein processing (Figure 5). In addition,33 proviral genes were identified as having lost their function including endogenous retrovirus group members ERVV-2 and ERVMER34-1, Bcl2/-adenovirus receptors, influenza virus NS1A binding protein (IVNS1ABP) and angiotensin I converting enzyme 2 (ACE2) involved in SARS-CoV-2 cell entry mechanism[8] (Table 3).

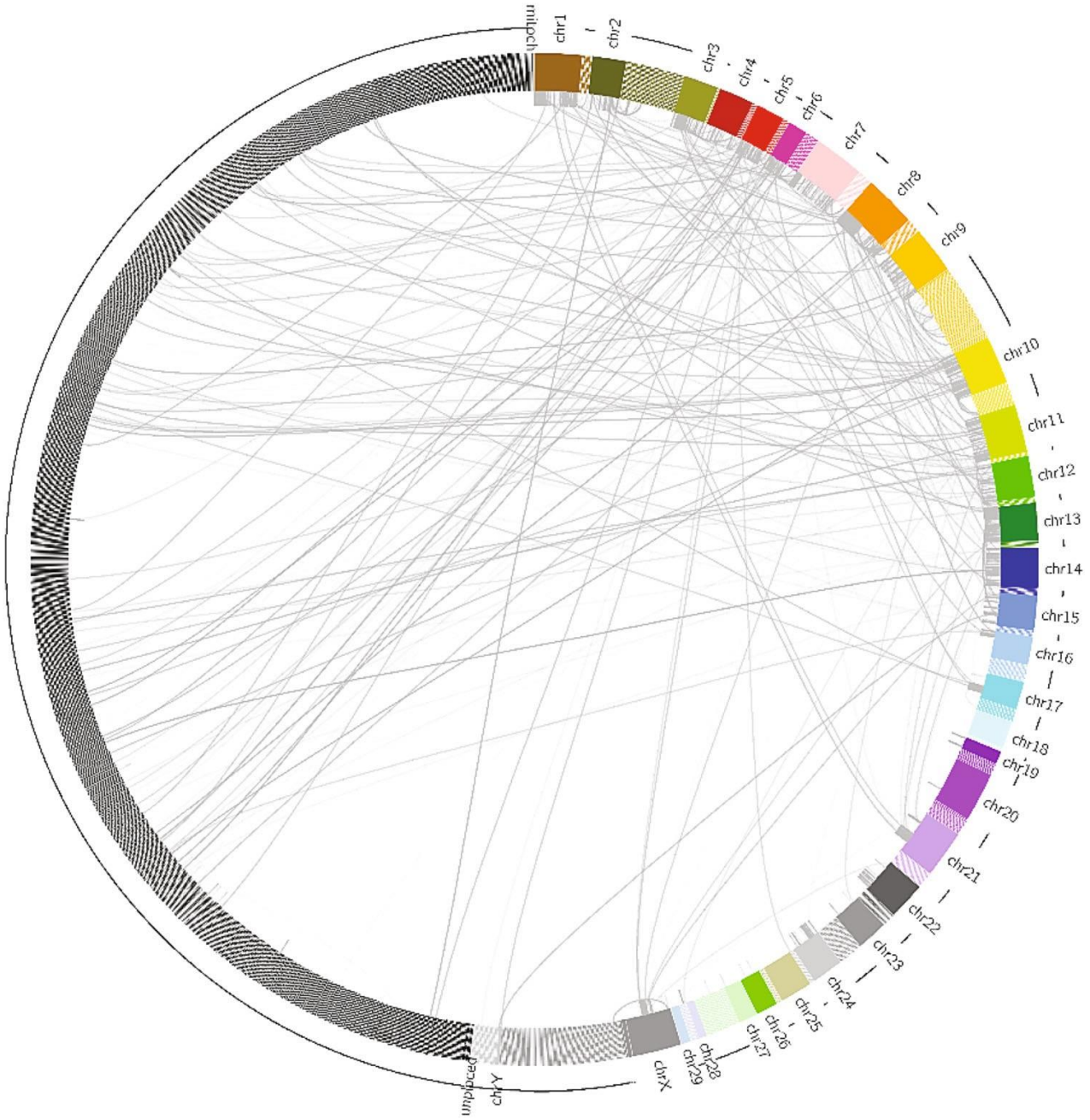


Figure 4 Circos plot of Manta called large-scale variants. The outer circle represents the African Green Monkey genome including its unplaced scaffolds (left dark gray) to highlight the interchromosomal translocations in the Vero genome relative to the African Green Monkey genome.

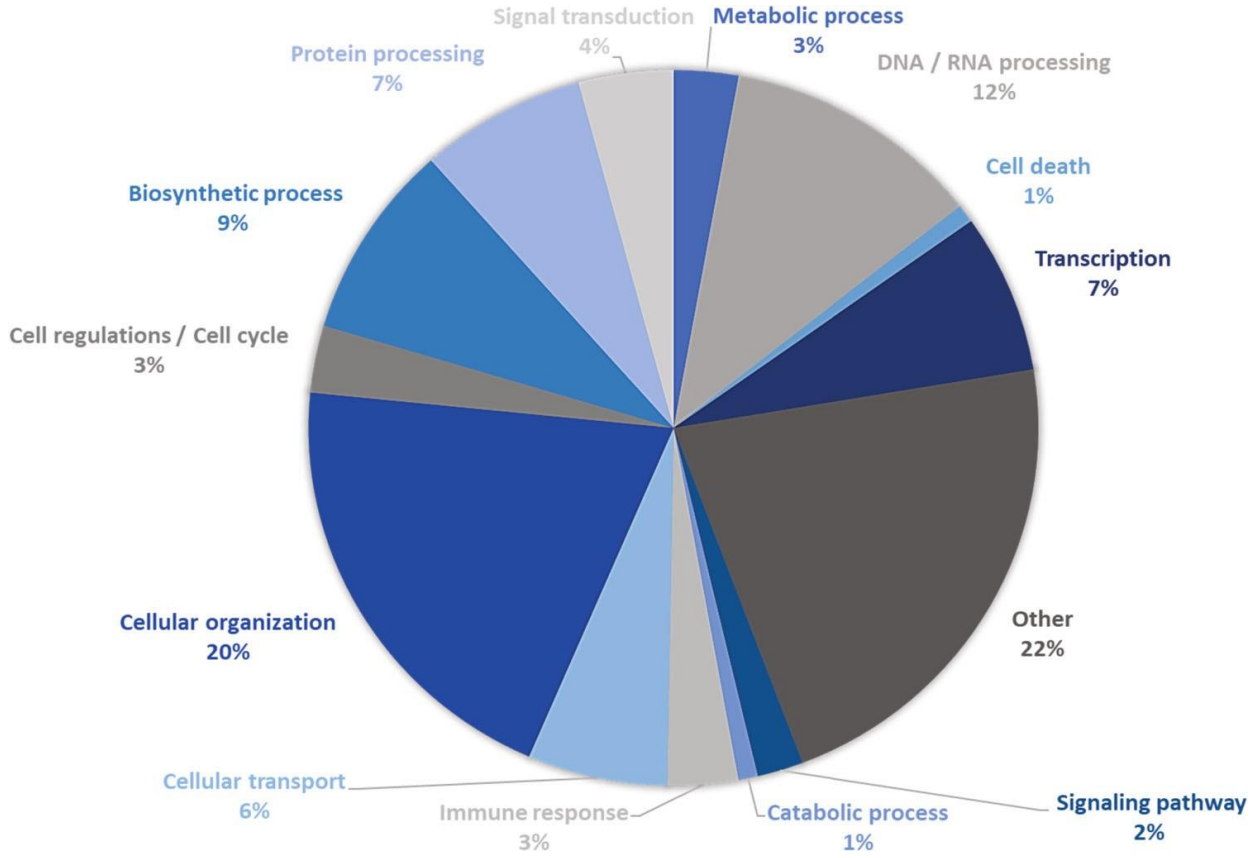


Figure 5 Function lost genes clustering. Biological clustering of genes identified as having lost their function based on their functional annotation.

Table 3. Proviral genes which lost their function due to structural variants.

NCBI ID	Gene name	LOF related variant type
103235187	endogenous retrovirus group V member 2(ERVV-2)	Deletion: Frameshift variant
103231639	angiotensin I convening enzyme 2(ACE2)	Insertion: Splice acceptor & intron variant (multiple stop codons)Deletion: Frameshift variant
103227905	endogenous Bomavirus-like nucleoprotein 2(EBLN2)	Insertion: Frameshift variant (multiple stop codons)
103229363	feline leukemia virus subgroup C cellular receptor family member2(FLVCR2)	Insertion: Frameshift variant
103225011	human immunodeficiency virus type I enhancer-binding protein3(HIVEP3)	Deletion: Frameshift variant
103230431	influenza virus NSIA binding protein(IVNSIABP)	Insertion: Frameshift variant
103240251	murine retrovirus integration site 1 homolog(MRVI1)	Deletion: Feature ablation (multiple stop codons)
103235692	endogenous retrovirus group MER34 member HERVMER34–1)	Deletion: Frameshift variant (multiple stop codons)
103237620	solute carrier family 52 member 2(SLC52A2)	Insertion: Frameshift variant
103215978	solute carrier family 52 member 3(SLC52A3)	Deletion: Frameshift variant
103227211	zinc finger and SCAN domain containing 10(ZSCANIO)	Insertion: Frameshift variant
103222367	zinc finger and SCAN domain containing 12(ZSCANI2)	Deletion: Frameshift variant
103221929	zinc finger and SCAN domain containing 16(ZSCANI6)	Deletion: Frameshift variant
103246758	zinc finger and SCAN domain containing 21(ZSCAN21)	Insertion: Frameshift variant
103221914	zinc finger and SCAN domain containing 23(ZSCAN23)	Insertion: Frameshift variant
103245744	zinc finger and SCAN domain containing 29(ZSCAN29)	Insertion: Frameshift variant
103227463	zinc finger and SCAN domain containing 32(ZSCAN32)	Insertion: Frameshift variant
103235415	zinc finger and SCAN domain containing 4(ZSCAN4)	Deletion: Frameshift variant
103227450	zinc finger protein 174(ZNF174)	Insertion: Frameshift variant
103227495	zinc finger protein 197(ZNF197)	Insertion: Frameshift variant (multiple stop codons)
103248713	zinc finger protein 202(ZNF202)	Insertion: Frameshift variant
103241144	zinc fmger protein 215(ZNF215)	Deletion: Frameshift variant (no start codon)
103242227	zinc finger protein 232(ZNF232)	Deletion: Frameshift variant
103222539	zinc finger protein 24(ZNF24)	Insertion: Frameshift variant
103227313	zinc fmger protein 263(ZNF263)	Deletion: Frameshift variant
103222546	zinc finger protein 397(ZNF397)	Deletion: Frameshift variant
103235347	zinc finger protein 444(ZNF444)	Deletion: Frameshift variant
103227490	zinc finger protein 445(ZNF445)	Deletion: Frameshift variant
103235451	zinc fmger protein 446(ZNF446)	SNP: Splice acceptor & intron variant
103221931	zinc finger with KRAB and SCAN domains 8(ZKSCANS)	Deletion: Frameshift variant
103233722	ATCAY, caytaxin(ATCAY)	Insertion: Frameshift variant
103247216	BCL2 interacting protein-like(BNIPL)	Insertion: Frameshift variant
103233646	adapter related protein complex 3 delta 1 subunit(AP3D1)	

LOF loss of function.

Table 3 Proviral genes which lost their function due to structural variants.

2.3 Identification of viral sequences

Following a BLASTN[18] search on the custom-made viral sequences database, several viral genomic sequences were identified for both Vero genome pseudohaplotypes and the African Green Monkey genome (Table 4) with an E-value cutoff of 10^{-50} to account for only highly similar sequences. These sequences include as expected retroviral sequences such as the simian retroviral[19] and human endogenous retroviruses. In addition, complete viral genomes of the

Adeno-associated virus serotypes 3, 4, 7, 8, sarcomas, blastomas and leukemia viruses were identified, among others.

Table 4. Viral genomic sequences inserted in the Vero cell line genome.

RefSeq release number	Viral sequence	E-value
NC_002665.1	Bovine herpesvirus 4 long unique region, complete sequence	0.0
NC_009889.1	RD114 retrovirus, complete genome	0.0
NC_022518.1	Human endogenous retrovirus K113 complete genome	0.0
NC_031326.1	Simian retrovirus 8 strain SRV8/SUZ/2012, complete genome	0.0
NC_014474.1	Simian retrovirus 4 strain SRV4/TEX/2009/V1, complete genome	0.0
NC_022517.1	Baboon endogenous virus strain M7 proviral DNA, complete genome	0.0
NC_001550.1	Mason-Pfizer monkey virus, complete genome	0.0
NC_001829.1	Adeno-associated virus - 4, complete genome	0.0
NC_001729.1	Adeno-associated virus - 3, complete genome	0.0
NC_006260.1	Adeno-associated virus - 7, complete genome	1E-175
NC_006261.1	Adeno-associated virus - 8, complete genome	6E-164
NC_001499.1	Abelson murine leukemia virus, complete genome	5E-163
NC_001350.1	Saimiriine herpesvirus 2 complete genome	4E-132
NC_032111.1	BeAn 58058 virus, complete genome	2E-114
NC_038922.1	Avian sarcoma virus CT10 genomic sequence	8.00E-97
NC_038858.1	FBR murine osteosarcoma, complete proviral sequence	2.00E-83
NC_001506.1	Murine osteosarcoma virus, complete genome	2.00E-83
NC_009424.5	Woolly monkey sarcoma virus	2.00E-81
NC_041925.1	Proteus phage VB_PmiS-Isfahan, complete genome	5.00E-79
NC_003678.1	Pestivirus giraffe-1 H138 complete genome	1.00E-66
NC_008094.1	Y73 sarcoma virus, complete genome	2.00E-66
NC_001461.1	Bovine viral diarrhea virus 1, complete genome	5.00E-65
NC_043404.1	Avian myeloblastosis virus RNA-dependent DNA polymerase gene, partial cds; transforming protein gene, complete cds; and long terminal repeat, complete sequence.	2.00E-63
NC_043382.1	Snyder-Theilen feline sarcoma virus genomic sequence	1.00E-59
NC_038923.1	Hardy-Zuckermann 4 feline sarcoma virus (H24-FeSV) kit oncogene	3.00E-57
NC_038668.1	Harvey murine sarcoma virus p21 v-has protein gene	5.00E-54
NC_001885.3	Gibbon ape leukemia virus gag, pol, and env genes, complete cds	1.00E-50
NC_007815.2	PreXMRV-1 provirus, complete genome	5.00E-50

Table 4 Viral genomic sequences inserted in the Vero cell line genome.

2.4 ACE2 preliminary analyses

A comparison of Vero ACE2(vACE2) and human ACE2(hACE2) protein sequences showed 43 residues mutations (Table 5), 94.71% identity and respectively, a molecular weight of 92427 Da and 92463 Da for vACE2 and hACE2. As shown in Table 5, the mutations were highlighted on the protein 3D structure. A preliminary experiment to assess the loss of function prediction for ACE2 in Vero cells at the final gene product level was conducted. Indeed, ACE2 activity was assessed via the activity assay and vACE2 cells showed no activity for three different cell culture samples in triplicates (Figure 6).

Table 5. Residues mutations between vACE2(alteration) and hACE2(original) and mutations positions on vACE2 3D structure(red).

Position	Original	Alteration	Position on vCAE2 3D structure
3	S	G	
67	D	E	
87	E	A	
136	D	N	
145	E	D	
153	A	E	
154	N	K	
167	S	G	
197	E	K	
218	S	N	
220	G	D	
228	H	R	
259	I	T	
299	D	N	
303	D	N	
342	A	V	
359	L	I	
555	F	L	
559	R	K	
630	D	A	
631	R	N	
634	E	K	
657	K	E	
658	V	N	
660	N	H	
662	M	T	
674	N	D	
684	F	Y	
702	K	E	
706	M	F	
716	R	Q	
729	P	S	
732	G	V	
735	N	Y	
737	P	S	
739	V	I	
740	S	T	
741	I	T	
751	G	A	
755	V	A	
759	I	V	
773	K	Q	
777	G	E	

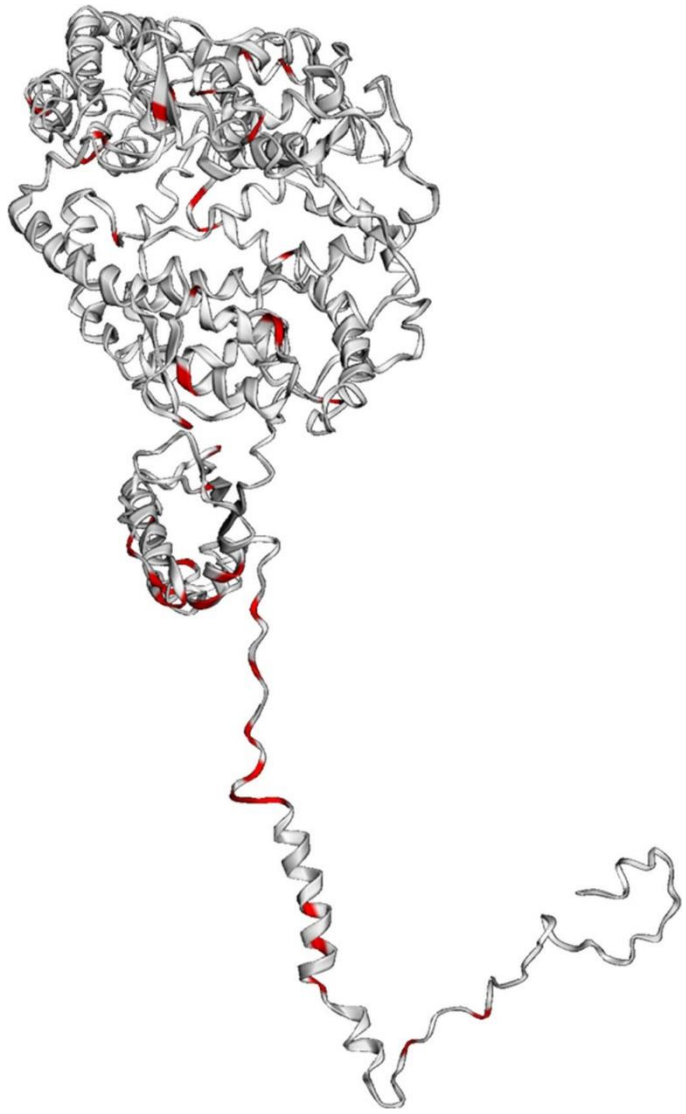


Table 5 Residues mutations between vACE2(alteration) and hACE2(original) and mutations positions on vACE2 3D structure(red).

	Replicate 1	Replicate 2	Replicate 3
Background Control			
Negative Control			
Positive Control			
Vero ACE2 Sample 1			
Vero ACE2 Sample 2			
Vero ACE2 Sample 3			

Figure 6 ACE2 enzymatic activity assessment across Vero ACE2 samples. Vero ACE2 activity assay Reading Matrix (Fluorometric).

3.0 DISCUSSION

In this work, we present a de novo assembly and annotation of two pseudohaplotypes for the Vero cell line, providing a genomic tool for a better understanding of the Vero cell line and its interactions with viruses but also for the design of more efficient cell engineering strategies such as CRISPR[20] screenings and gene editing(by proposing an annotated Vero genome that can be used as a baseline for screening approaches such as more rigorous gene expression profiling, pathway enrichment analysis and even the generation of a CRISPR screening library for Vero cells). Given the identification of Vero cells as an aneuploid cell line, it is necessary to fully characterize the heterogeneity of the Vero cell population using tools such as single cell analysis. Indeed, bulk genome analysis of cell populations tend to conceal differences between changes in expression from changes in the cellular composition of the population. Thus, single cell analysis has emerged as a tool to have a detailed understanding of genomic, epigenomic and transcriptional variations at the single cell level in order to detects cellular heterogeneity and more precisely identify major subpopulations emerging from that heterogeneity. Due to the lack

of a fully annotated reference genome to assist in single cell analysis, we took a first preliminary step proposing this haplotype-resolved draft assembly genome giving a first glance into the heterogeneity of this continuous cell line by comparing the two haplotypes. Comparing this assembly with the African Green Monkey[10] genome helped unravel genomic events explaining Vero cell characteristics as a cell culture platform. It is necessary to note that despite the stringent statistical parameters used for variant calling, errors due to sequencing technologies can still affect variant calls, thus loss of function predictions. Therefore, it is necessary for end users of this draft assembly to validate variants related to their genes of interest before downstream applications.

Notably, the interchromosomal translocations previously discovered via karyotyping[10], were identified alongside other large-scale and small-scale structural variants through variant calling (Figure 4, Supplementary Data 4) and can be used to fully characterize the interchromosomal events in Vero cells. Furthermore, the effects of those genomic rearrangements on gene products functionality further explained the continuous nature of the cell line and its relative high susceptibility to infection[8]. Indeed, several genes involved in the cytopathic pathway, cell regulations, immune response and pro tumorigenic genes lost their function due to frameshift variants, features ablation, splice acceptor variant and intron variants, among others.

In light of the extensive rearrangements observed both in our analysis and in the 2014 Landscape of Vero cells paper[10], we don't think that the African Green Monkey genome should be used as a reference to search for missassemblies and correct a genome assembly of the Vero cells because, if the African Green Monkey is used as a reference for that purpose, the extensive rearrangements will appear in the missassembly report and correcting those to comply with the reference, as per traditional reference-based genome assemblies, will not generate an assembly

that accurately describe the Vero cells genome. As expected, given the significant amount of rearrangements, when running QUAST[28] with the African Green Monkey as reference, NGA50 is 1.2Mb, which could be due to either significant genomic rearrangements or chimeric contigs present in either the African Green Monkey assembly or our Vero assembly. But, alignment of short reads to our assembly detected no chimeric read and given the extensive rearrangements observed in the 2014 Landscape paper[10] and in our analysis (which are detailed in the supplementary Data and visually shown in the figure 4) are the main cause of such NGA50 value. Nonetheless, the African Green Monkey assembly, is a valuable tool to identify rearrangements in the Vero cell genome compared to the African Green Monkey genome from which it is derived in order to shed some light on the peculiar characteristics of this cell line that makes it a valuable candidate for virus production and vaccine manufacturing.

In addition to the Simian retroviral sequence insertions previously identified[19], the analysis of this Vero cell genome further showed the insertion of several complete viral genomes including the human endogenous retroviral sequence, proviral sequences, sarcoma viruses and adeno-associated viruses which might provide insights on developing efficient downstream processing steps and quality control tools for manufacturing biologics. Furthermore, the availability of an annotated Vero cell line genome provides new possibilities for viral sequence clearance through gene editing. Also, the proviral genes identified as having lost their function are involved in key stages of retroviruses, leukemia viruses, influenza virus and adenoviruses reproduction cycle despite Vero cells wide use as a susceptible to highly susceptible cell line. Given the success of the Vero cell line as a virus production platform, the predicted loss of function of proviral genes might appear as counter intuitive but these proviral genes being specific to their associated type of virus, on one hand the loss of function might affect that specific virus production rate but

other virus production rates might not be affected. For example, in the case of influenza virus strains, the Vero cell line was one of the first cell lines considered as a vaccine production platform for influenza vaccines but its relatively low viral particles production yield limited the use of Vero cell line as a manufacturing platform. Our analysis of the Vero genome demonstrated that Influenza virus NS1A binding protein (IVNS1ABP) also lost its function, hindering the production of M2 protein and the overall Influenza virus production yield. Thus, with an annotated genome now available, new strategies can be designed to re-engineer high-yield Vero cell line for influenza vaccine production. On the other hand, the predicted loss of function might hint towards an alternative cell-virus mode of interaction used by Vero cells to bypass this predicted loss of function and still produce the virus at high rates. This might be the case for ACE2 where despite the ACE2 gene (receptor for SARS-CoV and SARS-CoV-2 host cell entry[8]) predicted loss of function, Vero cell line has a high susceptibility to SARS-CoV and SARS-CoV-2 infection and is consequently used for the production of inactivated and attenuated COVID-19 vaccines. It is important to note that the loss of function for ACE2 in Vero cells remains a prediction rather than a fact. While the ACE2 protein is predicted to at least lose its catalytic function in Vero cells which is corroborated by preliminary experimental results(Tables 3,5), further experiments are necessary to determine whether or not the receptor binding function of the ACE2 protein is maintained or not in Vero cells, via, for instance, detailed mass-spectroscopy analyses of viral binding sites, which are beyond the scope of this paper. Overall, these proviral genes predicted loss of functions need to be investigated on a case-by-case viral infection in order to fully comprehend the Vero cell line as a viral production platform. Vero cell's ability to be infected at high multiplicity of infection (MOI), without instantly triggering the cytopathic pathway, was first explained with the loss of type I interferon[10], which was confirmed with the identification of genes involved in the pathway,

from interferons and caspases loss to chemokines hindrances, losing their functions. In addition, a BLAST[16] search against viral databases revealed the insertion of BeAn 58058 virus complete genome which contains cytokine response-modifying protein B, surface antigen S, chemokine-binding protein, interferon antagonist K1L and serine proteinase inhibitor 1, among other cell death inhibitors, thus strengthening the cytopathic pathway inhibition. Gene profiling of infected Vero cells might provide additional insights on the balance between the effect of the host cell antiviral genes loss of function and the insertion of virus proviral gene sequences into the host cell genome.

Furthermore, the Adeno-associated complete viral sequences insertion into the Vero genome might provide new alternatives for the study of Vero cell-based Adeno-associated virus and eventually design of alternative production platforms for AAV serotypes.

In addition, gene editing tools have been used to create Vero suspension cell lines, which are considered to facilitate bioprocess development efforts. Nevertheless, singular genome modifications did not seem to lead to successful and sustainable results[1]. Using the annotated genome presented here, further studies can be conducted to investigate suspension adaptation efforts, for example large-scale screens could identify if a combination of modifications is needed. Additionally, already adapted suspension Vero cell lines still exhibit low cell growth rates with doubling times of more than 40 h and the frequent formation of aggregates[21]. Genome-wide screens or comparative transcriptomic studies could further investigate factors that would lead to improved Vero suspension systems.

In the long run, the findings of this study and previous ones[6,19] are sought to pave the way for widespread application of genome analysis, screening and editing tools for the Vero cell line. Taking into account the use of Vero cells in vaccine manufacturing processes and in particular

the acceptance of this cell line by regulatory authorities, successful applications of genome editing can significantly improve virus production and ultimately lower the cost of vaccine manufacturing.

4.0 METHODS

4.1 Cell lines and culture media

The Vero WHO cell line studied in this work was at passage 138(Neovacs). This cell line was itself derived from a vial of Vero ATCC CCL-81 which was send to WHO at passage 124 for analysis and establishment of the Vero WHO master cell bank approved for vaccine production. The cells were grown in static culture at 37 °C and 5 % CO2 in a humidified incubator (Infors HT, Switzerland). Cells were passaged twice weekly using TrypLE Express (Thermo Fisher Scientific) as dissociation reagent. A serum-free adapted sub-cell line grown in OptiPRO medium (Thermo Fisher Scientific) supplemented with 4 mM GlutaMAX (Thermo Fisher Scientific) was cryopreserved at a passage number of 151 in OptiPRO medium supplemented with 4 mM GlutaMAX and 10 % DMSO (Sigma, USA).

For genome analysis, Vero WHO cells at passage 153 where washed in PBS (Wisent, Canada), harvested using TrypLE Express and centrifuged at 300 × g for 5 minutes. Cell pellets of around 6 million cells were quickly frozen in a mixture of dry ice/ethanol and stored at −80 °C until further analysis.

4.2 De novo genome assembly and annotation

The 10x Genomics linked read libraries were sequenced on three HiSeqX lanes with paired end 151 reads. The reads were first processed with 10x Genomics Long Ranger[22] basic to flag all the reads with a valid molecule barcode. These processed reads were then used to filter the original demultiplexed reads keeping only the reads that carry valid barcodes. The assembly was then run using 10x Genomics Supernova[23] run with the following options: "--bcfrac=1 --maxreads=1490M --localcores=16 --localmem=327". Following the assembly step fasta files with the assembly sequence were generated using Supernova mkoutput with these options: "--style=pseudohap2 --minsize=250 --headers=full" to generate a principal and an alternate pseudohaplotype. 10x Genomics assembler, Supernova, initially outputs the assembly in the form of a graph where edges are assembled sequences. These sequences are linked together at the ends by overlaps of K-1 bases (K=48). To transform the assembly graph into fasta format the graph can be traversed concatenating the sequences along each visited edge. When using the pseudohap Supernova option to extract the fasta version of the assembly, the graph is traversed once based on the highest coverage edges. With the pseudohap2 option the traversal is performed by choosing the second highest coverage edges. This results in two very similar assembly files that differ only where large variations are present.

With the barcode information generated during 10X Linked Read sequencing protocol, ARCS[24], pipelined with LINKS[25] and Tigmint[26], was used to pair the Supernova draft assembly sequences by processing input alignments for sets of read pairs from the same barcode that aligned to different sequences and formed a link between sequence contigs. The iteration parameters m= 50-20000/ e= 90000 and m= 50-10000/ e= 30000 were applied for respectively the principal pseudohaplotype and the alternate pseudohaplotype. Following this scaffolding protocol, Nanopore long reads were used for gap filling via RAILS/Cobbler[27] (iteration parameters: i= 0.7,

0.65/d=250-50/l=1 for the principal pseudohaplotype; i=0.7,0.65/d=250-10/l=1 for the alternate pseudohaplotype) then the scaffolds were polished with ntEdit[27] (k=64,50,40). The assembly quality control metrics were calculated using QUAST[29] and a preliminary gene prediction was done via AUGUSTUS[30].

The principal pseudohaplotype annotation was performed using NCBI's in-house Eukaryotic Annotation Pipeline[31].

The QV value was calculated using Merqury[14] to first assess the optimal k-mer value based on the genome size, followed by Meryl[15] short read database build and QV evaluation.

4.3 Genomic rearrangements detection

Previously generated interleaved reads were mapped to the African Green Monkey annotated genome[10] using the BWA-MEM algorithm[32] and the resulting BAM file was generated via SAMTOOLS[33] view. Deeptools[34] was used to plot the genome coverage. Large-scale indels, duplication tandem and interchromosomal translocations were called using Manta[17] with default parameters. SNVSniffer[16] was used to call the remaining small-scale indels and SNPs (exec_mode parameter = 2 for most accurate variant calling where following variant call, all reads are realigned to calculate per-base alignment quality (BAQ) scores before inputting those alignments to the calling engine[16]). Besides the description of Vero cells as an aneuploid cell line and its major karyotyping presented in the Landscape of Vero cells[6], given the lack of additional information on Vero cells heterogeneity, we decided to use variant caller that are designed for both somatic and germline variant calls(SNVSniffer[16] and Manta[17]) and apply their suggested default parameters for variant calls. In addition, both pipelines score variant candidates relative

to the reference to identify and filter out (if needed) variants due to the Vero cell genome heterogeneity.

The effect of those called variants were predicted using Galaxy's SNPEff[35] to extract all genes that lost their functions. Those resulting genes were functionally annotated via DAVID[36], filtered and clustered in biological groups. Large-scale structural variants called with Manta were plotted via Circos[37]. Variant calls statistics were calculated for both SNVSniffer and Manta using bcftools stats[38].

4.4 Viral genomic sequences detection

To identify and characterize the viral genomic insertions in the Vero genome, all viral sequence releases from RefSeq were used to create a blastn database and a BLAST[18] search was run for both Vero genome pseudohaplotypes. To ensure that no false-positives were included in the results, the African Green Monkey genome[10] was also ran against the created viral database and all resulted viral sequences were identified in both the Vero genome and the African Green Monkey genome.

4.5 ACE2 preliminary analyses

Vero ACE2 protein sequence was obtained from NCBI annotation of our assembly (NCBI Chlorocebus sabaeus (Vero cell) Annotation Release 102 (AR 102)) and a BLAST[18] search was ran on NCBI portal to identify residues mutations. The 3D structure of vACE2 was modeled using Phyre2 server[39], mutations were marked using EzMol[40].

For ACE2 activity assessment, 3 separate cultures of Vero cells were prepared in the same conditions and used with abcam Angiotensin II Converting Enzyme (ACE2) Activity Assay Kit

(Fluorometric) (ab273297) as per the provided protocol all conditions were analysed in triplicates. Fluorometric reading was performed for an hour at 1 minute intervals.

5.0 ACKNOWLEDGEMENTS

This work was supported by funding from the Natural Sciences and Engineering Research Council (NSERC-Grant-CRDPJ 511957). M-A-S is supported by the McGill Engineering Doctoral Award (MEDA), SK is funded by a doctoral scholarship from the Fonds de Recherche du Québec – Santé (FRQS) and A.A.K by a Canada Research Chair (CRC-240394).

6.0 COMPETING INTERESTS STATEMENT

The authors declare that they have no financial or nonfinancial competing interests.

7.0 DATA AVAILABILITY STATEMENT

The assembly and annotation data have been deposited in NCBI under the accession numbers JACDXN000000000 and JACDXO000000000, for the principal and alternate pseudohaplotypes respectively. The NCBI Chlorocebus sabaeus (Vero cell) Annotation Release 102 can be directly accessed through the following page:

https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Chlorocebus_sabaeus/102/

The full comparison table between the Vero annotation 102 and the African Green Monkey annotation 100 can be downloaded at through the following link :

https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/60711/102/GCF_015252025.1_

Vero_WHO_p1.0/Annotation_comparison/GCF_015252025.1_Vero_WHO_p1.0_compare_prev.txt.gz

Supplementary Information: Supplementary Data 1-4 are available for this paper.

Viral genome sequences were downloaded from RefSeq release FTP page.

The African Green Monkey genome was obtained from NCBI (accession number: PRJDB2865).

All other relevant data are available upon request.

8.0 CODE AVAILABILITY STATEMENT

This study did not make use of custom code.

7.0 REFERENCES

1. Sascha Kiesslich, Amine A. Kamen, Vero cell upstream bioprocess development for the production of viral vectors and vaccines, *Biotech Adv.*, Volume 44, 2020, 107608, ISSN 0734-9750.
2. Barrett PN, Mundt W, Kistner O, Howard MK. Vero cell platform in vaccine production: moving towards cell culture-based viral vaccines. *Exp review of vacc*. 2009 May 1;8(5): 607-18.
3. Ammerman NC, Beier-Sexton M, Azad AF. Growth and maintenance of Vero cell lines. *Curr Protoc Microbiol*. 2008;Appendix 4:Appendix 4E.
4. Suder E, Furuyama W, Feldmann H, Marzi A, de Wit E. The vesicular stomatitis virus-based Ebola virus vaccine: From concept to clinical trials. *Hum Vaccin Immunother*. 2018;14(9):2107-2113.
5. Van der Sanden SM, Wu W, Dybdahl-Sissoko N, et al. Engineering Enhanced Vaccine Cell Lines To Eradicate Vaccine-Preventable Diseases: the Polio End Game. *J Virol*. 2016;90(4):1694-704.
6. Osada N, Kohara A, Yamaji T, et al. The genome landscape of the african green monkey kidney-derived vero cell line. *DNA Res*. 2014;21(6):673-83.
7. Liu et al. A recombinant VSV-vectored MERS-CoV vaccine induces neutralizing antibody and T cell responses in rhesus monkeys after single dose immunization. Antiviral Res. 2018 Feb;150:30-38.
8. Hoffmann et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Mar 4. pii: S0092-8674(20)30229-4.
9. Draft landscape of COVID-19 candidate vaccines. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines (2020)
10. Warren WC, Jasinska AJ, García-Pérez R, et al. The genome of the vervet (Chlorocebus aethiops sabaeus). *Genome Res*. 2015;25(12):1921–1933

11. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics (Oxford, England). 2015 Oct;31(19):3210-3212.
12. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics (Oxford, England). 2007 May;23(9):1061-1067.
13. Nishimura O, Hara Y, Kuraku S. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*. 2017;33(22):3635-3637.
14. Rhie, A., Walenz, B.P., Koren, S. et al. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol 21, 245 (2020). https://doi.org/10.1186/s13059-020-02134-9.
15. Jason R. Miller, Arthur L. Delcher, Sergey Koren, Eli Venter, Brian P. Walenz, Anushka Brownley, Justin Johnson, Kelvin Li, Clark Mobarry, Granger Sutton. Aggressive assembly of pyrosequencing reads with mates,Bioinformatics 2008, vol. 24 (pg. 2818-2824).
16. Liu Y, Loewer M, Aluru S, Schmidt B. SNVSniffer: an integrated caller for germline and somatic single-nucleotide and indel mutations. *BMC Syst Biol*. 2016;10 Suppl 2(Suppl 2):47. Published 2016 Aug 1.
17. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220-1222.
18. Zhang Z., Schwartz S., Wagner L., & Miller W. (2000), "A greedy algorithm for aligning DNA sequences" J Comput Biol 2000; 7(1-2):203-14.
19. Sakuma, C., Sekizuka, T., Kuroda, M. *et al.* Novel endogenous simian retroviral integrations in Vero cells: implications for quality control of a human vaccine cell substrate. *Sci Rep* **8,** 644 (2018).
20. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9,** 1911 (2018).
21. Shen CF, Guilbault C, Li X, et al. Development of suspension adapted Vero cell culture process technology for production of viral vaccines. *Vaccine*. 2019;37(47):6996-7002
22. Zheng, G.X.; Lau, B.T.; Schnall-Levin, M.; Jarosz, M.; Bell, J.M.; Hindson, C.M.; Kyriazopoulou-Panagiotopoulou, S.; Masquelier, D.A.; Merrill, L.; Terry, J.M. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* **2016**, *34*, 303.

23. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB (2017). Direct determination of diploid genome sequences. *Genome Research* 27: 757-767 .
24. Yeo S, Coombe L, Warren RL, Chu J, Birol I. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics*. 2018;34(5):725-731.
25. Warren, R.L., Yang, C., Vandervalk, B.P. *et al.* LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *GigaSci* **4,** 35 (2015).
26. Shaun D. Jackman, Lauren Coombe, Justin Chu, Rene L. Warren, Benjamin P. Vandervalk, Sarah Yeo, Zhuyi Xue, Hamid Mohamadi, Joerg Bohlmann, Steven J.M. Jones and Inanc Birol (2018). Tigmint: correcting assembly errors using linked reads from large molecules. *BMC Bioinformatics*, 19(1). doi:10.1186/s12859-018-2425-6.
27. Warren RL. 2016. RAILS and Cobbler: Scaffolding and automated finishing of draft genomes using long DNA sequences. *The Journal of Open Source Software*.
28. René L Warren, Lauren Coombe, Hamid Mohamadi, Jessica Zhang, Barry Jaquish, Nathalie Isabel, Steven J M Jones, Jean Bousquet, Joerg Bohlmann, Inanç Birol, ntEdit: scalable genome sequence polishing, *Bioinformatics*, Volume 35, Issue 21, 1 November 2019, Pages 4430–4432.
29. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075.
30. Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*. 2005;33(Web Server issue):W465-W467.
31. Thibaud-Nissen F, Souvorov A, Murphy T, et al. Eukaryotic Genome Annotation Pipeline. 2013 Nov 14. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-.
32. Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).
33. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009 Aug;25(16):2078-2079.
34. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):W187-W191.

35. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92.
36. Dennis, G., Sherman, B.T., Hosack, D.A. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4,** R60 (2003).
37. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639-1645.
38. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993.
39. Kelley LA et al.The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10, 845-858 (2015).
40. Reynolds CR, Islam SA, Sternberg MJE.EzMol: A Web Server Wizard for the Rapid Visualization and Image Production of Protein and Nucleic Acid Structures, *J Mol Biol, Vol. 430* (2018).

AUTHOR CONTRIBUTION STATEMENT

MAS designed all experiments, executed most experiments, and prepared the manuscript. SK cultured the cells and prepared the samples for sequencing at the McGill Genome Centre. HD performed the raw reads filtering and Supernova first draft assembly. IR, YX, and AAK provided guidance, supervision, and critical reading of the manuscript. AAK provided project original design and supervision.

Preface to Chapter 4

In chapter 3 we assembled and annotated the genome for Vero cells to achieve highly contiguous genome reference, unravelling significant information on the genomics of Vero cells, explaining notably its peculiar nature when it comes to its ability to be infected a high multiplicity of infection, its significant viral genomic insertions observed and the loss of function of some key genes or pathways. It is now possible to carry on functional genomics analysis using this reference genome and transcriptomics analysis in order to link the phenotypes observed during the adaptation to suspension and Vero cells genomic and transcriptomic information in order to draw some key conclusions explaining the cause behind the limitations observed during the adaptation of Vero cells to suspension but also propose some new strategies to control those limitations and achieve a successful adaptation to suspension.

Chapter 4

Comparative transcriptomics analyses of a Vero cell line in suspension versus adherent culture conditions

Authors: **Marie-Angélique Sène[1]**, Yu Xia[1], Amine A. Kamen[1,*]

Author affiliation : Department of Bioengineering, McGill University, Montréal, QC, Canada

Notes: * Corresponding author.

Journal: Scientific Reports

Year: 2021

DOI:

Status: Under Review

ABSTRACT

The Vero cell line is the most used continuous cell line for viral vaccine manufacturing. Its anchorage-dependent use renders scaling-up challenging and operations very labor intensive which affects cost effectiveness. Thus, efforts to adapt Vero cells to suspension cultures have been invested but hurdles such as the long doubling time and low cell viability remain to be addressed. In this study, building on the recently published Vero cell line annotated genome, a functional genomics analysis of the Vero cells adapted to suspension is performed to better understand the genetic and phenotypic switches at play during the adaptation of Vero cells from anchorage-dependent to suspension cultures. Results show a downregulation of the epithelial to mesenchymal transition (EMT) pathway, highlighting the dissociation between the adaptation to suspension process and EMT. Surprisingly, an upregulation of cell adhesion components is observed, notably the CDH18 gene, the cytoskeleton pathway, and the extracellular pathway. Moreover, a downregulation of the glycolytic pathway are balanced by an upregulation of the asparagine metabolism pathway, promoting cell adaptation to nutrient deprivation. A downregulation of the adherens junctions and the folate pathways alongside with the FYN gene are possible explanations behind the currently observed low cell viability and long doubling time.

1.0 INTRODUCTION

Derived from a female *Chlorocebus sabaeus* (African Green Monkey) kidney, the Vero cell line is susceptible to infection by a wide range of viruses. Consequently, the Vero cell line served as a platform for the development and production of approved vaccines against dengue fever, influenza, Japanese encephalitis, polio, rabies, rotavirus, smallpox and more recently, Ebola (using a recombinant Vesicular Stomatitis Virus) and Covid-19 (Sinopharm, Sinovac, CoronaVac) [1-3], thus representing the most widely used continuous cell line for the production of viral vaccines over more than 40 years of manufacturing experience [4].

Most cell-based vaccine production platforms, including Vero cells, are adherent cells cultured in an anchorage-dependent manner thus hindering scale-up attempts due to the time and cost challenges posed by the development of large-scale anchorage-dependent cell culture platforms [5-7]. Therefore, in an effort to circumvent those challenges, stable cell lines adapted to suspension culture have been proposed over the years [8]. In the case of Vero cells in particular, significant efforts have been dedicated to developing micro-carriers cultures [9], agitation [10], and medium engineering [11] to achieve scalable processes and industrialization. Despite several reports stating that Vero cells adapted to suspension showed a higher production rate for viruses such as measles virus, rabies virus and vesicular stomatitis virus (VSV) [10,11], some issues such as the low cell viability and long doubling time of these cells were reported underlining the needs for more effective adaptation of Vero cells to suspension by exploring new avenues such as genetic engineering [10,12].

Recent advances in functional genomics and gene editing paved the way for a better understanding and high-throughput engineering of vaccine production cell lines, thus providing new possibilities

for cell line development and vaccine bioprocessing intensification. With the recent publication of an annotated assembly of the Vero cell genome [13], we propose a transcriptomics analysis of the differences between adherent and suspension Vero cells developed by Shen et al. [11]. This process will help highlight key differentially expressed genes and their impact on the cell's phenotype and their metabolic pathways for a better understanding of the process of adaptation to suspension, thus providing insight for new strategies to successfully adapt Vero cells to suspension cultures and enable streamlined scale-up and industrialization.

2.0 RESULTS

2.1 Differential expression analysis and Pathway enrichment analysis

Following the DESeq2 differential expression (DE) analysis (using adherent Vero cell samples as control and suspension Vero cell samples as case) and an applied p-value cut-off of 10^{-4}, among the 6627 DE genes that have been identified as highly significant (p-value < 10^{-4}), 1753 genes were identified as highly differentially expressed, with a |log2 fold change| > 2 (Figure 7, Supplementary table S1).

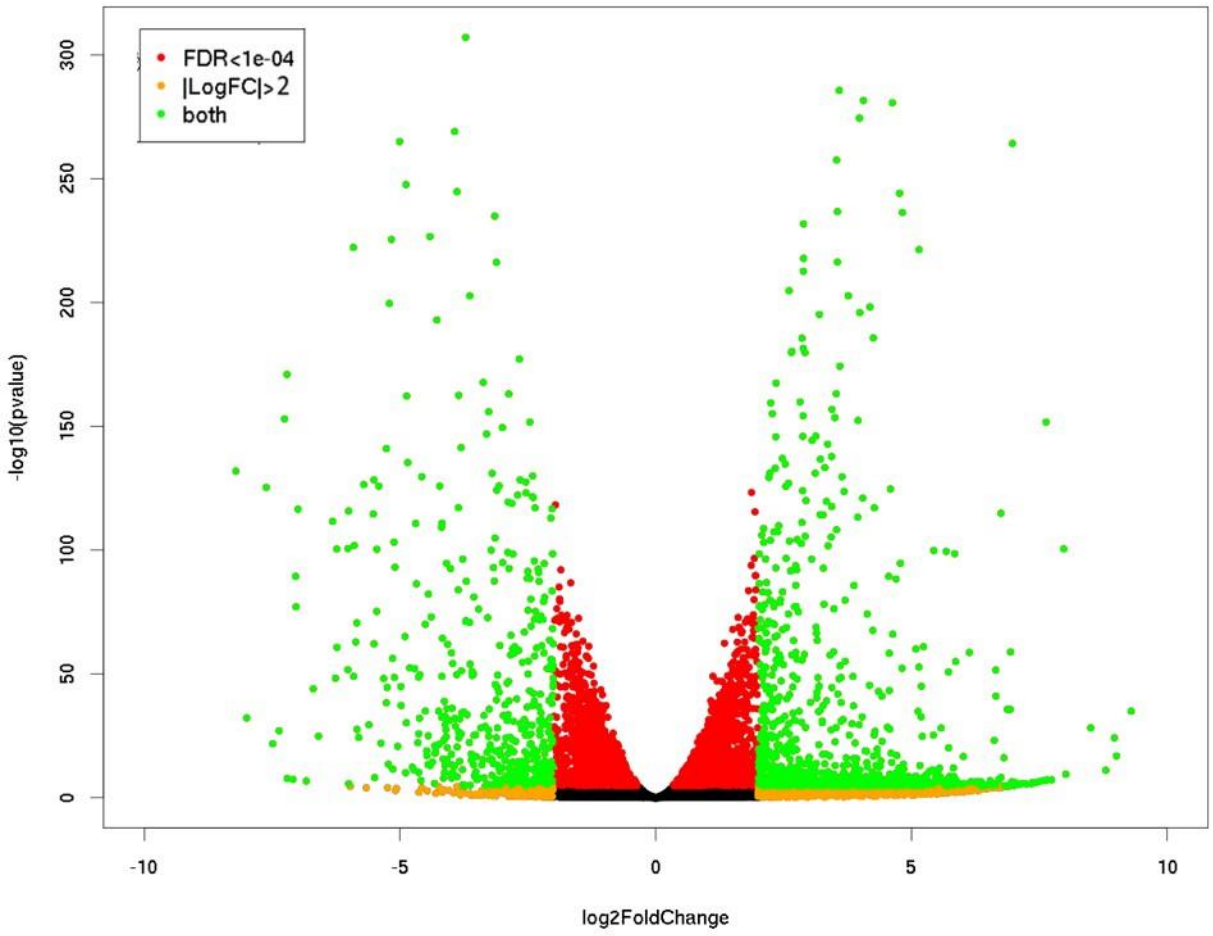


Figure 7 DESeq2 differential expression data quality control: Volcano plot of DE genes with applied p-value and log2 fold change cut-offs. FDR: False Discovery Rate; LogFC: log2 fold change.

The upregulated DE genes are highly enriched in key pathways such as vacuole, regulated exocytosis, plasma membrane bounded cell projection morphogenesis, actin filament-based process and regulation of cell adhesion with a p-value < 10^{-10} (Figure 8). On the other hand, downregulated DE genes are highly enriched in adherens junction, MYC targets, RNA processing and cell cycle related pathways with a p-value < 10^{-10} (Figure 9).

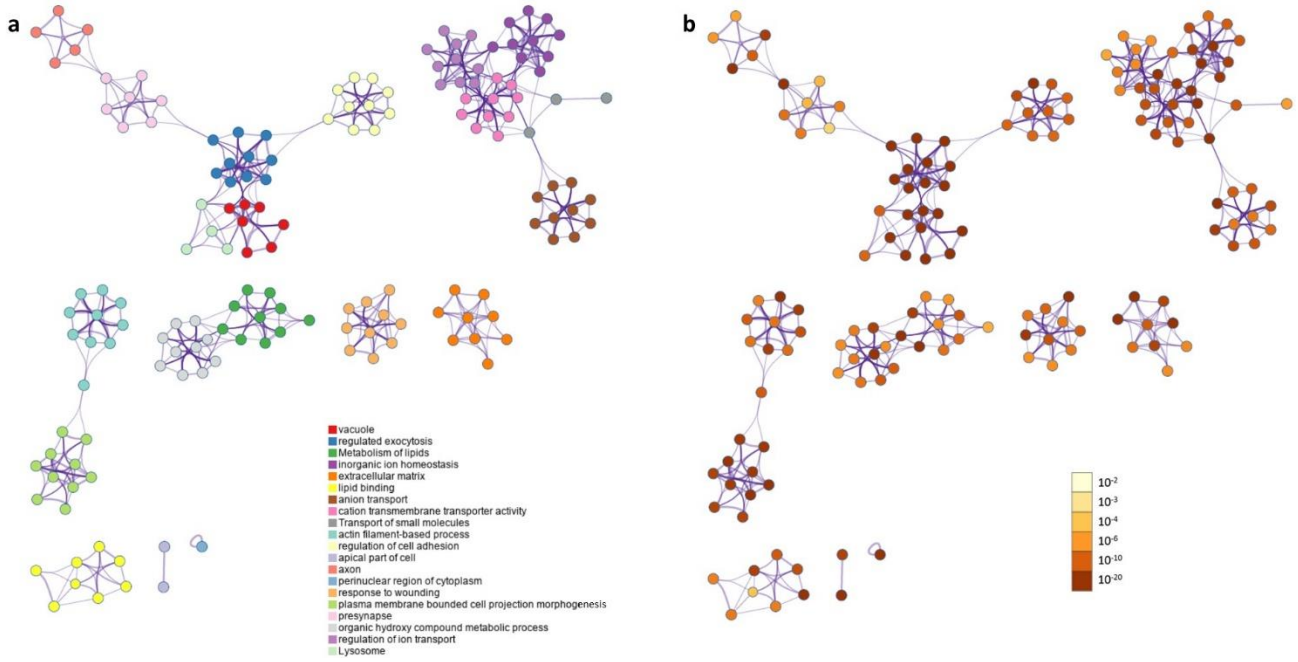


Figure 8 Network of terms enriched by upregulated DE genes: (a) colored by cluster ID, where nodes that share the same cluster ID are typically close to each other; (b) colored by p-value, where terms containing more genes tend to have a more significant p-value. Each node represents one enriched term and edges link similar terms (the edge thickness is proportional to the similarity between terms).

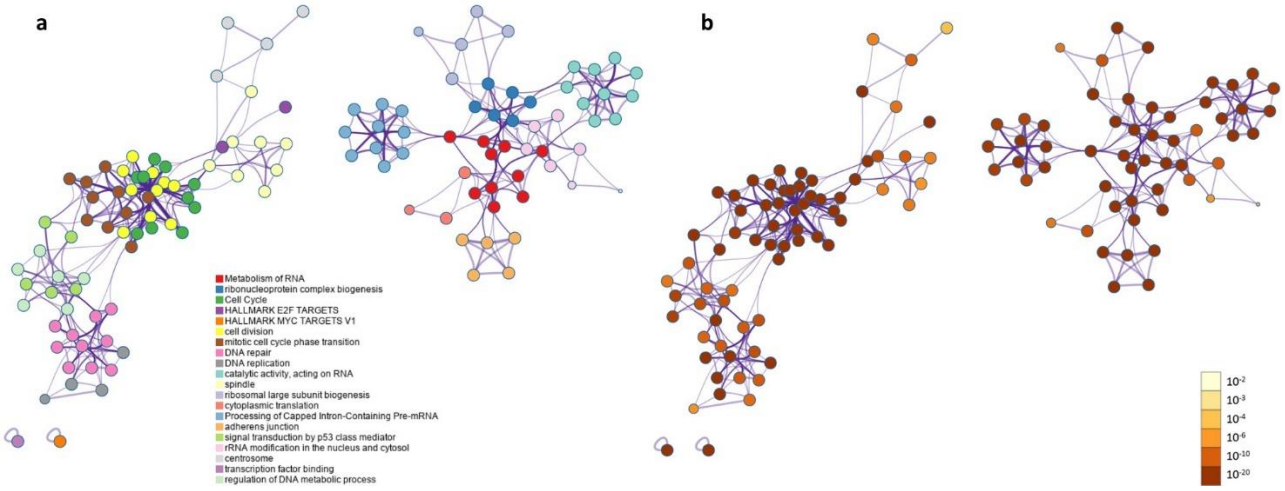


Figure 9 Network of terms enriched by downregulated DE genes: (a) colored by cluster ID, where nodes that share the same cluster ID are typically close to each other; (b) colored by p-value, where terms containing more genes tend to have a more significant p-value. Each node represents one enriched term and edges link similar terms (the edge thickness is proportional to the similarity between terms).

2.2 Metabolic pathway analysis

The extraction of metabolic genes and their scoring based on their relation with established metabolic pathways revealed an upregulation of the metabolic pathway of gluconeogenesis, galactose, pyrimidine, glycine, threonine, but also the serine pathway which plays a key role in unrestrained cell cycle progression [14], the asparagine metabolism pathway that is linked to cells adaptation to nutrient deprivation and/or hypoxia [15] and the mitochondrial one-carbon metabolism pathway which is implicated in rapid cell proliferation [16]. On the other hand, the adaptation to suspension led to a downregulation of key metabolic pathways such as proline, folate, aspartate, lipids, and glycolytic pathways (Figure 10, Supplementary table S2). Notably, deficiencies in the folate metabolism pathway were linked to growth limitations of BHK-21 in suspension culture [17].

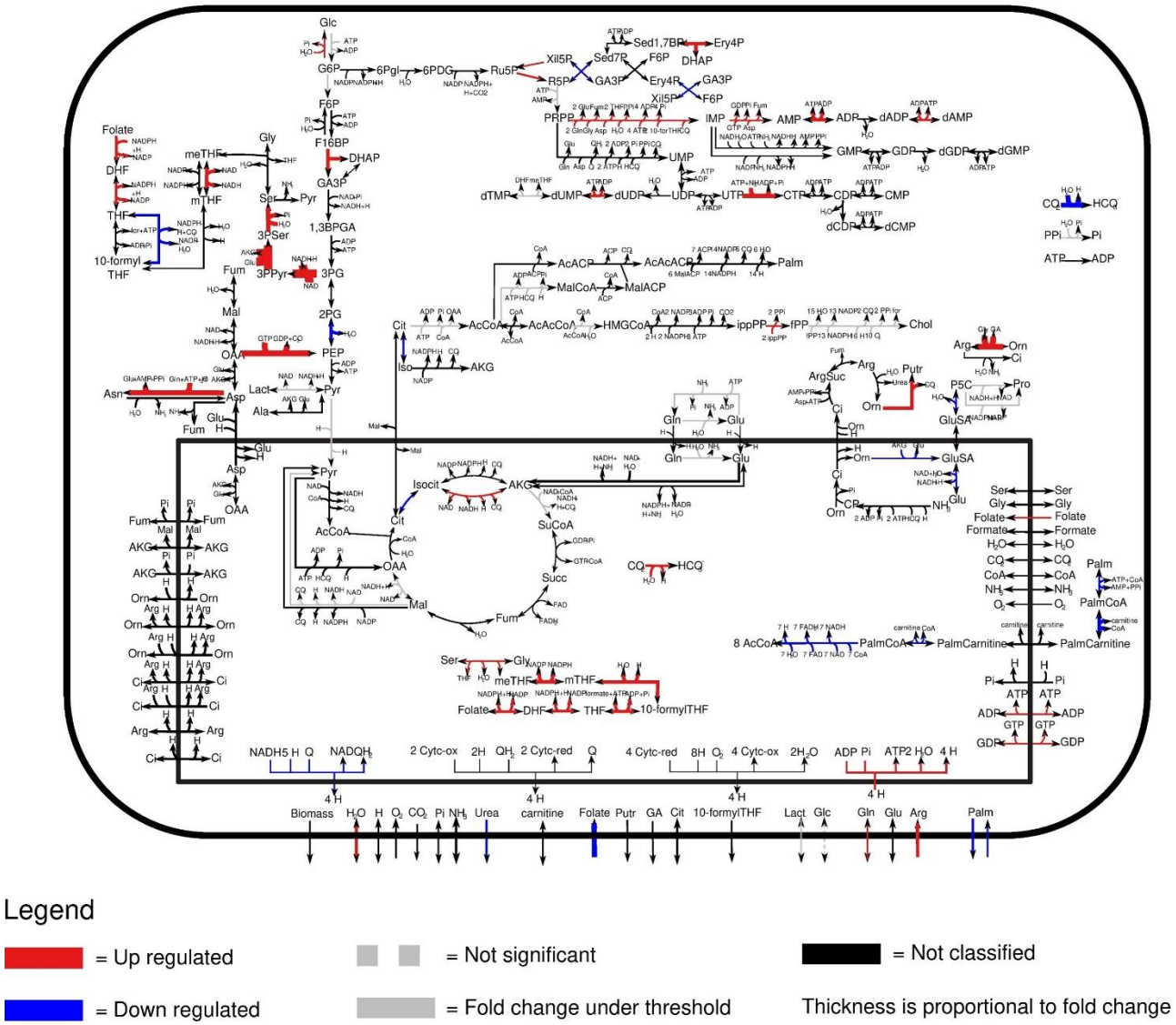


Figure 10 Adherent and suspension comparative metabolic map. Blue and red arrows refer respectively to downregulated and upregulated reactions. Dashed gray arrows refer to non-significant dysregulations according to Kolmogorov-Smirnov test with p-value 0.01. Solid gray arrows refer to reactions with a variation lower than 20%.

2.3 Gene set enrichment and network topology analysis

Gene set enrichment analysis (GSEA) with the hallmark gene set showed a significant upregulation of the interferon gamma response pathway and a down regulation of MYC targets (variant 1 and 2), E2F targets pathways but most importantly the epithelial mesenchymal transition pathway (Figure 11). Only pathways with an FDR(False discovery rate)< 0.05 were considered.

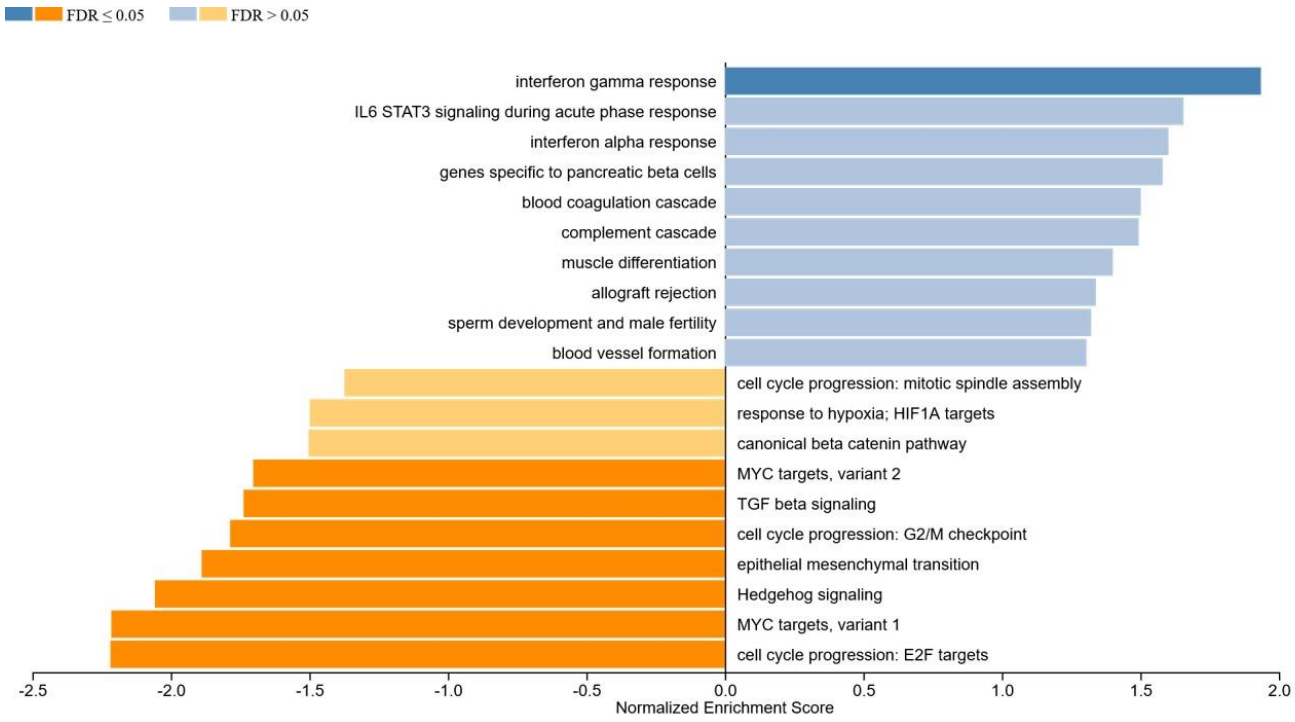


Figure 11 GSEA bar chart with significantly enriched hallmark pathways highlighted (FDR< 0.05).

In order to identify key genes involved in protein-protein interaction (PPI) networks, network topology analysis(NTA) was done for both upregulated genes (Table 6) and down-regulated genes (Table 7). The signaling, response to chemical and anatomical structure pathway associated with CRYAB, RHOU, ESR1, and C3 are upregulated alongside the cell adhesion pathway which is associated with CDH18. On the other hand, pathways associated with FYN and SMAD9 (cell

differentiation, system development, cellular response to growth factor) and the positive regulation of cell migration pathway associated with TRIP6 are downregulated.

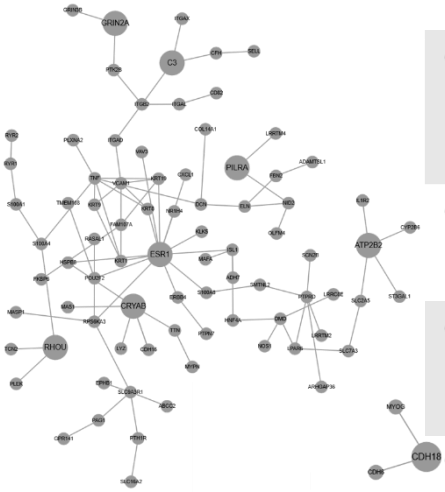
Sub-network Layout	Pathway GO ID	Pathway GO Name	Adjusted P-value	Top ranking associated Genes
	GO:0042221	Response to chemical	1.4515e-9	CRYAB, RHOU, ESR1, GRIN2A
	GO:0048731	System development	1.4515e-9	CRYAB, ESR1, C3, GRIN2A
	GO:0023052	Signaling	5.052e-9	PILRA, CRYAB, RHOU, ESR1, C3, GRIN2A, ATP2B2
	GO:0009605	Response to external stimulus	5.052e-9	ESR1, C3, GRIN2A
	GO:0009653	Anatomical structure morphogenesis	5.052e-9	CRYAB, RHOU, ESR1, C3
	GO:0007154	Cell communication	1.5582e-8	PILRA, CRYAB, RHOU, ESR1, C3, GRIN2A
	GO:0007155	Cell adhesion	1.7218e-8	CDH18

Table 6 Key upregulated networks and their top associated genes (NTA).

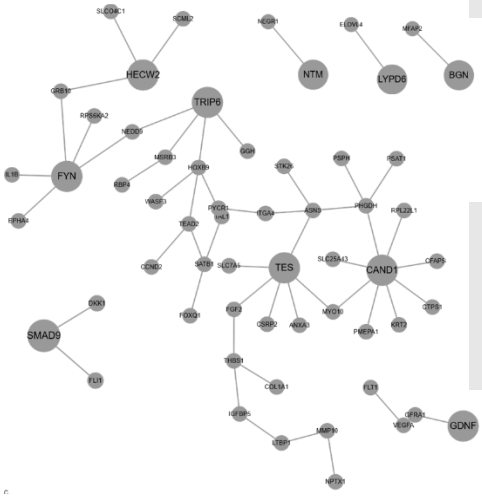
Sub-network Layout	Pathway GO ID	Pathway GO Name	Adjusted P-value	Top ranking associated Genes
	GO:0009653	Anatomical structure morphogenesis	8.6426e-8	FYN, GDNF, HECW2
	GO:0007167	Enzyme linked receptor protein signaling pathway	0.0000018879	FYN, SMAD9
	GO:0030154	Cell differentiation	0.0000030284	CAND1, FYN, GDNF, HECW2, NMT
	GO:0048731	System development	0.0000031743	FYN, GDNF, HECW2, NMT
	GO:0030335	Positive regulation of cell migration	0.0004	TRIP6
	GO:0071363	Cellular response to growth factor stimulus	0.0002	FYN, SMAD9
	GO:0045860	Positive regulation of protein kinase activity	0.0001	FYN

Table 7 Key downregulated networks and their top associated genes (NTA).

3.0 DISCUSSION

Over the years significant efforts have been made to adapt Vero cells to suspension in order to engineer high-throughput and scalable vaccine production platforms, however, those efforts were limited by hurdles such as low cell viability and long doubling time. In order to design more efficient strategies for successful adaptation to suspension cultures maintaining acceptable cell viability and doubling time, it is necessary to better understand the genetic and phenotypic changes triggered by the adaptation to the suspension state. Thus, we propose in this paper a comparative functional genomics analysis of adherent and suspension Vero cells from the gene to the protein interaction network level.

Indeed, correlations between differential expression analysis (Figure 8, 9), metabolic pathway enrichment analyses (Figure 10), gene set enrichment analyses (Figure 11) and network topology analyses (Tables 6, 7) highlight key events such as the upregulation of immune response, exocytosis and vacuole pathways which are involved in the storage of waste and other exocytosis molecules.

Furthermore, several events of checks-and-balances were observed across the different analysis methods used (Table 8): lipids metabolism and lipid pathways are also upregulated via pathway enrichment while the metabolic analysis revealed a downregulation of CPT1 and palmitate associated metabolic reactions which regulate fatty acid oxidation in mitochondria and whose upregulation impairs glucose homeostasis [18]. In parallel, the gluconeogenesis metabolic pathway is upregulated while the glycolytic pathway is downregulated thus hindering ATP generation. That resulting stress is met with not only a downregulation of the proline metabolic pathway which constitute a checkpoint that is reported to promote proline accumulation during

stress [19], but also an upregulation of asparagine metabolism which promotes the cell adaptation to nutrient depravation and/or hypoxia [15] thus promoting cell survival.

Key Finding	Analysis methods confirming finding
Lipid Metabolism	✓ PEA: lipids metabolism and lipid pathways upregulated ✓ MPA: downregulation of CPT1 and palmitate associated metabolic reactions
Response to stress	✓ PEA: Response to wounding pathway upregulated ✓ MPA: downregulation of the proline metabolic pathway and upregulation of asparagine metabolism
Low cell density	✓ PEA: adherens junction, MYC and E2F pathways downregulated ✓ MPA: aspartate metabolic pathway and the mitochondrial 1-carbon metabolic pathway downregulated ✓ GSEA: MYC and E2F pathways downregulated
Long doubling time	✓ PEA: cell division and mitotic cell cycle phase transition pathways downregulated ✓ MA: folate metabolic pathway downregulated ✓ GSEA: G2/M checkpoint downregulated ✓ NTA: FYN related PPI networks downregulated
Attempt to balance long doubling time	✓ MPA: glycine, threonine and serine pathway upregulated ✓ NTA: RHOU, ESR1 related PPI networks upregulated
Upregulation of cell adhesion	✓ PEA: actin filament-based process, regulation of cell adhesion, apical part of cell, plasma membrane bounded cell projection morphogenesis and extracellular matrix pathways upregulated ✓ NTA: upregulation of the cell adhesion PPI network(CDH18)

Table 8 Correlation between analysis methods for key findings. PEA: Pathway Enrichment Analysis; MPA: Metabolic Pathway Analysis; GSEA: Gene Set Enrichment Analysis; NTA: Network Topology-based Analysis.

Surprisingly, as observed in HEK293 cells adapted to suspension[20], cellular component organization pathways associated with cell adhesion such as actin filament-based process, regulation of cell adhesion, apical part of cell, plasma membrane bounded cell projection morphogenesis and extracellular matrix are upregulated via pathway enrichment analysis, while NTA showed an upregulation of the cell adhesion PPI network with CDH18 as its central gene and which can be considered as possible target for engineering in order to improve the cells adaptation to suspension. This upregulation of cell adhesion related genes could be due to the cells' attempt to restore the attachment to culture surfaces and surrounding cells which could explain the aggregates that are often observed in Vero cells suspension cultures and the cell rings that form on the suspension culture dishes.

On the other hand, the adherens junction pathway, which regulates cell-cell adhesion and is essential for viability (via the control of cell proliferation, polarity, shape, motility and survival) [21], is downregulated alongside the aspartate metabolic pathway, the mitochondrial 1-carbon metabolic pathway, the MYC and E2F targets pathway, which could explain the low cell density observed during Vero cells adaptation to suspension.

Moreover, the pathways related to cell division, mitotic cell cycle phase transition are downregulated via pathway enrichment, which is confirmed by the downregulation of the folate metabolic pathway which leads to cell cycle arrest at G0/G1 [22] and the downregulation of FYN which is central to the networks associated with the control of cell growth, thus, providing some insight in the origin of the long doubling time observed in Vero cells adapted to suspension culture. Nonetheless, Vero cells attempt to balance that effect on doubling time via the upregulation of the glycine, threonine and serine pathway which associated with an unrestrained cell cycle progression [14], and the upregulation of RHOU, ESR1 which are central to the anatomical

structure morphogenesis pathway and more precisely cell proliferation and antiapoptotic regulations.

In this study, we presented possible gene candidates for CRISPR/Cas gene editing in order to reduce the tendency of Vero cells adapted to suspension to upregulate their adhesion-related pathways. Inspired by Ley et al. AA catabolism engineering in CHO cells using CRISPR/Cas9[23], several culture conditions of suspension Vero cells with a variation in amino acids intake (based on the results obtain from our metabolic pathways analysis and more precisely the regulation of the serine, threonine and glycine pathway) can be sampled and analyzed via transcriptomics to highlight key gene targets for gene editing thus combining medium development and gene engineering to improve the growth and fitness of Vero cells adapted to suspension. The results reported by Ley et al. show that this strategy might be also promising for Vero cells but a prior, more in depth genomic analysis for the specific case of Vero cells is necessary before drawing any conclusions.

Lastly the epithelial to mesenchymal transition (EMT) pathway is downregulated in Vero cells adapted to suspension, thus highlighting the fact that the adaptation to suspension is not associated with EMT as previously shown with HEK293 cells adapted to suspension [20].

To conclude, we present in this paper key genes pathways at play during the adaptation of Vero cells to suspension and their complex checks-and-balances which could assist in the successful adaptation of Vero cells to suspension. Indeed, those key genes, notably associated with cell adhesion or hindering cell viability or doubling time could be potentially targeted via gene editing as new strategies for the adaptation to suspension, but also the observed competitions between the regulation of competing pathways can be studied more in detail via targeted perturbations using gene editing tools such as CRISPR [24].

4.0 MATERIAL AND METHODS

4.1 Cell lines and culture media

The adherent Vero WHO cell line studied in this work was at passage 138 (Neovacs). This cell line was itself derived from a vial of Vero ATCC CCL-81 which was send to WHO at passage 124 for analysis and establishment of the Vero WHO master cell bank approved for vaccine production. The cells were grown in static culture at 37 °C and 5% CO_2 in a humidified incubator (Infors HT, Switzerland). Cells were passaged twice weekly using TrypLE Express (Thermo Fisher Scientific) as dissociation reagent. A serum-free adapted sub-cell line grown in OptiPRO medium (Thermo Fisher Scientific) supplemented with 4 mM GlutaMAX (Thermo Fisher Scientific) was cryopreserved at a passage number of 151 in OptiPRO medium supplemented with 4 mM GlutaMAX and 10% DMSO (Sigma, USA).

The suspension Vero cell line was provided by the National Research Council (NRC) of Canada [11]. The cells were maintained at 37°C, 135 rpm and 5% CO_2 in a humidified Multitronorbital shaker (Infors HT, Switzerland) and were cultivated in 20 mL working volume of either IHM03 medium, provided by the NRC, or in MDXK medium (Xell AG, Germany), supplemented with 4 mM GlutaMAX (Thermo Fisher Scientific, USA) [4].

For transcriptome analysis, the media used for adherent and suspension cultures were serum-free. Vero WHO cells at passage 153 were washed in PBS (Wisent, Canada), harvested using TrypLE Express and centrifuged at 300 × g for 5 minutes. Suspension cells were also harvested and centrifuged in similar conditions. Cell pellets of around 6 million cells were quickly frozen in a

mixture of dry ice/ethanol and stored at −80 °C until further analysis. Adherent and suspension samples were generated in triplicates.

4.2 Differential gene expression analysis

Total RNA sequencing (TrueSeq) was performed using Illumina NovaSeq6000 Sprime v1.5, PE100. Following standard quality control, the reads were first aligned to the recently published Vero cell genome [13] using STAR [25] and the resulting BAM files were sorted by name using SAMtools [26] before read count. Transcripts were quantified using featureCounts [27]. Differential expression analysis of the raw read counts was done using DESeq2 [28] and quality control graphs were produced using DESeq2 and R package. The resulting differentially expressed gene list was filtered with a p-value cut-off of 0.0001.

4.3 Downstream analysis of differentially expressed genes

The differentially expressed (DE) genes were ranked based on their log2 fold change. Upregulated and downregulated DE genes between adherent and suspension cell lines were used for pathway enrichment analysis via the Metascape webtool and plotted using default parameters [29].

In order to identify the metabolic deregulations that distinguish adherent cells from suspension cells, metabolic genes from our list of differentially expressed genes were extracted and their Reaction Activity Score (RAS) were computed by solving Gene-Protein-Reaction (GPR) association rules based on the HMRcore metabolic network model via MaREA (Metabolic Reaction Enrichment

Analysis) [30]. A statistical comparison of the RAS computed for adherent samples and suspension samples was done via the Kolmogorov-Smirnov test and a metabolic map was generated.

WebGestalt (WEB-based GEne SeT AnaLysis Toolkit) [31] was used for gene set enrichment analysis with the hallmark gene set collection. To find differentially expressed pathways of genes between adherent and suspension cell lines, gene sets were filtered and the top 20 gene sets with an adjusted p-value lower than 0.05 were considered as significantly changed.

The upregulated part of the gene list generated by DESeq2 was filtered to consider genes with a |log2 fold change| > 2 for Network Topology Analysis (NTA) based on the Network Retrieval & Prioritization construction method [31] by first using random walk analysis to calculate random walk probability for the input gene IDs (seeds), then identifying the relationships among the seeds in the selected network to return a retrieval sub-network where the top 20 genes with the top random walk probability are highlighted. Indeed, assuming a tight connection between mechanistically important genes and a random distribution of other genes on the network, the Network Topology-based Analysis (NTA) uses random walk-based network propagation by identifying those genes which are potentially biologically significant. Our input gene IDs (upregulated genes previously filtered) were used as seeds and, based on its overall proximity (quantified by the random walk similarity) to the input seeds, each gene in the protein-protein interaction (PPI) network was attributed a score. Then the statistical significance of those scores was calculated via two p-values: a global p-value which significance is the result of a non-random association between the gene in the PPI network and the input seeds; and a local p-value which significance ensures that the gene did not acquire a significant association with the input seeds simply because of network topology.

Finally, enrichment analysis of the retrieved sub-networks was done using the PPI BIOGRID [32] database and Gene Ontology (GO) Biology Process terms[31]. The GO terms were first ranked based on their adjusted p-value and only the top 10 highly significant terms with an adjusted p-value cut-off of 0.01 were considered.

7.0 REFERENCES

1. Ammerman NC, Beier-Sexton M, Azad AF. Growth and maintenance of Vero cell lines. *Curr Protoc Microbiol*. 2008;Appendix 4:Appendix 4E.

2. Suder E, Furuyama W, Feldmann H, Marzi A, de Wit E. The vesicular stomatitis virus-based Ebola virus vaccine: From concept to clinical trials. *Hum Vaccin Immunother*. 2018;14(9):2107-2113.

3. COVID-19 vaccine tracker and landscape: https://www.who.int/teams/blueprint/covid-19/covid-19-vaccine-tracker-and-landscape

4. Kiesslich S, Kamen A, Vero cell upstream bioprocess development for the production of viral vectors and vaccines, *Biotech Adv*., Volume 44, 2020, 107608, ISSN 0734-9750.

5. Madeline B, Ribaud S, Xenopoulos A, Simler J, Schwamborn K, Léon A. Culturing a duck es-derived cell line in single-use bioreactors: a rapid, efficient, and cost-effective vaccine manufacturing system based on suspension culture. *Bioprocess Int*, 13 (3) (2015), (pp. 26-33).

6. Peschel B, Frentzel S, Laske T, Genzel Y, Reichl U. Comparison of influenza virus yields and apoptosis-induction in an adherent and a suspension MDCK cell line. *Vaccine*, 31 (2013), (pp. 5693-5699).

7. Yang J, Guertin P, Jia G, Lv Z, Yang H, Ju D. Large-scale microcarrier culture of HEK293T cells and Vero cells in single-use bioreactors. *AMB Exp*, 9 (2019), (pp. 70-83).

8. Merten OW. Development of serum-free media for cell growth and production of viruses/viral vaccines: safety issues of animal products used in serum-free media. *Dev Biol* (Basel) 2002; 111:233–257.

9. Chen A, Poh SL, Dietzsch C, Roethl E, Yan ML, Ng SK. Serum-free microcarrier based production of replication deficient influenza vaccine candidate virus lacking NS1 using Vero cells. *BMC Biotechnol*. 2011;11:81.

10. Rourou S, Ben Ayed Y, Trabelsi K, Majoul S, Kallel H. An animal component free medium that promotes the growth of various animal cell lines for the production of viral vaccines. *Vaccine*. 2014;32:2767–2769.

11. Shen CF, Guilbault C, Li X, et al. Development of suspension adapted Vero cell culture process technology for production of viral vaccines. *Vaccine*. 2019;37:6996–7002.

12. M. Logan, M. Aucoin. Media formulation to support the growth of Vero cells in suspension. *Proceedings of vaccine technology VII conference*, June 17-22, Mont Tremblant, Quebec, Canada (2018).

13. Sène MA, Kiesslich S, Djambazian H, Ragoussis J, Xia Y, Kamen AA. Haplotype-resolved de novo assembly of the Vero cell line genome. NPJ Vaccines. 2021 Aug 20;6(1):106. doi: 10.1038/s41541-021-00358-9. PMID: 34417462.

14. Amelio I., Cutruzzolá F., Antonov A., Agostini M., Melino G.. Serine and glycine metabolism in cancer. *Trends Biochem Sci*. 2014;39(4):191-198. doi:10.1016/j.tibs.2014.02.004.

15. Balasubramanian MN, Butterworth EA, Kilberg MS. Asparagine synthetase: regulation by cell stress and involvement in tumor biology. *Am J Physiol Endocrinol Metab*. 2013;304(8):E789-E799. doi:10.1152/ajpendo.00015.2013.

16. Nilsson R, Jain M, Madhusudhan N, et al. Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat Commun*. 2014;5:3128. doi:10.1038/ncomms4128.

17. Neugut, A., & Weinstein, I. (1979). Growth Limitation of BHK-21 Cells and Its Relation to Folate Metabolism. *In Vitro*, 15(5), 363-367. Retrieved March 29, 2021, from http://www.jstor.org/stable/4292213.

18. Assimacopoulos-Jeannet F, Thumelin S, Roche E, Esser V, McGarry JD, Prentki M. Fatty acids rapidly induce the carnitine palmitoyltransferase I gene in the pancreatic beta-cell line INS-1. *J Biol Chem*. 1997 Jan 17;272(3):1659-64. doi: 10.1074/jbc.272.3.1659. PMID: 8999842.

19. Liang X, Zhang L, Natarajan SK, Becker DF. Proline mechanisms of stress survival. *Antioxid Redox Signal*. 2013;19(9):998-1011. doi:10.1089/ars.2012.5074.

20. Malm, M., Saghaleyni, R., Lundqvist, M. et al. Evolution from adherent to suspension: systems biology of HEK293 cell line development. Sci Rep 10, 18996 (2020). doi: 10.1038/s41598-020-76137-8

21. Ivanov AI, Naydenov NG. Dynamics and regulation of epithelial adherens junctions: recent discoveries and controversies. *Int Rev Cell Mol Biol*. 2013;303:27-99. doi: 10.1016/B978-0-12-407697-6.00002-7. PMID: 23445808.

22. Yang Y, Li X, Sun Q, He B, Jia Y, Cai D, Zhao R. Folate deprivation induces cell cycle arrest at G0/G1 phase and apoptosis in hippocampal neuron cells through down-regulation of IGF-1 signaling pathway. *Int J Biochem Cell Biol*. 2016 Oct;79:222-230. doi: 10.1016/j.biocel.2016.08.040. Epub 2016 Aug 31. PMID: 27592453.

23. Ley D, Pereira S, Pedersen LE, et al. Reprogramming AA catabolism in CHO cells with CRISPR/Cas9 genome editing improves cell growth and reduces byproduct secretion. Metab Eng. 2019;56:120-129. doi:10.1016/j.ymben.2019.09.005

24. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 9, 1911 (2018).

25. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics*, Volume 29, Issue 1, January 2013, (pp. 15–21).

26. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (Oxford, England). 2009 Aug;25(16):2078-2079.

27. Liao Y, Smyth GK, Shi W, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics*, Volume 30, Issue 7, 1 April 2014, (pp. 923–930).

28. Love M.I., Huber W., Anders S. (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, 15, 550. doi: 10.1186/s13059-014-0550-8.

29. Zhou et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* (2019) 10(1):1523.

30. Damiani, C., Rovida, L., Maspero, D., Sala, I., Rosato, L., Di Filippo, M., Pescini, D., Graudenzi, A., Antoniotti, M., & Mauri, G. (2020). MaREA4Galaxy: Metabolic reaction enrichment analysis and visualization of RNA-seq data within Galaxy. *Computational and Structural Biotechnology Journal*, 18, 993–999.

31. Liao Y., Wang J., Jaehnig E. J., Shi Z., Zhang B., WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, (pp. W199–W205).

32. Stark C. et al. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. (2006) 34:D535-539.

8.0 ACKNOWLEDGEMENTS

This work was supported by funding from the Natural Sciences and Engineering Research Council (NSERC-Grant-CRDPJ 511957). M-A-S is supported by the McGill Engineering Doctoral Award (MEDA), and A.A.K by a Canada Research Chair (CRC-240394).

The authors would like to thank Sascha Kiesslich for providing cell culture samples for sequencing.

9.0 AUTHOR CONTRIBUTION STATEMENT

M.-A.S. designed, executed all experiments and prepared the manuscript. Y.X., and A.A.K. pro-vided guidance, supervision, and critical reading of the manuscript.

10.0 COMPETING INTERESTS STATEMENT

The authors declare that they have no financial or nonfinancial competing interests.

11.0 DATA AVAILABILITY STATEMENT

The Vero genome assembly and generated RNA sequencing reads are accessible under the Bioproject PRJNA644395. All other relevant data are available upon request.

12.0 CODE AVAILABILITY STATEMENT

This study did not make use of custom code.

Preface to Chapter 5

In Chapter 4, the transcriptomics analysis of Vero cells adapted to suspension helped understand the process further and point out the causes behind the low cell density, long doubling time and formation of aggregates.

Notably, based on those findings, the following avenues could be taken to improve the adaptation of Vero cells to suspension via media formulation by reducing cell aggregates by increasing Calcium and magnesium intake; reducing cell doubling time by increasing the glycine, threonine, serine, and folate intake; increasing cell density by increasing aspartate and glucose intake. Or via gene editing by targeting CDH18 or FYN for knockout.

Aside from the study in chapter 4 of the adaptation to suspension, the second main hypothesis of this thesis is: by studying the genomics aspect of host-virus interaction during infection, it is possible to detect key host factors triggered during viral infection of Vero cells, to find best gene targets for CRISPR gene editing to increase virus production in Vero cells, thus, producing an engineered cell line for high yield virus production. Therefore, in this Chapter 5, the reference genome generated in Chapter 3 and the computation protocol used in Chapter 4 were combined to first study the interactions between Vero cells and two viruses of interest during infection and choose the most attractive antiviral gene target for genomic deletion to increase Vero cells viral yield.

Chapter 5

From Functional Genomics of Vero Cells to CRISPR-based Genomic Deletion: Understanding Host-Virus Interactions for Improved Viral Production Rates

Authors: **Marie-Angélique Sène[1]**, Yu Xia[1], Amine A. Kamen[1,*]

Author affiliation : [1] Department of Bioengineering, McGill University, Montreal, QC, Canada

* Corresponding author

Journal: Biotechnology and Bioengineering

Year: 2022

DOI: pending

Status: Under review

ABSTRACT

Despite their wide use in the vaccine manufacturing field for over 40 years, one of the main limitations to recent efforts to develop Vero cells as high-throughput vaccine manufacturing platforms is the lack of understanding of virus-host interactions during infection and cell-based virus production in Vero cells. To overcome this limitation, this manuscript uses the recently generated reference genome for the Vero cell line to identify the factors at play during Influenza A virus (IAV) and recombinant vesicular stomatitis virus (rVSV) infection and replication in Vero host cells. The best antiviral gene candidate for gene editing was selected using Differential Gene Expression analysis, Gene Set Enrichment Analysis and Network Topology-based Analysis. After selection of the ISG15 gene for targeted CRISPR genomic deletion, the ISG15 genomic sequence was isolated for CRISPR guide RNAs design and the guide RNAs with the highest knockout efficiency score were selected. The CRISPR experiment was then validated by confirmation of genomic deletion via PCR and further assessed via quantification of ISG15 protein levels by western blot. The gene deletion effect was assessed thereafter via quantification of virus production yield in the edited Vero cell line. A 70-fold and an 87-fold increase of total viral particles productions in ISG15-/- Vero cells was achieved for, respectively, IAV and rVSV while also increasing infectious viral particles release that was around 3.2-5.4% to up to 65.3-67.9%.

1.0 INTRODUCTION

Vero cells are a female African Green Monkey kidney derived cell line widely used over 40 years for viral vaccine production[1] including the vaccine against dengue fever, influenza, Japanese encephalitis, polio, rabies, rotavirus, smallpox and more recently, Ebola (using a Vesicular Stomatitis Virus (VSV) recombinant)[2-4]

The Vero cell line is part of the various cell culture-based platforms used to produce influenza vaccines to counter the drawbacks of egg-based vaccine production methods by providing a high throughput, robust and cross-contamination-free alternative. Nonetheless, the Vero cell line's low influenza virus production rate significantly hinders its potential for wider use.

Among the numerous approved Vero cell-based vaccines, an Ebola vaccine designed using a pseudotyped recombinant vesicular stomatitis virus (rVSV) has been shown to be safe for human administration[5] which further stresses the potential of rVSV as an effective vaccine production platform given the generally asymptomatic nature of VSV infections in humans[6].

Furthermore, not only were Vero cells identified as the cell line with the highest susceptibility to MERS-CoV[7], SARS-CoV and recently SARS-CoV-2[8], but also, several inactivated COVID-19 vaccines were approved and are used to immunize millions such as Sinopharm, Sinovac, CoronaVac[9]. Thus, successful engineering of Vero cell lines to significantly increase the viral production of several viruses including Influenza A virus and rVSV would greatly impact global health.

Recent advances in gene editing made it possible to edit the genome of cell lines and notably those used for viral infection studies and vaccine production through process intensification exploiting

the available genomic information. Previous attempts were made to increase viral production rates in Vero cells through gene editing using a genome-wide RNA interference screen dataset to select gene targets[10] but no significant increase in production yields was observed after repeated experiments[11]. Possible reasons cited by the authors to explain such results include the use of a genome other than the Vero cell genome, and the fact that the phenotypes induced by transcriptional suppression (RNAi-based knockdown) and genetic deletion (CRISPR knockout) differ in such a way that the former increases virus production while the latter does not. Thus, it is necessary to propose a new approach that does not rely on screens to identify valuable gene targets for CRISPR/Cas9 editing.

With the availability of complete genome sequences and accurate annotations of genes and their products, the field of functional genomics has emerged as an alternative to classical molecular biology gene-by-gene-based approaches to study genes and proteins interactions and their phenotypic effects by using genome-wide methods combining genomics, epigenomics, transcriptomics, proteomics, and metabolomics among others[12].

With the recent publication of the Vero cell genome[13], we hereby propose a novel approach combining the study of host-virus interactions during Vero cells infection with Influenza A virus (IAV) Puerto Rico 8 and rVSV-GFP to identify gene targets and a genomic deletion method using CRISPR/Cas9[14] with subsequent single-cell cloning to generate an engineered cell line with high yield viral production rates.

2.0 MATERIALS AND METHODS

2.1 Cell lines and culture media

The Vero WHO cell line studied in this work was at passage 138. This cell line was derived from a vial of Vero ATCC CCL-81 which was sent to WHO at passage 124 for analysis and establishment of the Vero WHO master cell bank approved for vaccine production. The cells were grown in static culture at 37 °C and 5 % CO_2 in a humidified incubator (Infors HT, Switzerland). Cells were passaged twice weekly using TrypLE Express (Thermo Fisher Scientific) as a dissociation reagent. A serum-free adapted sub-cell line grown in OptiPRO medium (Thermo Fisher Scientific) supplemented with 4 mM GlutaMAX (Thermo Fisher Scientific) was cryopreserved at a passage number of 151 in OptiPRO medium supplemented with 4 mM GlutaMAX and 10 % DMSO (Sigma, USA).

A quality control kinetics experiment was run in which, Vero cells were infected with influenza A virus Puerto Rico 8 strain and rVSV-GFP at MOI 10 (MOI: Multiplicity of Infection) to quantify the viral production rate and the cells viability over time. That led to the identification of the best time window to harvest samples for RNA sequencing (highest viability and before induction of cell death to avoid having false enrichment of the cell death-related pathways).

The supernatant was harvested at several time points and cell viability was monitored. To quantify the virus production for each time point, $TCID_{50}$ and hemagglutination assay were used.

Following the kinetics analysis, for transcriptome analysis, Vero WHO cells at passage 153 were infected with Influenza A virus Puerto Rico 8, rVSV-GFP and harvested before visible cytopathic effect at 24 hpi and 6 hpi respectively. Due to the nature of rVSV infection which makes the

cytopathic effect clearly visible at the microscope at early stages, monitoring of cell viability was done using the microscope. The viability was assessed using a cell counter for the IAV, since the cells are able to withstand infection at MOI 10 for more than 48 hours without visible cytopathic effects.. The samples were harvested using TrypLE Express and centrifuged at 300 × g for 5 minutes. Cell pellets of around 6 million cells were lysed and quickly frozen in a mixture of dry ice/ethanol and stored at −80 °C until further analysis, samples made of non-infected cells were also prepared and sent to sequencing as the control batch. All the samples were generated in triplicates.

2.2 Differential gene expression analysis

Total RNA sequencing (TrueSeq) was performed using Illumina NovaSeq6000 Sprime v1.5, PE100. Following standard quality control, the reads were first aligned to the recently published Vero cell genome[13] using STAR[15] and the resulting BAM files were sorted by name using SAMtools[16] before read count. Transcripts were quantified using featureCounts[17]. Differential expression analysis of the raw read counts was done using DESeq2[18]. The resulting differentially expressed gene list was filtered with a p-value cut-off of 0.0001.

2.3 Downstream analysis of differentially expressed genes

The differentially expressed (DE) genes were ranked based on their log2 fold change.

WebGestalt (WEB-based GEne SeT AnaLysis Toolkit)[19] was used for gene set enrichment analysis with the Reactome gene set collection. To find differentially expressed pathways of genes between

non-infected and infected cells, gene sets were filtered and the top 20 gene sets with an adjusted p-value lower than 0.05 were considered as significantly changed.

The upregulated part of the gene list generated by DESeq2 was filtered to consider genes with a |log2 fold change| > 2 for Network Topology Analysis (NTA) based on the Network Retrieval & Prioritization construction method[20] by first using random walk analysis to calculate random walk probability for the input gene IDs (seeds), then identifying the relationships among the seeds in the selected network to return a retrieval sub-network where the top 20 genes with the top random walk probability are highlighted. Indeed, assuming a tight connection between mechanistically important genes and a random distribution of other genes on the network, the Network Topology-based Analysis (NTA) uses random walk-based network propagation by identifying those genes which are potentially biologically significant. Our input gene IDs (upregulated genes previously filtered) were used as seeds, and based on their overall proximity (quantified by the random walk similarity) to the input seeds, each gene in the PPI network was attributed a score. Then, the statistical significance of those scores was calculated via two p-values: a global p-value whose significance is the result of a non-random association between the gene in the PPI network and the input seeds; and a local p-value whose significance ensures that the gene did not acquire a significant association with the input seeds simply because of network topology.

Finally, enrichment analysis of the retrieved sub-networks was done using the protein-protein interaction (PPI) BIOGRID[21] database and Gene Ontology (GO) Biology Process terms[22]. The GO terms were first ranked based on their adjusted p-value and only the top 10 highly significant terms with an adjusted p-value cut-off of 0.01 were considered.

2.4 ISG15 protein sequences comparison

ISG15 protein sequences were retrieved from RefSeq for Vero cells (XP_007979280.1), human (NP_005092.1), mice (NP_056598.2) and canine (XP_003639101.1). The sequences were aligned using T-Coffee[23] and exported to the ESPript server[24] for sequence alignment graphic design. Regions known to interact with viruses are also highlighted.

2.5 Genomic deletion using CRISPR/Cas9

The strategy used for this genomic deletion protocol relies on cellular delivery of a pair of chimeric single guide RNAs (sgRNAs) to create two double-strand breaks (DSBs) at a locus to delete the intervening DNA segment by non-homologous end joining (NHEJ) repair. This method has been used to delete genes with a length between 1 to 10 kb[14] and is being applied here for the deletion of the genes CDS regions. Genomic deletions are more advantageous compared to HDR or single-site small indel production because not only does the high frequency of deletions limits the number of clones needed to be screened to find clones of interest, and monoallelic and biallelic deletions can be easily identified via PCR avoiding labor intensive methods, but also, given that a significant portion of the gene of interest is deleted, reliable loss-of-function alleles can be obtained.

A pair of guide RNAs was designed using freely available online tools CRISPOR[25] and EuPaGDT[26] which already included the Vero cell genome in their list of custom genomes. These tools helped identify guide sequences that minimize identical genomic matches or near-matches to reduce the risk of cleavage away from target sites (off-target effects). The guide sequences consist of a 20-mer ("protospacer sequence") upstream of an "NGG" sequence ("protospacer adjacent motif" or PAM)

at the genomic recognition site. The plasmid structures pX458 (Addgene plasmid ID 48138) containing our designed gRNAs (guide A: ACCAGCATTCGAGCAAGATCAAGG; guide B: GGAAACCGAAACTTGGCCACCGG), which include GFP as a selectable marker, were purchased from GenScript.

The delivery of CRISPR/Cas9 plasmids was done by electroporation. 4 vials containing 2.6×10^6 cells in 90 uL of growth medium were prepared for transfection. The cells were washed two times in ice-cold phosphate-buffered saline (PBS), resuspended and transferred to a 4 mm gapped cuvette. 4 tubes of 10 mL growth media were prepared and put into the incubator for 10 min. 5 µg of each CRISPR/Cas9 construct A and B were mixed with the Vero cell suspensions and the samples were immediately pulsed using an electroporator at 250 volts square wave for 20 ms. The cells were then diluted into the previously prepared 10 mL prewarmed complete growth media and plated in a T75 cm^2 flask before incubation at 37°C, 5% CO2 for 48 hours. For all studies, non-transfected cells were included as a negative control.

2.6 FACS sorting, single cell cloning and gene editing validation

The top ~3% of GFP positive cells were sorted using FACS to enrich for cells that received high levels of the CRISPR/Cas9 constructs. The sorted cells were individually plated into 96-well plates containing 100 µl per well of cell culture media using FACS sorter. The clones were incubated at 37 °C for 3 weeks. The resulting monoclonal colonies were passaged and split to proceed with validation steps.

PCR was used to validate the intended genomic deletion of the ISG15 CDS region. One set of primers internal to the sequence to be deleted (non-deletion band) and another set of primers upstream and downstream of the sgRNA cleavage sites (deletion band) were designed (Figure 12, Table 9). In the absence of deletion, the deletion band is often too large to efficiently amplify. Primers at least 100 bp

from the predicted cleavage site were used to ensure detection would not be impacted by a small indel at the sgRNA target site. The genomic DNA was extracted from each clone using Invitrogen PureLink Genomic DNA Mini Kit and DNA concentration was measured. Each clone was screened for both non-deletion band and deletion band detection using the following PCR protocol: for each detection, a 25 µL PCR reaction containing 12.5 µL master mix, 0.5 µL forward primer (10 µM), 0.5 µL reverse primer (10 µM), 100 ng gDNA, and H_2O up to 25 µL was run in the thermocycler (98 °C for 30s, 35 cycles of (98 °C for 10 sec, 60 °C for 30s, 72 °C for 1 min), and 72 °C for 2 min). The PCR products were then run on 2% agarose gel at 10 V/cm using 1x Tris-acetate-EDTA (TAE) buffer. The samples were examined for the detection of non-deletion and deletion bands using a Chemidoc (Biorad) and clones with biallelic deletions were passaged and split for cell banking and further validation analysis. This validation was repeated at a week's interval for quality control. 7 independent clones with biallelic deletions were used for the remaining validation steps.

Band to be detected	Forward Primer	Reverse Primer
Non-deletion Band	GTCCCAGCTCTGCAGACATTA	GAGCTCGGCCAGGTTCTAAG
Deletion Band	CCTCGAGGCTGTAACTGCAA	ACCATAGGGGTGTTTTCCGT

Table 9 Primers designed for genomic deletion validation

Following the validation of the intended genomic deletion of the ISG15 CDS region, validation at the protein level was done via western blot to ensure that the ISG15 protein is indeed deleted. 20 µL of each cell lysate sample were mixed with SDS loading buffer, separated on SDS-PAGE gels (BioRad Criterion TGX Precast gels), and transferred to polyvinylidene difluoride (PVDF) membranes. Immunoblotting was performed using the relevant antibodies (anti-ISG15, Invitrogen). Horseradish peroxidase coupled secondary antibodies were detected with the BioRad

Clarity Western ECL substrate. The resulting signals were imaged with a Chemidoc (BioRad) and analyzed by ImageJ.

Following the confirmation of ISG15 deletion, the next validation step consisted of verifying the phenotypic impact of such deletion on virus reproduction. Therefore, triplicates of wild-type Vero cells and ISG15 -/- Vero cells were cultured and infected at MOI 10 with, on one hand, IAV PR8 and on the other hand rVSV-GFP. The supernatant for each sample was harvested 24 hours post-infection and virus production was quantified via ddPCR (viral genome to quantify the total number of particles) and TCID50 (to quantify the number of infectious viral particles) as previously described[27].

3.0 RESULTS

3.1 Preliminary kinetics study

The optimal harvesting time point for RNA sequencing was selected based on the infectious viral particles production level and the viability of the cells. Thus, for IAV (Figure 12), the selected time points were at the peak of infectious viral particles production (i.e. 24hpi). For rVSV-GFP (Figure13), given the observation of significant cytopathic effects at the initial stages (8 hpi), the time point selected was 6hpi to ensure that pathways such as cell death are not falsely enriched simply due to sampling quality issues.

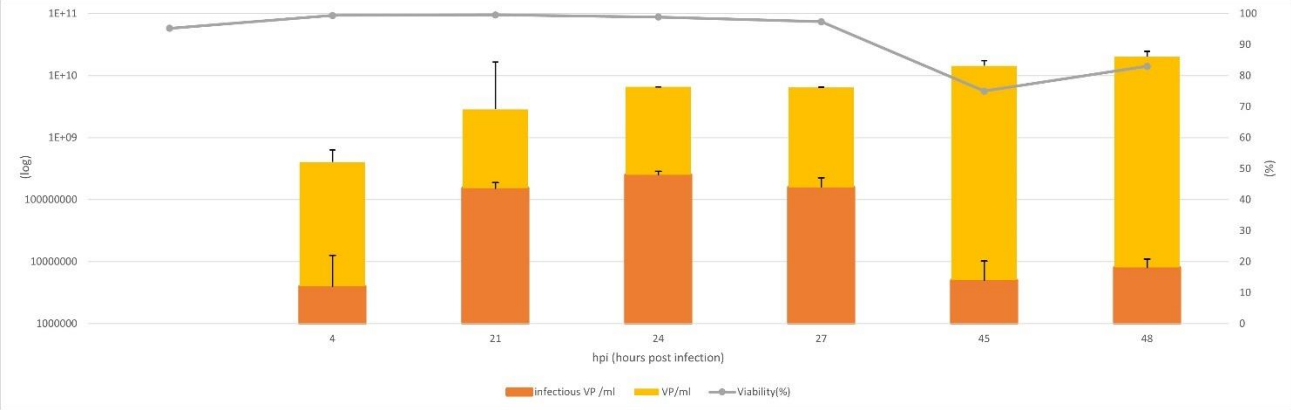


Figure 12 Kinetics of IVA PR8 production in Vero cells

Quantification of both viral particles (VP) and infectious particles and monitoring of cells viability.

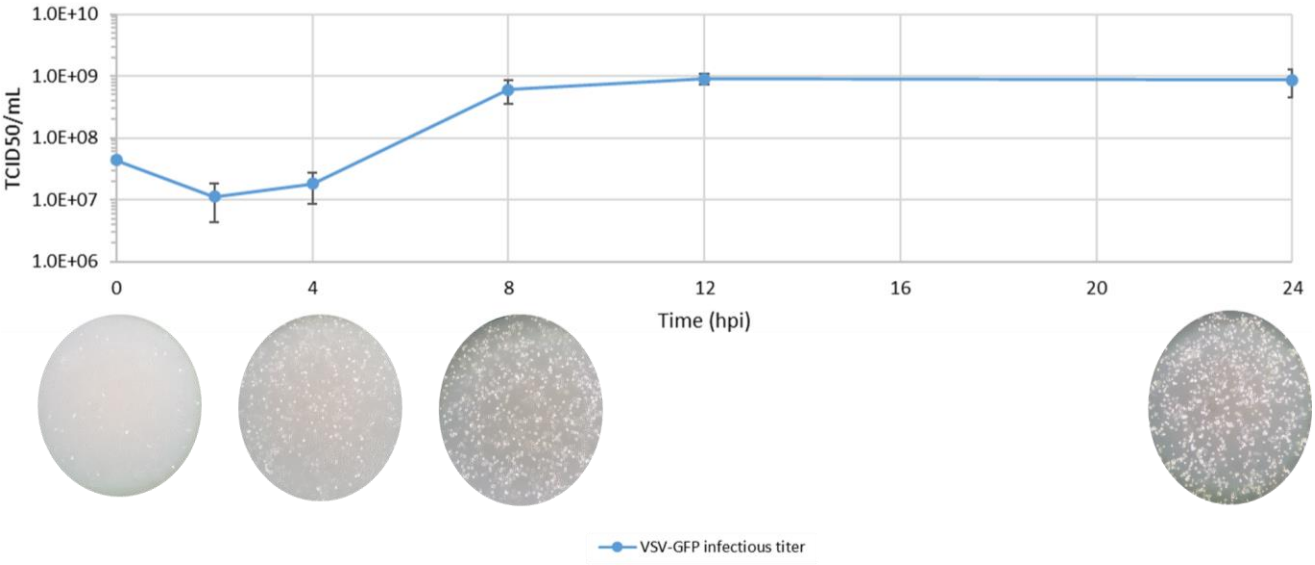


Figure 13 Kinetics of rVSV-GFP production in Vero cells

Quantification of infectious particles and monitoring of cytopathic effects via microscope.

2.2 Functional genomics analysis and gene target selection

Following the DESeq2[18] differential expression (DE) analysis and an applied p-value cut-off of 10^{-4}, Gene set enrichment analysis (GSEA) using Reactome[28] as a gene set showed, in the case of IAV infection at 24hpi (Figure 14),a downregulation of major RNA processing gene sets such as the Influenza viral RNA transcription and replication gene set, the activation of mRNA upon binding of the cap-binding complex and eIFs and L13a-mediated translational silencing of Ceruloplasmin expression which correlates with viruses evasion strategies via cap snatching and the host cell's attempts to counter that evasion. On the other hand, selenium related pathways are also downregulated such as selenoamino acid metabolism and selenocysteine synthesis. Indeed, it was previously shown that selenium and selenoproteins deficiency leads to increased host susceptibility to viral infection[29]. Meanwhile, key immune response related pathways are significantly upregulated such as interferon signalling, cytokine signalling and chemokine

signalling.

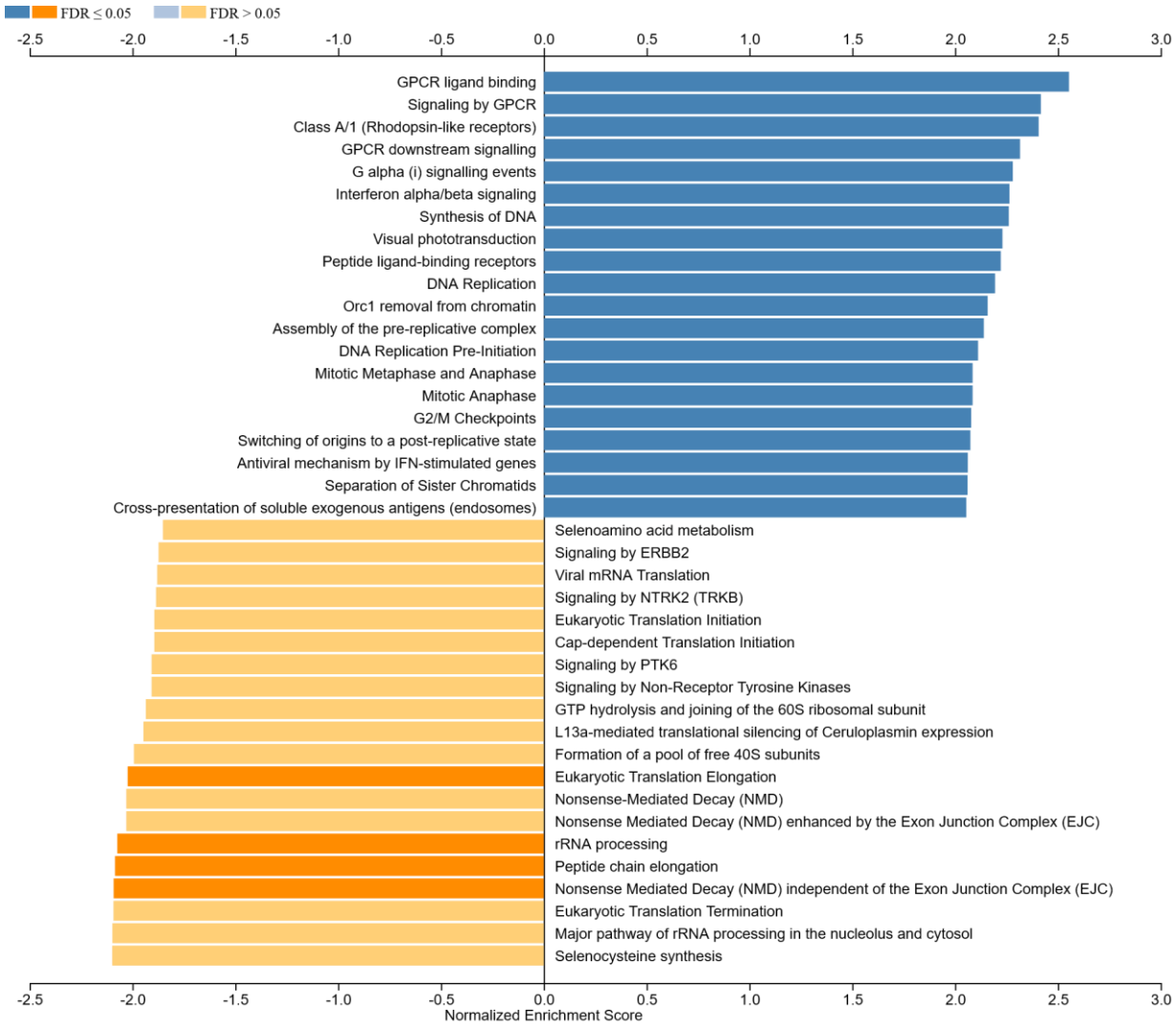


Figure 14 GSEA bar chart with significantly enriched hallmark pathways for 24hpi Influenza infection highlighted (FDR< 0.05)

More precisely an upregulation of IFN-stimulated genes and ISG15 antiviral mechanism are upregulated. Similar to IAV, rVSV-GFP interaction with Vero cells at 6hpi showed (Figure 15) downregulation of one of the key quality control mechanisms of RNA processing: the nonsense mediated decay (NMD)[30] thus promoting viral reproduction alongside the downregulation of eukaryotic translation elongation. Moreover, the previously identified antiviral pathways related

to interferons are also upregulated in the case of rVSV-GFP infection, notably, the antiviral mechanism by IFN-stimulated genes such as ISG15.

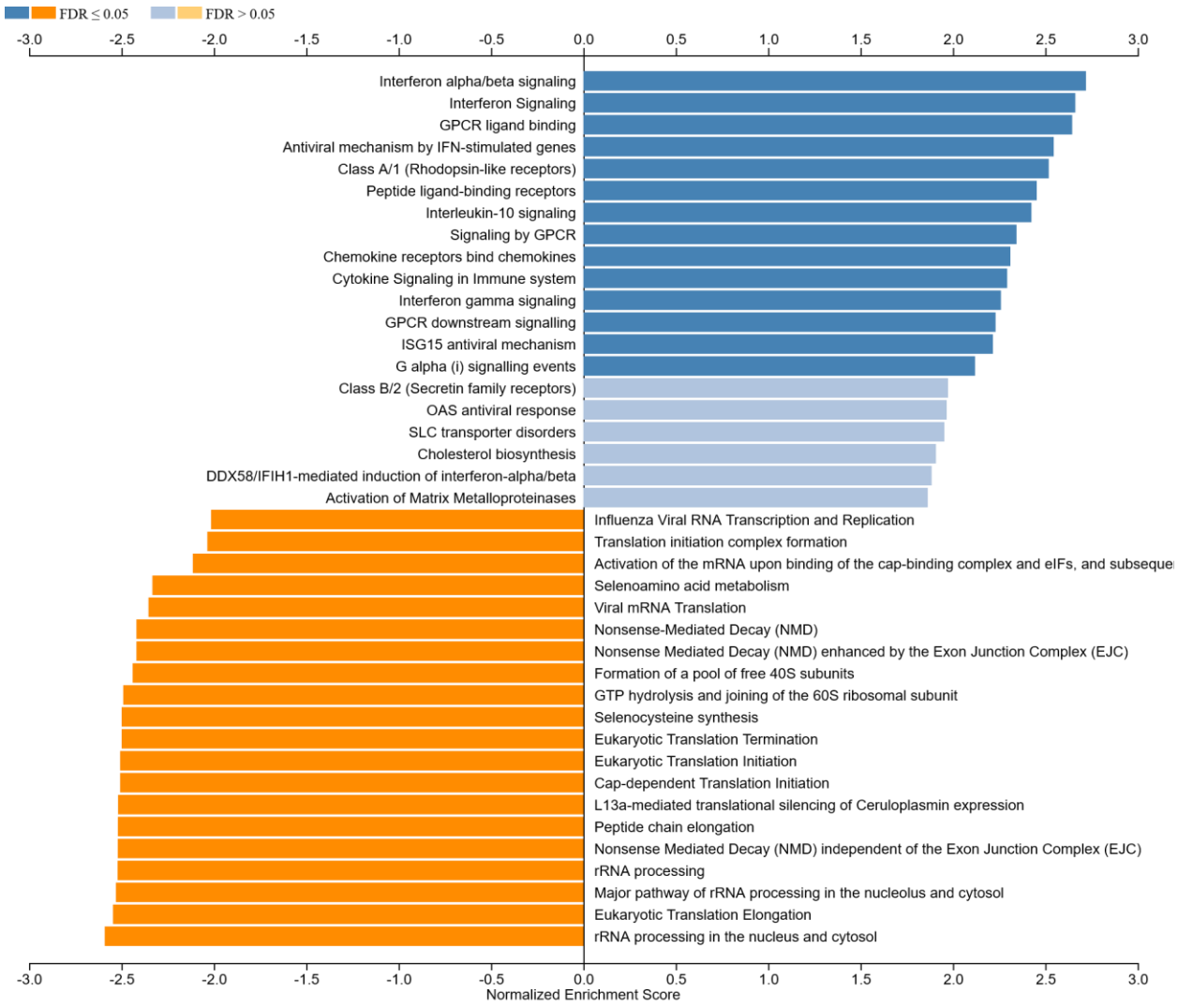


Figure 15 GSEA bar chart with significantly enriched hallmark pathways for 6hpi VSV-GFP infection highlighted (FDR< 0.05)

To go beyond gene sets and pathways and identify key antiviral genes involved in protein-protein interaction (PPI) networks, a Network Topology Analysis was done for the previously identified significantly upregulated genes (264 genes for IAV 24hpi and 235 for rVSV-GFP 6hpi). For all 2 cases (Tables 10,11), in any of the pathways identified such as defence response, viral life cycle,

response to cytokine, interferons, negative regulation of viral genome regulation among others, ISG15 plays a central role thus emerging as an attractive candidate for knockout via CRISPR/Cas9.

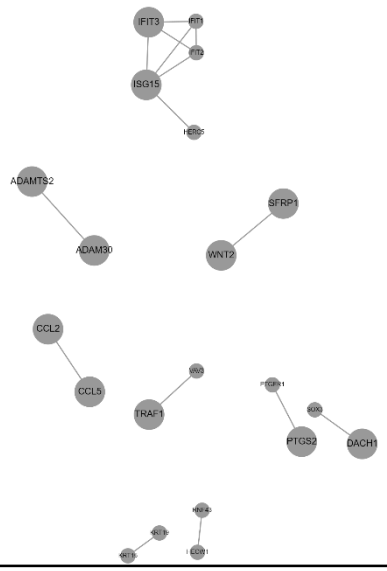
Sub-network Layout	Pathway GO ID	Pathway GO Name	Top ranking associated Genes
	GO:0009615	response to virus	CCL5, ISG15, IFIT1, IFIT2, IFIT3, HERC5
	GO:0019079	viral genome replication	CCL2, CCL5, ISG15, IFIT1
	GO:0034340	response to type I interferon	ISG15, IFIT1, IFIT2, IFIT3
	GO:0045071	negative regulation of viral genome replication	CCL5, ISG15, IFIT1
	GO:0051607	defense response to virus	ISG15, IFIT1, IFIT2, IFIT3, HERC5
	GO:0071345	cellular response to cytokine stimulus	CCL2, CCL5, ISG15, IFIT1, IFIT2, IFIT3, TRAF1, SFRP1, PTGS2

Table 10 Key upregulated networks and their top associated genes for IVA PR8 infection 24hpi (NTA)

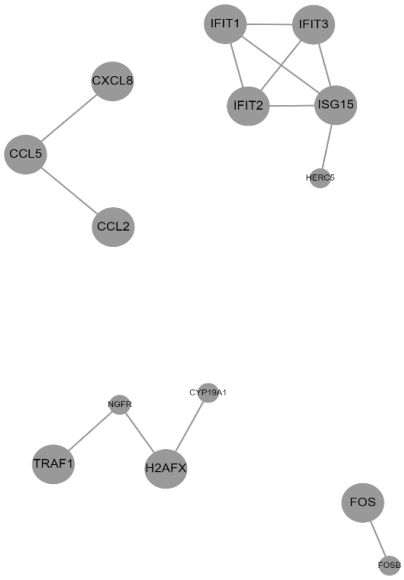
Sub-network Layout	Pathway GO ID	Pathway GO Name	Top ranking associated Genes
	GO:0006952	defense response	ISG15, IFIT1, IFIT2, IFIT3, HERC5, CCL2, CCL5, CXCL8, FOS, CYP19A1
	GO:0009615	response to virus	ISG15, IFIT1, CCL2, CCL5, CXCL8
	GO:0019079	viral genome replication	ISG15, IFIT1, CCL2, CCL5, CXCL8
	GO:0034097	response to cytokine	ISG15, IFIT1, IFIT2, IFIT3, CCL2, CCL5, CXCL8, FOS, TRAF1
	GO:0034340	response to type I interferon	ISG15, IFIT1, IFIT2, IFIT3
	GO:0045069	regulation of viral genome replication	ISG15, IFIT1, CCL5, CXCL8
	GO:0051607	defense response to virus	ISG15, IFIT1, IFIT2, IFIT3, HERC5

Table 11. Key upregulated networks and their top associated genes for rVSV-GFP infection 6hpi (NTA)

Indeed, following these results, extensive literature search showed that ISG15 is a 17kDa antiviral protein (15kDa after maturation[31]) that protects the host via the inhibition of viral replication in a conjugation-dependent manner and is implicated in antiviral responses to various viruses such as SARS-CoV, Influenza virus, HIV, Hepatitis virus among others.

By conjugating host and viral proteins via ISGylation, ISG15 was reported to enhance pathogenesis, inhibit nuclear translocation, budding and release of viral particles, impede viral RNA synthesis

and viral protein translation, decrease infectivity of produced viral particles and suppress viral growth[32].

A functional diversity across species was also previously reported with ISG15-deficient patients showing no increased viral yield following viral infection compared to ISG15-deficient mice. Thus, looking further into that diversity especially for Vero cells to ensure that gene editing of ISG15 will lead to phenotypic modifications with regards to viral infection, the ISG15 protein sequences were compared (Figure 16) between species of interest (i.e. from which cell lines used in vaccine production are derived, such as HEK293 and MDCK). Mutations between human ISG15 and mice ISG15 were like those between human ISG15 and Vero ISG15, especially at position 89 which was previously highlighted as a key player in the ability of Old-World Monkey ISG15 (including Vero cells) to more efficiently ISGylate proteins compared to human ISG15[33], thus, giving some indications concerning the desired effects of ISG15 deletion in Vero cells.

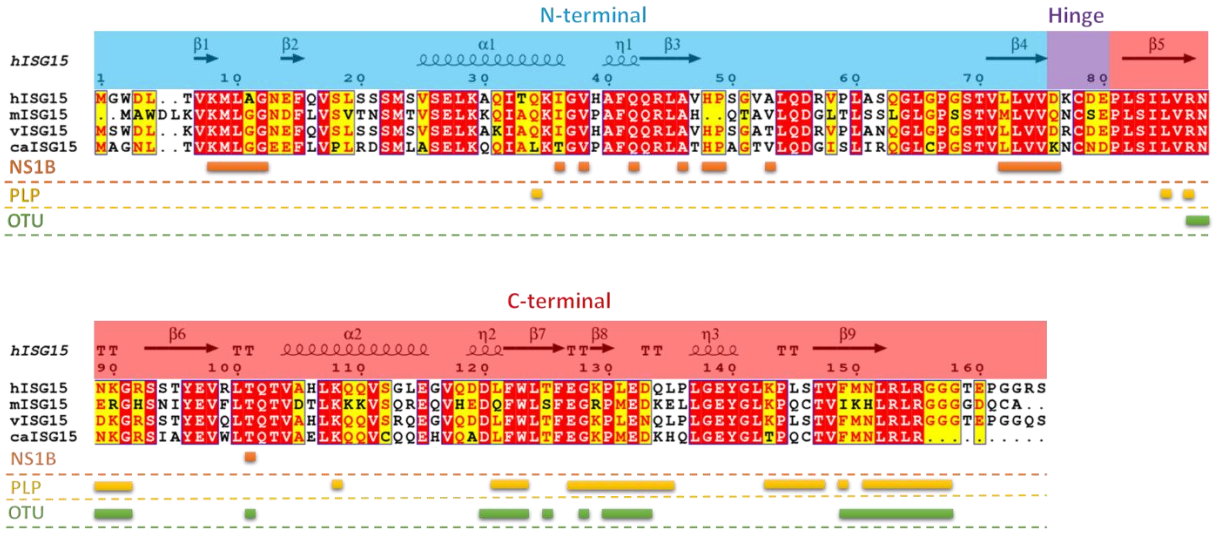


Figure 16 Comparison of ISG15 protein sequences across species

A sequence alignment of human ISG15 (hISG15), mouse ISG15 (mISG15), vero ISG15 (vISG15) canine ISG15 (caISG15). The residues of ISG15 known to interact with the influenza NS1 protein, coronavirus PLPs, and nairovirus OTUs are indicated[34].

3.3 ISG15 deletion validation at the genomic/proteomic level and virus production quantification

Following CRISPR/Cas9-based genomic deletion of ISG15 CDS region, several validation steps were designed to confirm plasmid delivery using a GFP reporter and cell sorting, intended deletion using PCR, protein deletion using western blot and deletion phenotypic effects via viral infection and virus production quantification.

To confirm the genomic deletion, two pairs of PCR primers were designed (Figure 17) (on flanking inside the deletion region and one outside) to screen for deletion bands and non-deletion bands (Figure 18) and among 100 clones screened, 6 were identified with a biallelic deletion and good fitness (via monitoring of the clones doubling time).

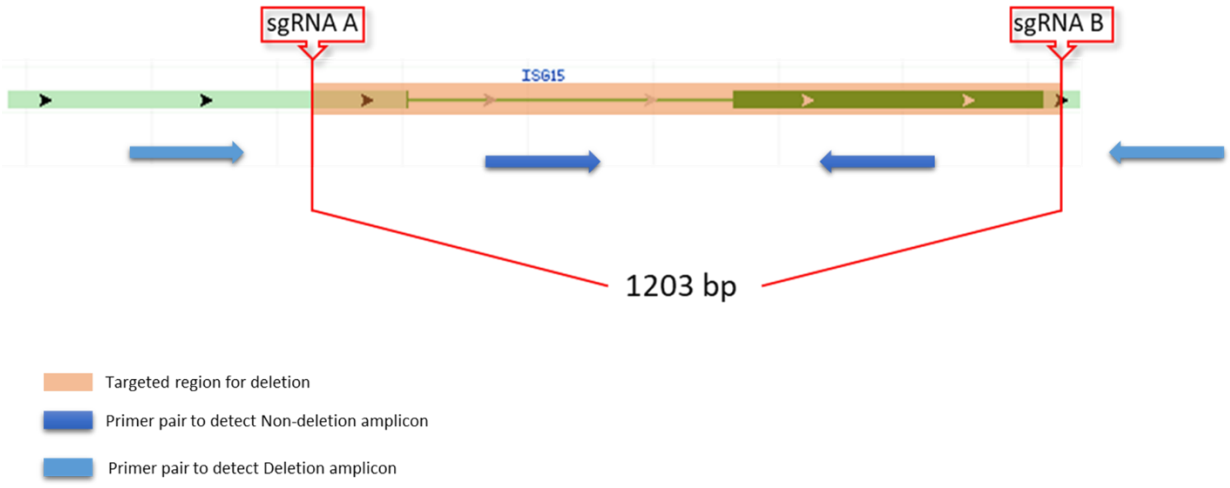


Figure 17 PCR primers design

For the detection of non-deletion and deletion bands.

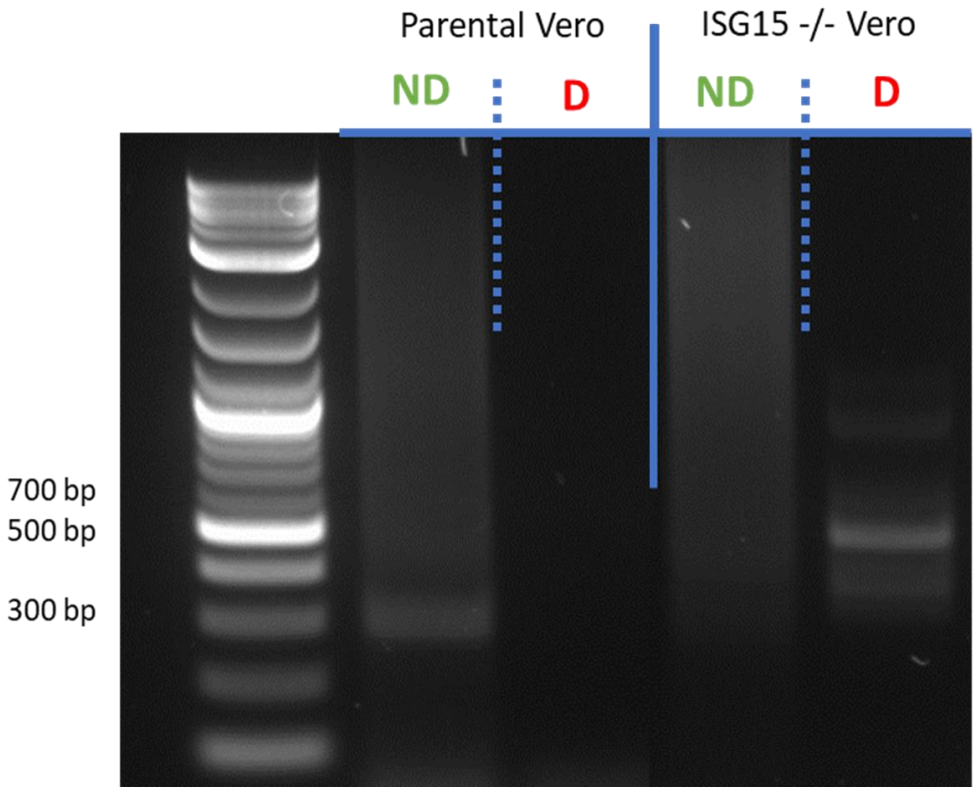


Figure 18 PCR Screening of clones with biallelic deletions of ISG15 CDS region

Biallelic clones selected based on the absence of non-deletion band and the presence of deletion band.

At the protein level, western blot showed that none of the previously selected clones had a band at 15-17kDa which was visible in parental or wild type Vero cells.

Infection of the engineered clone (Figure 19) showed a significant increase not only in total viral particle production but also in infectious viral particles. Indeed, an increase of 70.3-fold of total viral particles was observed for IAV infection and an increase of 87-fold was shown for rVSV-GFP. Interestingly, the ratio of infectious viral particles/total viral particles also significantly increased from 0.0316 to 0.653 for IAV and from 0.0542 to 0.679 for rVSV-GFP.

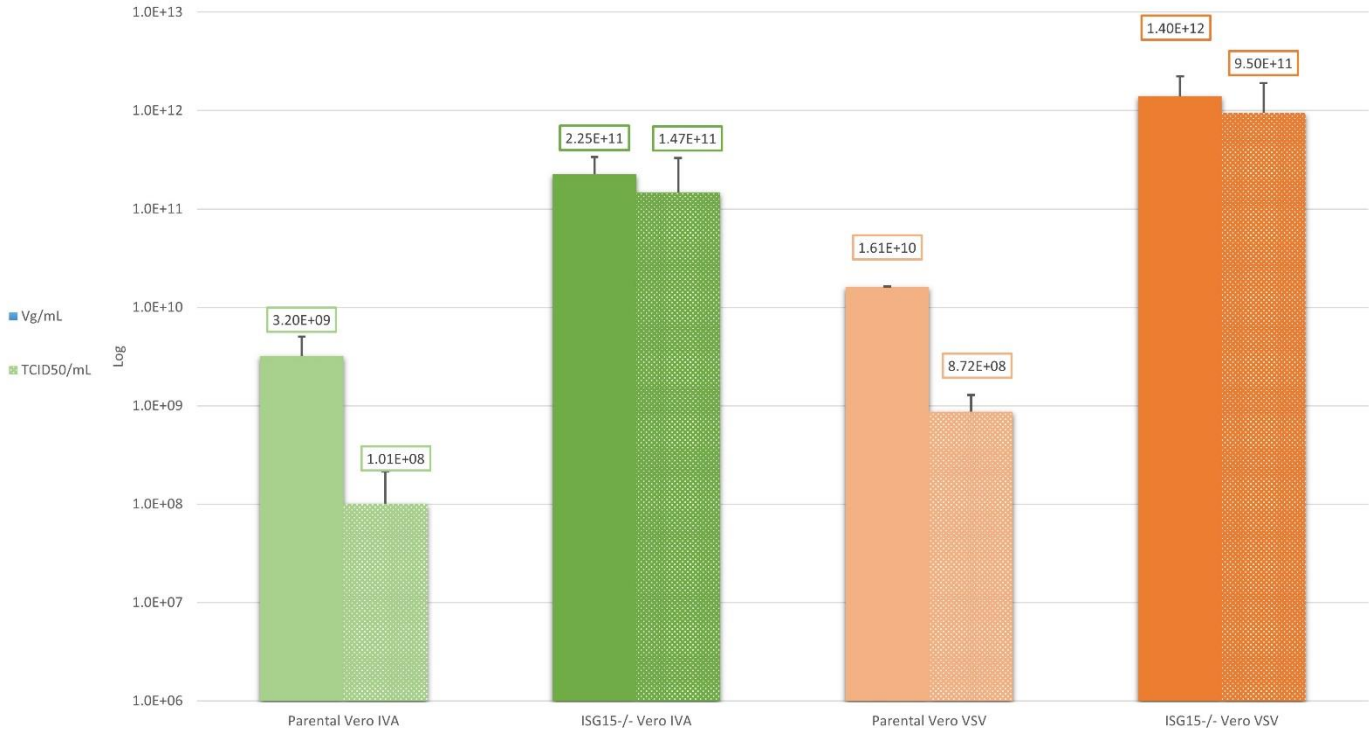


Figure 19 Effect of ISG15 deletion on Virus Production Rate

Quantification of parental Vero and ISG15-/- Vero cells IVA and rVSV-GFP viral genomes and infectious particles production (Vg: Viral genome).

4.0 DISCUSSION

In this work, we present a detailed analysis of virus-host interactions during IAV and rVSV-GFP infection of Vero cells to better understand the dynamics taking place between the host attempting to minimize the impact of viral infection and the viruses attempting to evade host immune responses. Using a novel approach combining functional genomics and cell biology, we propose a new strategy for a more efficient targeted CRISPR[34] gene editing. Opening new possibilities for Vero cells' establishment as a pandemic-ready high throughput vaccine production platform.

Indeed, previous attempts to generate engineered cell lines for high yield virus production relied mainly on screening data to choose gene targets and then knock them out via CRISPR to verify that the previously reported viral production yield in the screening data is obtained[10]. But, in most cases, the knockout did not lead to a significant viral production yield increase due to several factors. These factors include: (i) knock-down and knockout do not lead to the same phenotypic effects; (ii) up to 4 different guide RNAs are used for screening which increase the possibility of off-target effects that could lead to an increased viral production yield in unknown ways; and (iii) another genome or cell line was used for screening compared to the actual cell line used for targeted gene editing. With the recent publication of the de novo assembled and annotated Vero genome[13] and the use of functional genomics, we were able to reach a deeper understanding of the mechanisms at play during infection and of the gene target candidates before selection thus, giving better control and oversight on the gene-editing experiment.

Notably, the deletion of a whole genomic region[14], here the CDS region, was preferred over a single guide RNA-based cut to increase the probability of getting a biallelic deletion, to ensure that the deletion will lead to the desired loss of function of the targeted gene product but also to make the validation step easier thus, ensuring a rapid and high throughput gene-editing protocol.

Interestingly, our deletion of ISG15 in Vero cells led to, alongside the overall increase in total particles production to up to 87-fold, an increase of infectious particles production ratio from 3.2% to 65.3% for IAV and from 5.4% to 67.9% for rVSV-GFP, which confirm previous reports that ISG15 modulates the released infectious particles ratio while intracellular viral replication remains intact[21]. In addition, during the preliminary study of IAV infection kinetics, it was shown that infectious particles production increases until 24hpi before decreasing dramatically (Figure 12), which was also observed during influenza virus infection in HEK-293 and MDCK cells[35]. Thus, this

newly engineered cell line could present even more attractive advantages particularly to produce live vaccines. Therefore, as future work, it would be interesting to study the kinetics of IAV infection in the engineered ISG15-/- Vero cell line to monitor the effects of ISG15 deletion on infectious particles release especially after 24hpi.

Nonetheless, it is important to note that, while being a valuable candidate for gene editing in cell lines used for vaccines production, it is necessary to be cautious about ISG15's variability across species. Indeed, ISG15 knockout showed no effects in human cells while showing increased viral production yield in mice and Vero cells for instance. Thus, in the case of HEK293 cells, given that they are derived from human cells, before applying the ISG15 deletion protocol presented in this paper to HEK293 cells, it is necessary to at least investigate the ISG15 protein sequence of HEK293 to verify the mutations and their similarity with mice and Vero cells to have a rough idea of the possible effects of ISG15 deletion in HEK293 cells. Similarly, MDCK cells being derived from canine, the corresponding protein sequence was also compared to the ISG15 protein from Vero cell, mice and human with key mutation located at sites known to interact with viruses (Figure 16) which increases the chances of successful gene knockout. But again, as shown with Vero cells which present significant differences with the African Green Monkey genome[13], it is important to verify that, indeed, the MDCK ISG15 protein sequence is like Canis lupus familiaris ISG15 at least in the regions known to interact with viruses.

Overall, the Vero cell line was one of the first cell lines considered as a vaccine production platform with various vaccines produced in Vero cells being approved and used to immunize large populations including in the current COVID-19 pandemic. With this new approach, we successfully engineered a cell line capable of increasing virus production yield to up to 87-fold while also increasing infectious viral particles release that was around 3.2-5.4% to up to 65.3-67.9%. Thus,

the key findings of this study open new avenues for the development of pandemic-ready vaccine production platforms in line with the global preparedness plan.

5.0 ACKNOWLEDGEMENTS

This work was supported by funding from the Natural Sciences and Engineering Research Council (NSERC-Grant-CRDPJ 511957). M-A-S is supported by the McGill Engineering Doctoral Award (MEDA) and A.A.K by a Canada Research Chair (CRC-240394).

The authors would like to thank Denis Flipo (UQAM) for assistance with FACS sorting and Shantoshini Dash (McGill University) for performing the western blot.

6.0 COMPETING INTERESTS STATEMENT

The authors declare that they have no financial or nonfinancial competing interests.

7.0 DATA AVAILABILITY STATEMENT

All relevant data are available upon request.

8.0 CODE AVAILABILITY STATEMENT

This study did not make use of custom code.

9.0 REFERENCES

1. Sascha Kiesslich, Amine A. Kamen, Vero cell upstream bioprocess development for the production of viral vectors and vaccines, *Biotech Adv.*, Volume 44, 2020, 107608, ISSN 0734-9750.

2. Barrett PN, Mundt W, Kistner O, Howard MK. Vero cell platform in vaccine production: moving towards cell culture-based viral vaccines. *Exp review of vacc*. 2009 May 1;8(5): 607-18.

3. Ammerman NC, Beier-Sexton M, Azad AF. Growth and maintenance of Vero cell lines. *Curr Protoc Microbiol*. 2008; Appendix 4:Appendix 4E.

4. Suder E, Furuyama W, Feldmann H, Marzi A, de Wit E. The vesicular stomatitis virus-based Ebola virus vaccine: From concept to clinical trials. *Hum Vaccin Immunother*. 2018 ;14(9):2107-2113.

5. Agnandji ST, Huttner A, Zinser ME, et al. Phase 1 Trials of rVSV Ebola Vaccine in Africa and Europe. *N Engl J Med*. 2016;374(17):1647-1660. doi:10.1056/NEJMoa1502924

6. Kiesslich S, Vila-Chã Losa JP, Gélinas JF, Kamen AA. Serum-free production of rVSV-ZEBOV in Vero cells: Microcarrier bioreactor versus scale-X™ hydro fixed-bed. *J Biotechnol*. 2020; 310:32-39. doi: 10.1016/j.jbiotec.2020.01.015

7. Liu et al. A recombinant VSV-vectored MERS-CoV vaccine induces neutralizing antibody and T cell responses in rhesus monkeys after single dose immunization. Antiviral Res. 2018 Feb;150:30-38.

8. Hoffmann et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Mar 4. pii: S0092-8674(20)30229-4.

9. COVID-19 vaccine tracker and landscape : https://www.who.int/teams/blueprint/covid-19/covid-19-vaccine-tracker-and-landscape

10. Van der Sanden SM, Wu W, Dybdahl-Sissoko N, et al. Engineering Enhanced Vaccine Cell Lines To Eradicate Vaccine-Preventable Diseases: the Polio End Game. J Virol. 2016;90(4):1694-704.

11. Hoeksema F, Karpilow J, Luitjens A, et al. Enhancing viral vaccine production using engineered knockout vero cell lines - A second look. Vaccine. 2018;36(16):2093-2103. doi:10.1016/j.vaccine.2018.03.010

12. Bunnik EM, Le Roch KG. An Introduction to Functional Genomics and Systems Biology. *Adv Wound Care (New Rochelle)*. 2013;2(9):490-498. doi:10.1089/wound.2012.0379

13. Sène MA, Kiesslich S, Djambazian H, Ragoussis J, Xia Y, Kamen AA. Haplotype-resolved de novo assembly of the Vero cell line genome. NPJ Vaccines. 2021 Aug 20;6(1):106. doi: 10.1038/s41541-021-00358-9. PMID: 34417462.

14. Bauer DE, Canver MC, Orkin SH. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J Vis Exp*. 2015;(95):e52118. Published 2015 Jan 3. doi:10.3791/52118

15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR, STAR: ultrafast universal RNA-seq aligner, Bioinformatics, Volume 29, Issue 1, January 2013, (pp. 15–21).

16. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009 Aug;25(16):2078-2079.

17. Liao Y, Smyth GK, Shi W, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics, Volume 30, Issue 7, 1 April 2014, (pp. 923–930).

18. Love M.I., Huber W., Anders S. (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, 15, 550. doi: 10.1186/s13059-014-0550-8.

19. Liao Y., Wang J., Jaehnig E. J., Shi Z., Zhang B., WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs, Nucleic Acids Research, Volume 47, Issue W1, 02 July 2019, (pp. W199–W205).

20. Wang J, Ma Z, Carr SA, et al. Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction. *Mol Cell Proteomics*. 2017;16(1):121-134. doi:10.1074/mcp.M116.060301

21. Stark C. et al. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. (2006) 34:D535-539.

22. Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32(Database issue):D258-D261. doi:10.1093/nar/gkh036

23. Di Tommaso P, Moretti S, Xenarios I, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology

extension. *Nucleic Acids Res*. 2011;39(Web Server issue):W13-W17. doi:10.1093/nar/gkr245

24. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(Web Server issue):W320-W324. doi:10.1093/nar/gku316

25. Concordet JP, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res*. 2018;46(W1):W242-W245. doi:10.1093/nar/gky354

26. Peng D, Tarleton R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. *Microb Genom*. 2015;1(4):e000033. Published 2015 Oct 30. doi:10.1099/mgen.0.000033

27. Dzimianski, John & Scholte, Florine & Bergeron, Eric & Pegan, Scott. (2019). ISG15: it's Complicated. *Journl of Mol Biol*. 431. 10.1016/j.jmb.2019.03.013.

28. Fulber JPC, Farnós O, Kiesslich S, et al. Process Development for Newcastle Disease Virus-Vectored Vaccines in Serum-Free Vero Cell Suspension Cultures. *Vaccines (Basel)*. 2021;9(11):1335. doi:10.3390/vaccines9111335

29. Croft D, Mundo AF, Haw R, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2014;42(Database issue):D472-D477. doi:10.1093/nar/gkt1102

30. Guillin OM, Vindry C, Ohlmann T, Chavatte L. Selenium, Selenoproteins and Viral Infection. *Nutrients*. 2019;11(9):2101. Published 2019 Sep 4. doi:10.3390/nu11092101

31. Popp MW, Cho H, Maquat LE. Viral subversion of nonsense-mediated mRNA decay. *RNA*. 2020;26(11):1509-1518. doi:10.1261/rna.076687.120

32. Perng YC, Lenschow DJ. ISG15 in antiviral immunity and beyond. *Nat Rev Microbiol*. 2018;16(7):423-439. doi:10.1038/s41579-018-0020-5

33. Pattyn E, Verhee A, Uyttendaele I, et al. HyperISGylation of Old World monkey ISG15 in human cells. *PLoS One*. 2008;3(6):e2427. Published 2008 Jun 18. doi:10.1371/journal.pone.0002427

34. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9,** 1911 (2018).

35. Perng YC, Lenschow DJ. ISG15 in antiviral immunity and beyond. *Nat Rev Microbiol*. 2018;16(7):423-439. doi:10.1038/s41579-018-0020-5

36. Frensing T, Kupke SY, Bachmann M, Fritzsche S, Gallo-Ramirez LE, Reichl U. Influenza virus intracellular replication dynamics, release kinetics, and particle morphology during propagation in MDCK cells. *Appl Microbiol Biotechnol*. 2016;100(16):7181-7192. doi:10.1007/s00253-016-7542-4

AUTHOR CONTRIBUTION STATEMENT

MAS designed all experiments, executed all experiments, and prepared the manuscript. YX, and AAK provided guidance, supervision, and critical reading of the manuscript.

Chapter 6: General Discussion & Conclusion

In this thesis we present a project focusing mostly on the field of vaccine bioprocessing and more precisely the engineering of the Vero cell line as high throughput viral vaccine manufacturing platform. Indeed, we used genetic engineering methods to optimize the viral production in Vero cells to bypass the previous low virus production and limited scalability possibilities previously observed with this cell line.

First, vaccines are the method of choice for responding to viral infectious disease by producing either proteins, genetic materials, or weakened viral particles that are delivered to the subject to create an immune response and generate the necessary antibodies for a successful immune protection of the patients.

Previously, the antigens particles needed to produce those vaccines were generated in chicken eggs or even primary cells. First, the use of chicken eggs was faced with some hindrances notably because one chicken egg could produce a limited amount of material needed for vaccine doses thus increasing not only the cost of vaccine production but also the production time. Another concern was the cross-contamination and allergic reactions that can come from material produced in chicken eggs. Following those limitations, primary cell lines seemed to be a good alternative alleviating the limited material production issue but also the contamination, time, and cost of production limitations. But due to the nature of primary cells, their use raised significant ethical concerns thus complicating the overall vaccine production process. Therefore, bioprocess researchers turned their efforts towards continuous cell lines which seems an attractive alternative for vaccine production.

Indeed, given the strict guidelines used by regulatory agencies to recommend vaccine manufacturing platforms only few cell lines were recommended for vaccine manufacturing including the Vero cell line. Indeed, this cell line is the most used continuous cell line for virus isolation, replication, and even vaccine manufacturing with over 40 years of experience. Vero cells became an attractive candidate notably due to their high susceptibility to a wide range of viruses but also their capacity to be infected at high multiplicity of infection (MOI) without inducing the cell death pathway. Nonetheless, production of vaccines in Vero cells faced challenges such as the

low virus productivity. The issue of low productivity could have been alleviated by scaling up vaccine manufacturing capacities to increase the amount of Vero cells in culture. But Vero cells being an adherent cell line, there are limited possibilities for scale up. Although use of microcarriers partially addressed this limitation, the complexity of the process restrained their broad industrial adoption. A very promising alternative is the adaptation of Vero cells to suspension, providing a cost effective and high-throughput volumetric scale up. Another complementary strategy is to increase the cell-specific viral production yield by editing the cell line genome to reduce its antiviral response. Both avenues have been explored over decades.

Chapter 2 presents a literature overview of the efforts made to shape Vero cells into the go-to vaccine manufacturing cell-based platform like its antibody production platform counterpart (CHO cells) by focusing mostly on the genomics aspects of those efforts. Looking at the suspension adaptation, through medium formulation, several groups achieved partial adaptation to suspension and most recently Chen and al. [Chapter 4-11] generated a suspension Vero cell line that successfully passed the bioprocessing proof of concept. But even that cell line had still some limitations notably the aggregation of cells, the high doubling time, and the low cell density. On the other hand, to increase the virus production in Vero cells, genetic engineering avenues using CRISPR were explored. However, those attempts were not successful mainly since a comprehensive contiguous Vero cell genome was not available at the time. Rather, the African Green Monkey genome was used in some attempts since Vero cells are derived from a female African Green Monkey kidney.

This chapter highlighted the necessity to generate a reference-grade genome assembly for Vero cells to not only better understand the adaptation to suspension and its hindrances for a more successful suspension cell line production, but also to have the tools and the understanding of host-virus interactions necessary for a successful genetic engineering to generate a high-throughput high-yield vaccine production cell line.

In this context, Chapter 3 presents the work done to generate that reference-grade genome. Indeed, we assembled a highly contiguous genome , thus providing comprehensive list of the Vero cell genomic, transcriptomic, and proteomic components. Through this assembly, we unraveled significant characteristics that helped us better understand the peculiar nature of Vero cells. By

comparing the Vero genome to the African green monkey genome, we were able to identify significant genomic variants leading to the loss of function (LOF) of hundreds of genes, notably genes involved in the cell death pathway activation and the cytokine pathway for instance, thus highlighting one of the reasons behind the high susceptibility of Vero cells to various viruses and its ability to be infected at high MOI without triggering the cell death pathway.

Among those genes which lost their function, several proviral genes were identified, explaining the low productivity of certain viruses such as Influenza in Vero cells but also opening new implication for the engineering of Vero cell-based manufacturing platforms.

Indeed, viral production of some viruses such as Influenza can be increased through genetic engineering by targeting those LOF genes that led to the low viral production yields.

Moreover, in the context of the current COVID-19 pandemic, Vero cells came into the spotlight as being the cell line with the highest susceptibility to SARS-CoV-2 with the reported entry route being ACE2 despite other entry routes coming to light with new investigations. In our study, we showed that ACE2 was also listed as a gene having lost its function due to structural variants. The partial loss of function was demonstrated with an absence of enzymatic activity of ACE2 in Vero cells and further investigations need to be done in an appropriate safety level facility to fully study Vero cells interactions with SARS-CoV-2 to better understand Vero cells characteristics when it comes to coronavirus infections. In addition, as demonstrated in this thesis with ISG15 for increased viral production, a genomic deletion strategy of the defective ACE2 in Vero cells can be considered to assess its effects on the cell line susceptibility to coronaviruses, going as far as inserting a fully functional ACE2 gene to observe its effects and run a comparison study.

Furthermore, through the study of this newly generated Vero cell genome, we discovered viral genomic sequences inserted not only in the Vero genome but also in the African Green Monkey genome, notably the complete genomes of several AAV serotypes. In addition to the regulatory aspects, given that AAV is an important viral vector for key applications such as gene delivery, developing novel strategies for producing functional recombinant AAV particles at high yields will be an important contribution to the field of AAV biology and potential application in gene therapy. Therefore, by further studying the inserted complete genomes of AAV serotypes in Vero cells and their characteristics, it might be possible to generate tailored Vero cell sublines specific to high yield production of AAV.

In summary, this chapter shows the significant potential of Vero cells to become a highly valued

vaccine manufacturing and eventually gene therapy manufacturing platform, through the generation of a highly continuous haplotype-resolved reference-grade genome, to better understand the genetic identity of this cell line.

Notably, with this newly generated genetic information and its rich annotation, there are now new possibilities to further study Vero cells through genomic, transcriptomic, proteomic, and even metabolomic lenses via the use of functional genomics to better understand the phenotypic effects triggered during Vero cells adaptation to suspension or during viral infection of Vero cells, for example.

In Chapter 4, we investigated these new possibilities by applying transcriptomics statistical tools to RNA data generated for Vero cells in adherent state and in suspension state. This process helped better understanding the pathways at play during Vero cells adaptation to suspension and provide tailored solutions to the hindrances previously reported such as the high doubling time, cells aggregation and low cell density.

To achieve a higher level of understanding, we combined several functional genomics tools that complement each other in a synergistic fashion, namely, gene expression, pathway enrichment analysis, metabolic pathway analysis, gene set enrichment analysis and network topology assay.
Indeed, several events of checks-and-balances were consistently observed across the different analytical methodologies used. For instance, despite the accumulation of stress caused by the upregulation of the gluconeogenesis pathway and the downregulation of glycolytic pathway causing a hindered ATP generation, Vero cells adapting to suspension have their proline metabolic pathway downregulated promoting proline accumulation, and their asparagine metabolic pathway upregulated thus promoting the cell survival.
Some counter intuitive events were also observed:

- Cell adhesion related genes are upregulated during the process of adaptation to suspension of Vero cells but also of HEK293 cells as previously reported by Malm et al. [Chapter 4, Reference 20]. This upregulation, even though counter-intuitive at first glance, explains the cells aggregates observed in suspension cultures and the necessity to reduce the clumping of the cells in culture and are mainly due to the cells' attempt to restore their natural attachments to culture surfaces.

- In the attempts to generate suspension cell lines, the epithelial to mesenchymal transition (EMT) pathway was previously described and studied as an avenue to explore for a successful adaption of adherent cells to suspension. But our study and previous study [Chapter 4, Reference 20] on HEK293 showed a downregulation of that particular pathway thus dissociating the EMT event and the adaptation to suspension process.

The cause of the low cell density and long doubling time were also found via the downregulation of the adherens junction pathway, which is essential for cell viability, and cell cycle related pathways and genes. Even though the cells attempt to counter that long doubling time by upregulating metabolic pathways and genes associated with unrestrained cell cycle progression. Thus, this chapter not only provides some insights in the overall adaptation of Vero cells to suspension at the phenotypic level, but also key genes associated with cell adhesion or hindering cell viability or doubling time are highlighted as potential targets for gene editing to achieve a successful adaptation of Vero cells to suspension. Furthermore, the different checks-and-balances events observed between competing pathways could be further analyzed using perturbation-based methods and genomic libraries that can be generated from the reference-grade Vero genome presented in the previous chapter of this thesis. Also, the insights in the metabolic process taking place open new avenues with regards to medium formulation and chemical combinations that could be used for a better adaption to suspension. Overall, this study on Vero cells adaptation to suspension has the potential to pave the way for a successful adaptation thus providing valuable technology transfer opportunities for developing countries. The current COVID-19 pandemic highlighted the necessity for all countries, regardless of their development stage, to have the tools necessary to counter the spread of SARS-CoV-2 by notably having enough vaccine stocks to immunize their population. Thus, providing developing countries with the tools necessary to have their own vaccine production facilities is a great step towards a successful global response to pandemics.

Not only were efforts made to adapt Vero cells to suspension, also, as shown in Chapter 2, studies were carried to increase the cell-specific viral production yield via genetic screen and CRISPR/Cas9 gene editing which had limited success in prior studies.

Building on the functional genomic approach used in Chapter 4, Chapter 5 covers the application of those statistical methods to study the host-virus interactions during Vero cells infection with

Influenza A (IVA) PR8 and rVSV-GFP. Through this interdisciplinary approach, we propose a more efficient gene target selection process through combination of several statistical tools and intensive literature search on the gene background and the previously attempted gene editing leading to the successful production of a high yield virus production Vero subline.

First, the detection of the ISG15 pathway in the different statistical methods used and at the gene, pathway, gene set and even the protein-protein interaction pathway levels led to a deep metadata investigation of this gene and its characteristics using evidence from the literature. ISG15 surfaced as an attractive candidate given its important role in the host antiviral response, even though it presented significant variations from one species to the other thus explaining the discrepancies when studying the effects of knockout of this gene in human or mice. A systematic comparison of the ISG15 protein sequences for human, mice, Vero and even *canis lupus* further confirm that the knockout in Vero cells would lead to a successful outcome.

On the other hand, a whole genomic region deletion using pairs of guide-RNAs was preferred to single guide-RNA cut (that could randomly lead to loss of function structural variants) for easier validation steps and to ensure that the desired deletion occurred by increasing the probability of getting a biallelic deletion.

As predicted, not only did ISG15 coding sequence deletion led to an increase in virus production yield of up to 87-fold, but also the infectious particles production ratio was increased from 3.16% to 65.3% for IVA and from 5.42% to 67.9% for rVSV-GFP.

But, before applying this genomic deletion of ISG15 to other continuous cell lines used in the field of vaccine manufacturing and gene therapy such as HEK293 or MDCK cell lines, it is important to consider the variations in ISG15 protein sequences and especially at the regions known to interact with viruses to ensure a successful cell engineering experiment.

In the case of Influenza, as it was previously reported with MDCK cells, our preliminary kinetics study showed that infectious particles production increases until 24hpi before decreasing drastically. Thus, given the significant increase in the infectious particles production ratio after genetic engineering, it would be interesting to conduct another kinetic study with this new cell line to see how the previously reported decrease after 24h is affected by the deletion of ISG15.

Furthermore, given the importance of producing infectious viral particles in quantity for field such as gene therapy and live-attenuated vaccines manufacturing, applying the overall protocol presented in Chapter 5 to the production of other viral vectors and vaccines could be another

highly relevant direction for future work.

The work presented in this thesis, contributed to the field in different aspects. First, a reference grade type of genome was generated, providing a new genomic tool for better understanding of the Vero cell line but also foundation for other studies related to this cell line. Also, the significant viral insertions identified during the building of this reference genome, highlight new possibilities for quality control using viral sequence clearance through gene editing and ensure the detection and purification of any undesirable by-products during vaccine manufacturing.

Moreover, the transcriptomic characterization of Vero cells adaptation suspension helped understand the limitations observed and propose some new avenues that can be explored to further improve the adaptation to suspension process providing valuable information to the field of cell-based vaccine manufacturing.

This work also proposes a new tool for the design of more efficient genome analysis and cell engineering strategy by giving a first glance into the heterogeneity of this cell line by building a haplotype-resolved reference genome but also, by proposing a targeted approach for genome editing not relying on the conventional perturbation-based screening approaches but relying mostly on interdisciplinary approach using functional genomics on one side and associating it novel gene editing strategies.

Ultimately, this thesis and this overall project highlights the versatility and significant potential of Vero cells with regard to its use in the global effort to contain current and emerging pathogens via the isolation and replication of viruses but also the manufacturing of vaccines and gene therapy tools among others.

Literature

Yasumura, Y., Kawakita, Y., 1963. The research for the SV40 by means of tissue culture technique. Nippon rinsho 21, 1201–1219.

Emeny JM, Morgan MJ. Regulation of the interferon system: evidence that Vero cells have a genetic defect in interferon production. *J Gen Virol*. 1979;43(1):247-252. doi:10.1099/0022-1317-43-1-247

World Health Organization, 1987a. Acceptability of cells substrates for production of biologicals. WHO Tech. Rep. Ser. 747, 5–24.

Barrett PN, Mundt W, Kistner O, Howard MK. Vero cell platform in vaccine production: moving towards cell culture-based viral vaccines. *Exp review of vacc*. 2009 May 1;8(5): 607-18.

Ammerman NC, Beier-Sexton M, Azad AF. Growth and maintenance of Vero cell lines. *Curr Protoc Microbiol*. 2008;Appendix 4:Appendix 4E.

Suder E, Furuyama W, Feldmann H, Marzi A, de Wit E. The vesicular stomatitis virus-based Ebola virus vaccine: From concept to clinical trials. *Hum Vaccin Immunother*. 2018;14(9):2107-2113.

Agnandji ST, Huttner A, Zinser ME, et al. Phase 1 Trials of rVSV Ebola Vaccine in Africa and Europe. *N Engl J Med*. 2016;374(17):1647-1660. doi:10.1056/NEJMoa1502924

COVID-19 vaccine tracker and landscape : https://www.who.int/teams/blueprint/covid-19/covid-19-vaccine-tracker-and-landscape

Lee JS, Grav LM, Lewis NE, Kildegaard HF. CRISPR/Cas9 Mediated Genome Engineering of CHO Cell Factories: Application and Perspectives. Biotechnol J 10 (7): 979-94.

Lee JS, Kallehauge TB, Pedersen LE, Kildegaard HF. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. Sci Rep. 2015;5:8572.

Arifin MA, Mel M, Abdul Karim MI, Ideris A. Production of Newcastle disease virus by Vero cells grown on cytodex 1 microcarriers in a 2-litre stirred tank bioreactor. *J Biomed Biotechnol*. 2010;2010:586363. doi:10.1155/2010/586363

Knop DR, Harrell H. Bioreactor production of recombinant herpes simplex virus vectors. *Biotechnol Prog*. 2007;23(3):715-721. doi:10.1021/bp060373p

Genzel Y. Designing cell lines for viral vaccine production: Where do we stand?. *Biotechnol J*. 2015;10(5):728-740. doi:10.1002/biot.201400388

Litwin J. The growth of Vero cells in suspension as cell-aggregates in serum-free media. *Cytotechnology*. 1992;10(2):169-174. doi:10.1007/BF00570893

Dee KU, Shuler ML. Optimization of an assay for baculovirus titer and design of regimens for the synchronous infection of insect cells. *Biotechnol Prog*. 1997;13(1):14-24. doi:10.1021/bp960086t

Paillet C, Forno G, Kratje R, Etcheverrigaray M. Suspension-Vero cell cultures as a platform for viral vaccine production. *Vaccine*. 2009;27(46):6464-6467. doi:10.1016/j.vaccine.2009.06.020

Rourou S, Ben Ayed Y, Trabelsi K, Majoul S, Kallel H. An animal component free medium that promotes the growth of various animal cell lines for the production of viral vaccines. *Vaccine*. 2014;32:2767–2769.

Shen CF, Guilbault C, Li X, et al. Development of suspension adapted Vero cell culture process technology for production of viral vaccines. *Vaccine*. 2019;37:6996–7002.

Jaluria P, Betenbaugh M, Konstantopoulos K, Frank B, Shiloach J. Application of microarrays to identify and characterize genes involved in attachment dependence in HeLa cells. *Metab Eng*. 2007;9(3):241-251. doi:10.1016/j.ymben.2006.12.001

Mehrbod, P., Fotouhi, F., G, I.M., Mohammadpour, A., Farahmand, B., 2015. Comparison of Vero and MDCK cell lines transfected with human siat7e gene for conversion to suspension culture. Vaccine Res. 2, 69–73. 10.18869/acadpub.vacres.2.4.69.

Van der Sanden SM, Wu W, Dybdahl-Sissoko N, et al. Engineering Enhanced Vaccine Cell Lines To Eradicate Vaccine-Preventable Diseases: the Polio End Game. J Virol. 2016;90(4):1694-704.

Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9,** 1911 (2018).

Wu, W., Orr-Burks, N., Karpilow, J. *et al.* Development of improved vaccine cell lines against rotavirus. *Sci Data* **4,** 170021 (2017). https://doi.org/10.1038/sdata.2017.21

Hoeksema F, Karpilow J, Luitjens A, et al. Enhancing viral vaccine production using engineered knockout vero cell lines - A second look. Vaccine. 2018;36(16):2093-2103. doi:10.1016/j.vaccine.2018.03.010

Osada N, Kohara A, Yamaji T, et al. The genome landscape of the african green monkey kidney-derived vero cell line. *DNA Res*. 2014;21(6):673-683. doi:10.1093/dnares/dsu029

Davies SL, Lovelady CS, Grainger RK, Racher AJ, Young RJ, James DC. Functional heterogeneity and heritability in CHO cell populations. Biotechnol Bioeng 2013;110:260–74.

Diaz, M.O., Ziemin, S., Le Beau, M.M., et al. 1988,Homozygous deletion of the a- and b interferon genes in human leukemia and derived cell lines, *Proc.Natl. Acad. Sci*. USA, 85, 5259–63.

Sakuma, C., Sekizuka, T., Kuroda, M. *et al.* Novel endogenous simian retroviral integrations in Vero cells: implications for quality control of a human vaccine cell substrate. *Sci Rep* **8,** 644 (2018).

Sène MA, Kiesslich S, Djambazian H, Ragoussis J, Xia Y, Kamen AA. Haplotype-resolved de novo assembly of the Vero cell line genome. *NPJ Vaccines*. 2021 Aug 20;6(1):106. doi: 10.1038/s41541-021-00358-9. PMID: 34417462.

Sascha Kiesslich, Amine A. Kamen, Vero cell upstream bioprocess development for the production of viral vectors and vaccines, *Biotech Adv.*, Volume 44, 2020, 107608, ISSN 0734-9750.

Liu et al. A recombinant VSV-vectored MERS-CoV vaccine induces neutralizing antibody and T cell responses in rhesus monkeys after single dose immunization. Antiviral Res. 2018 Feb;150:30-38.

Hoffmann et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Mar 4. pii: S0092-8674(20)30229-4.

Draft landscape of COVID-19 candidate vaccines. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines (2020)

Warren WC, Jasinska AJ, García-Pérez R, et al. The genome of the vervet (Chlorocebus aethiops sabaeus). *Genome Res*. 2015;25(12):1921–1933

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics (Oxford, England). 2015 Oct;31(19):3210-3212.

Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics (Oxford, England). 2007 May;23(9):1061-1067.

Nishimura O, Hara Y, Kuraku S. gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics*. 2017;33(22):3635-3637.

Rhie, A., Walenz, B.P., Koren, S. et al. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol 21, 245 (2020). https://doi.org/10.1186/s13059-020-02134-9.

Jason R. Miller, Arthur L. Delcher, Sergey Koren, Eli Venter, Brian P. Walenz, Anushka Brownley, Justin Johnson, Kelvin Li, Clark Mobarry, Granger Sutton. Aggressive assembly of pyrosequencing reads with mates,Bioinformatics 2008, vol. 24 (pg. 2818-2824).

Liu Y, Loewer M, Aluru S, Schmidt B. SNVSniffer: an integrated caller for germline and somatic single-nucleotide and indel mutations. *BMC Syst Biol*. 2016;10 Suppl 2(Suppl 2):47. Published 2016 Aug 1.

Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220-1222.

Zhang Z., Schwartz S., Wagner L., & Miller W. (2000), "A greedy algorithm for aligning DNA sequences" J Comput Biol 2000; 7(1-2):203-14.

Zheng, G.X.; Lau, B.T.; Schnall-Levin, M.; Jarosz, M.; Bell, J.M.; Hindson, C.M.; Kyriazopoulou-Panagiotopoulou, S.; Masquelier, D.A.; Merrill, L.; Terry, J.M. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* **2016**, *34*, 303.

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB (2017). Direct determination of diploid genome sequences. *Genome Research* 27: 757-767 .

Yeo S, Coombe L, Warren RL, Chu J, Birol I. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics*. 2018;34(5):725-731

.

Warren, R.L., Yang, C., Vandervalk, B.P. *et al.* LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *GigaSci* **4,** 35 (2015).

Shaun D. Jackman, Lauren Coombe, Justin Chu, Rene L. Warren, Benjamin P. Vandervalk, Sarah Yeo, Zhuyi Xue, Hamid Mohamadi, Joerg Bohlmann, Steven J.M. Jones and Inanc Birol (2018). Tigmint: correcting assembly errors using linked reads from large molecules. *BMC Bioinformatics*, 19(1). doi:10.1186/s12859-018-2425-6.

Warren RL. 2016. RAILS and Cobbler: Scaffolding and automated finishing of draft genomes using long DNA sequences. *The Journal of Open Source Software*.

René L Warren, Lauren Coombe, Hamid Mohamadi, Jessica Zhang, Barry Jaquish, Nathalie Isabel, Steven J M Jones, Jean Bousquet, Joerg Bohlmann, Inanç Birol, ntEdit: scalable genome sequence polishing, *Bioinformatics*, Volume 35, Issue 21, 1 November 2019, Pages 4430–4432.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075.

Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res*. 2005;33(Web Server issue):W465-W467.

Thibaud-Nissen F, Souvorov A, Murphy T, et al. Eukaryotic Genome Annotation Pipeline. 2013 Nov 14. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).

Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009 Aug;25(16):2078-2079.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42(Web Server issue):W187-W191.

Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92.

Dennis, G., Sherman, B.T., Hosack, D.A. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4,** R60 (2003).

Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative

genomics. *Genome Res*. 2009;19(9):1639-1645.

Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993.

Kelley LA et al.The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10, 845-858 (2015).

Reynolds CR, Islam SA, Sternberg MJE.EzMol: A Web Server Wizard for the Rapid Visualization and Image Production of Protein and Nucleic Acid Structures, *J Mol Biol, Vol. 430* (2018).

COVID-19 vaccine tracker and landscape: https://www.who.int/teams/blueprint/covid-19/covid-19-vaccine-tracker-and-landscape

Madeline B, Ribaud S, Xenopoulos A, Simler J, Schwamborn K, Léon A. Culturing a duck es-derived cell line in single-use bioreactors: a rapid, efficient, and cost-effective vaccine manufacturing system based on suspension culture. *Bioprocess Int*, 13 (3) (2015), (pp. 26-33).

Peschel B, Frentzel S, Laske T, Genzel Y, Reichl U. Comparison of influenza virus yields and apoptosis-induction in an adherent and a suspension MDCK cell line. *Vaccine*, 31 (2013), (pp. 5693-5699).

Yang J, Guertin P, Jia G, Lv Z, Yang H, Ju D. Large-scale microcarrier culture of HEK293T cells and Vero cells in single-use bioreactors. *AMB Exp*, 9 (2019), (pp. 70-83).

Merten OW. Development of serum-free media for cell growth and production of viruses/viral vaccines: safety issues of animal products used in serum-free media. *Dev Biol* (Basel) 2002; 111:233–257.

Chen A, Poh SL, Dietzsch C, Roethl E, Yan ML, Ng SK. Serum-free microcarrier based production of replication deficient influenza vaccine candidate virus lacking NS1 using Vero cells. *BMC Biotechnol*. 2011;11:81.

M. Logan, M. Aucoin. Media formulation to support the growth of Vero cells in suspension. *Proceedings of vaccine technology VII conference*, June 17-22, Mont Tremblant, Quebec, Canada (2018).

Amelio I., Cutruzzolá F., Antonov A., Agostini M., Melino G.. Serine and glycine metabolism in cancer. *Trends Biochem Sci*. 2014;39(4):191-198. doi:10.1016/j.tibs.2014.02.004.

Balasubramanian MN, Butterworth EA, Kilberg MS. Asparagine synthetase: regulation by cell stress and involvement in tumor biology. *Am J Physiol Endocrinol Metab*. 2013;304(8):E789-E799. doi:10.1152/ajpendo.00015.2013.

Nilsson R, Jain M, Madhusudhan N, et al. Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat Commun*. 2014;5:3128. doi:10.1038/ncomms4128.

Neugut, A., & Weinstein, I. (1979). Growth Limitation of BHK-21 Cells and Its Relation to Folate Metabolism. *In Vitro*, 15(5), 363-367. Retrieved March 29, 2021, from http://www.jstor.org/stable/4292213.

Assimacopoulos-Jeannet F, Thumelin S, Roche E, Esser V, McGarry JD, Prentki M. Fatty acids rapidly induce the carnitine palmitoyltransferase I gene in the pancreatic beta-cell line INS-1. *J Biol Chem*. 1997 Jan 17;272(3):1659-64. doi: 10.1074/jbc.272.3.1659. PMID: 8999842.

Liang X, Zhang L, Natarajan SK, Becker DF. Proline mechanisms of stress survival. *Antioxid Redox Signal*. 2013;19(9):998-1011. doi:10.1089/ars.2012.5074.

Malm, M., Saghaleyni, R., Lundqvist, M. et al. Evolution from adherent to suspension: systems biology of HEK293 cell line development. Sci Rep 10, 18996 (2020). doi: 10.1038/s41598-020-76137-8

Ivanov AI, Naydenov NG. Dynamics and regulation of epithelial adherens junctions: recent discoveries and controversies. *Int Rev Cell Mol Biol*. 2013;303:27-99. doi: 10.1016/B978-0-12-407697-6.00002-7. PMID: 23445808.

Yang Y, Li X, Sun Q, He B, Jia Y, Cai D, Zhao R. Folate deprivation induces cell cycle arrest at G0/G1 phase and apoptosis in hippocampal neuron cells through down-regulation of IGF-1 signaling

pathway. *Int J Biochem Cell Biol*. 2016 Oct;79:222-230. doi: 10.1016/j.biocel.2016.08.040. Epub 2016 Aug 31. PMID: 27592453.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics*, Volume 29, Issue 1, January 2013, (pp. 15–21).

Liao Y, Smyth GK, Shi W, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics*, Volume 30, Issue 7, 1 April 2014, (pp. 923–930).

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

Zhou et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* (2019) 10(1):1523.

Damiani, C., Rovida, L., Maspero, D., Sala, I., Rosato, L., Di Filippo, M., Pescini, D., Graudenzi, A., Antoniotti, M., & Mauri, G. (2020). MaREA4Galaxy: Metabolic reaction enrichment analysis and visualization of RNA-seq data within Galaxy. *Computational and Structural Biotechnology Journal*, 18, 993–999.

Liao Y., Wang J., Jaehnig E. J., Shi Z., Zhang B., WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs, *Nucleic Acids Research*, Volume 47, Issue W1, 02 July 2019, (pp. W199–W205).

Stark C. et al. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. (2006) 34:D535-539.

Kiesslich S, Vila-Chã Losa JP, Gélinas JF, Kamen AA. Serum-free production of rVSV-ZEBOV in Vero cells: Microcarrier bioreactor versus scale-X™ hydro fixed-bed. *J Biotechnol*. 2020;310:32-39. doi:10.1016/j.jbiotec.2020.01.015

Bunnik EM, Le Roch KG. An Introduction to Functional Genomics and Systems Biology. *Adv Wound Care (New Rochelle)*. 2013;2(9):490-498. doi:10.1089/wound.2012.0379

Bauer DE, Canver MC, Orkin SH. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J Vis Exp*. 2015;(95):e52118. Published 2015 Jan 3. doi:10.3791/52118

Croft D, Mundo AF, Haw R, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2014;42(Database issue):D472-D477. doi:10.1093/nar/gkt1102

Guillin OM, Vindry C, Ohlmann T, Chavatte L. Selenium, Selenoproteins and Viral Infection. *Nutrients*. 2019;11(9):2101. Published 2019 Sep 4. doi:10.3390/nu11092101

Popp MW, Cho H, Maquat LE. Viral subversion of nonsense-mediated mRNA decay. *RNA*. 2020;26(11):1509-1518. doi:10.1261/rna.076687.120

Pattyn E, Verhee A, Uyttendaele I, et al. HyperISGylation of Old World monkey ISG15 in human cells. *PLoS One*. 2008;3(6):e2427. Published 2008 Jun 18. doi:10.1371/journal.pone.0002427

Perng YC, Lenschow DJ. ISG15 in antiviral immunity and beyond. *Nat Rev Microbiol*. 2018;16(7):423-439. doi:10.1038/s41579-018-0020-5

Frensing T, Kupke SY, Bachmann M, Fritzsche S, Gallo-Ramirez LE, Reichl U. Influenza virus intracellular replication dynamics, release kinetics, and particle morphology during propagation in MDCK cells. *Appl Microbiol Biotechnol*. 2016;100(16):7181-7192. doi:10.1007/s00253-016-7542-4

Wang J, Ma Z, Carr SA, et al. Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction. *Mol Cell Proteomics*. 2017;16(1):121-134. doi:10.1074/mcp.M116.060301

Harris MA, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 2004;32(Database issue):D258-D261. doi:10.1093/nar/gkh036

Di Tommaso P, Moretti S, Xenarios I, et al. T-Coffee: a web server for the multiple sequence

alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*. 2011;39(Web Server issue):W13-W17. doi:10.1093/nar/gkr245

Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(Web Server issue):W320-W324. doi:10.1093/nar/gku316

Concordet JP, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res*. 2018;46(W1):W242-W245. doi:10.1093/nar/gky354

Peng D, Tarleton R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. *Microb Genom*. 2015;1(4):e000033. Published 2015 Oct 30. doi:10.1099/mgen.0.000033

Dzimianski, John & Scholte, Florine & Bergeron, Eric & Pegan, Scott. (2019). ISG15: it's Complicated. *Journl of Mol Biol*. 431. 10.1016/j.jmb.2019.03.013.