POOLED GENOME-WIDE CRISPR/Cas9 SCREENING AS A TOOL TO INTENSIFY THE YIELD OF CELL BASED INFLUENZA VACCINES AND OTHER VIRAL BIOLOGICS

Author: David Michael Sharon

Academic unit: Biological and Biomedical Engineering Program

Institution: McGill University, Montreal

Submission date: April 14, 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Biological and Biomedical Engineering.

CONTENTS

ABSTRACT (ENGLISH)

Viral biologics such as vaccines and gene transfer vectors are dominantly and increasingly produced in immortalized cell lines. Genetic modification of the host cell line may offer a means to increase viral yield, intensifying production for a wide range of established and emerging therapeutics. In this thesis I explore the use of genome-wide, pooled CRISPR/Cas/9 screening as a tool to identify targets for host cell line engineering. Using the HEK-293SF cell line and the A/PuertoRico/08/1934 (H1N1) influenza vaccine strain as a model, I designed and carried out a genome-wide knockout screen for cellular "restriction factors" that reduce influenza vaccine yield. The screen identified n=135 putative influenza restriction factors at a significance threshold of $p<0.01$. The results of the screen were then applied to create high yield multiple knockout cell lines for cell-based influenza vaccine production. Characterization of the best performing multiple knockout clones showed a roughly 10-fold improvement in volumetric yield in terms of packaged viral genomes and a 2-fold improvement in terms of active HA yield. In addition to the work conducted to increase influenza vaccine yield, attempts were made to apply genome-wide activation screening to increase the yield of another viral biologic - AAV2 vectors produced by transient transfection of HEK-293SF. This led to the novel observation that even in highly optimized systems with transfection efficiencies exceeding 60%, only ~7% of cells produce assembled AAV2 vector particles. Overall, the work presented here demonstrates the utility and drawbacks of pooled genome-wide CRISPR screening as a cell line development tool, and offers insights into the host determinants of viral yield for the cell-based production of influenza vaccines and AAV2 vectors.

ABSTRACT (FRANÇAIS)

Les produits biologiques viraux tels que les vaccins et les vecteurs de transfert de gènes sont principalement et de plus en plus produits dans des lignées cellulaires immortalisées. La modification génétique de la lignée cellulaire hôte peut offrir un moyen d'augmenter le rendement viral, en intensifiant la production virale pour un large éventail de thérapies déjà établies ou émergentes. Dans cette thèse, j'explore l'utilisation du criblage CRISPR/Cas/9 sur l'ensemble du génome pour identifier des cibles pour l'ingénierie de la lignée cellulaire hôte. En utilisant la lignée cellulaire HEK-293SF et la souche vaccinale antigrippale A/PuertoRico/08/1934 (H1N1) comme modèle, j'ai conçu et réalisé un criblage pangénomique pour les « facteurs de restriction » cellulaires qui réduisent le rendement du vaccin antigrippal. Le dépistage a identifié n = 135 facteurs de restriction présumés de la grippe à un seuil de signification de $p<0.01$. Les résultats du criblage ont ensuite été appliqués pour créer des lignées cellulaires a « knock-out » multiples avec un haut rendement pour la production de vaccins antigrippaux en cultures cellulaires. La caractérisation des clones « knock-out » multiples les plus performants a montré une amélioration d'environ 10 fois du rendement volumétrique en termes de génomes viraux enveloppés et une amélioration de 2 fois en termes de rendement de HA actif. En plus des travaux menés pour augmenter le rendement du vaccin contre la grippe, des essais ont été réalisés pour appliquer le dépistage d'activation à l'échelle du génome pour augmenter le rendement d'un autre vecteur viral biologique - AAV2 produit par transfection transitoire de HEK-293SF. Cela a conduit à observation originale importante indiquant que même pour des systèmes hautement optimisés avec des efficacités de transfection supérieures à 60 %, seulement ~ 7 % de la population cellulaire produisent des particules de vecteur AAV2 assemblées et fonctionnelles. Dans son ensemble, le travail présenté dans cette thèse démontre l'utilité en soulignant les avantages et les inconvénients du dépistage CRISPR à l'échelle du génome complet en tant qu'outil de développement de lignées cellulaires, et offre un aperçu des déterminants de l'hôte du rendement viral pour la production cellulaire de vaccins antigrippaux et de vecteurs AAV2.

ACKNOWLEDGEMENTS

The work completed within this thesis was supported by funding from the National Sciences and Engineering Research Council (NSERC), Canadian Institutes of Health Research (CIHR), the McGill Faculty of Engineering, and the McGill Department of Bioengineering. It also benefitted from both formal and informal collaborations with individuals at the National Research Council (NRC), Genome Québec, the McGill Genome Excellence Center, and Université du Québec à Montréal (UQAM). Significant scientific and technical contributions to the work presented here were provided by Sean Nesdoly, Shantoshini Dash, Michelle Tran, Hsin Yang, Sophie Broussau, and Denis Flipo (see co-author contributions for specific contributions to each chapter).

My PhD supervisory committee (David Juncker, Rongtuan Lin, and Yu Xia) as well as Alaka Mullick provided formal scientific feedback and guidance. Equally important was the informal feedback and support provided by Sascha Kiesslich, Michelle Tran, Omar Farnos, Milica Momčilović, and other members of the Kamen lab group, as well as from my partner, Aaron.

Finally, I would like to thank my supervisor, Amine Kamen, for his guidance and for leading by example as a compassionate, skilled, and pragmatic scientist

CONTRIBUTION TO ORIGINAL KNOWLEDGE

The pooled genome-wide CRISPR screen conducted in section 4.0 of this thesis represents the first time that pooled CRISPR screening has been used with the aim of increasing influenza vaccine yields. The results identified n=135 putative influenza restriction factors, n=41 of which were not previously known to impact influenza replication. Subsequent analysis of the genes identified demonstrates that the genetic determinants of influenza yield in the HEK-293SF cell line are primarily metabolic, a striking contrast to the innate immune factors identified in previous screens on other cell lines intended to model human infection. This finding has broader implications for cell line engineering efforts for HEK-293SF and other immortalized cell lines used to produce viral vaccines and gene therapy vectors. This work was published in *Scientific Reports*[1].

Attempts detailed in section 5.0 to apply screening data towards the creation of high-yield multiple knockout cell lines for influenza vaccine production were highly instructive on several fronts. It was shown that individual single guide RNA (sgRNA) efficiency can be adversely impacted by simultaneous expression of multiple constructs. This finding suggests that multiple knockouts should be induced sequentially, rather than in parallel, when constructing multiple knockout cell lines. Section 5.0 also highlights the dangers of using indirect reporters as a proxy for viral replication in pooled genome-wide screens. The results demonstrate that even with careful validation of the reporter system, the genetic perturbations induced by screening can cause decoupling of normally correlated biological processes, leading to unpredictable results.

This thesis is also the first reported use of conformation-specific antibody staining to quantify the percentage of cells that contain assembled viral capsids during the transfection-

based production of AAV2 vectors. This revealed the surprising result that despite achieving transfection efficiencies in excess of 60%, only ~7% of cells in culture produced detectable amounts of assembled AAV capsid. Subsequent experiments suggested that this is due to insufficient expression of AAV capsid monomers. This finding has important implications for future efforts to optimize this widely used vector production platform, and was published in *Biotechnology and Bioengineering*[2].

CO-AUTHOR CONTRIBUTIONS

Sean Nesdoly performed the initial analysis (i.e. QC of read quality and library coverage via the FastQC tool, as well as tabulation and normalization of raw read counts) of all deep sequencing data presented in sections 4.0, 5.0, and 6.0 and wrote the associated methods entry in section 3.14. Sean Nesdoly also completed Robust Rank Aggregation (RRA) analysis shown in Figure 9. Shantoshini Dash designed and performed the confocal microscopy study illustrated in Figure 22 and optimized the staining protocol for AAV2 capsids used in section 6.0 and wrote the associated methods entry in section 3.10. Michelle Tran operated the 1L bioreactor for experiments performed in section 5.3. Hsin Yang assisted in generating and validating the HEK-293SF$^{\Delta TBK1}$ knockout pool shown in section 4.2. Sophie Broussau and other staff at the NRC performed the single cell isolation of the multiple knockout (MK) clones shown in section 5.0. Denis Flipo assisted in the operation of the FACS instrument for experiments shown in sections 4.0 and 6.0 and advised on optimal sorting parameters. Amine Kamen translated the abstract of this thesis into French.

With the exception of the contributions stated above, all ideas, text, figures, design and execution of experiments, and analysis of results were carried out by myself

LIST OF FIGURES AND TABLES

Figure 21. Selected quality control metrics for pooled genome-wide pooled Calabrese activation screen

Figure 22. Expression of transfected plasmids during AAV2-GFP vector production

LIST OF ABBREVIATIONS

AAV	Adeno-associated virus
BEVS	Baculovirus expression vector system
CARM1	Coactivator-associated arginine methyltransferase 1
Cas	CRISPR associated proteins
CFS	Cell-free supernatant
CIHR	Canadian Institutes of Health Research
CPSF	Cleavage and polyadenylation specificity factor
CRISPR	Clustered regularly interspaced short palindromic repeat
crRNA	CRISPR RNA
CSP	Cell specific productivity
dCas9	(catalytically) dead Cas9
DDX6	DEAD-box helicase 6
DOE	Design of experiments
FACS	Fluorescence assisted cell sorting
FBS	Fetal bovine serum
gDNA	Genomic DNA
GFP	Green fluorescent protein
GO	Gene ontology
HA	Hemagglutinin
hAd5	Human adenovirus type 5
HDR	Homology directed repair
HEK-293	Human embryonic kidney cell line 293
Hpi	Hours post infection
Hpt	Hours post transfection/transduction
IAV	Influenza A virus
IBV	Influenza B virus
ICV	Influenza C virus
IDV	Influenza D virus
IFN	Interferon
ITR	Inverted terminal repeat
Lfc	Log fold change
M1	Matrix 1
M2	Matrix 2
MAVS	Mitochondrial antiviral-signaling protein
MDCK	Madin-Darby canine kidney
MK	Multiple knockout
MOI	Multiplicity of infection

NA	Neuraminidase
NEP	Nuclear export protein
NHEJ	Nonhomologous end joining
NLS	Nuclear localization signal
NMD	Nonsense mediated decay
NP	Nucleoprotein
NRC	Rational Research Council of Canada
NS1	Non-structural protein 1
NS2	Non-structural protein 2
NSERC	National Sciences and Engineering Research Council
NTC	Non-targeting control
NXF1	Nuclear RNA export factor 1
NXT1	Nuclear transport factor 2 like export factor 1
OFAT	One factor at a time
PA	Polymerase acidic
PAM	Protospacer adjacent motif
PB1	Polymerase basic 1
PB2	Polymerase basic 2
PEI	Polyethylenimine
PKR	Protein kinase R
PPI	Protein-protein interactions
PR/8	A/PuertoRico/8/1934 (H1N1)
rcAAV	Replication competent AAV
RdRp	RNA dependent RNA polymerase
RIG-I	Retinoic acid-inducible gene I
RRA	Robust rank aggregation
scCRISPR-seq	Single cell CRISPR sequencing
SEM	Standard error of the mean
SF-9	Spodoptera frugiperda 9 cells
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
SMG9	SMG9 nonsense mediated mRNA decay factor
SRID	Single radial immunodiffusion
TBK1	Tank binding kinase 1
TGN	Trans-golgi network
TIDE	Tracking of indels by decomposition
tracrRNA	Trans-activating RNA
UQAM	Université du Québec à Montréal
VG	Viral genomes
VLP	Virus-like particle
VP64	4x Herpes Simplex Viral Protein 16

vRNA	Viral RNA
vRNP	Viral ribonucleoprotein

1.0 INTRODUCTION

Viral biologics represent a significant and growing portion of the biopharmaceutical market, encompassing vaccines, gene therapy vectors, vectors for CAR-T cell therapies, and oncolytic viral therapies[1]. While the medical applications of viral biologics hold much promise, realizing that promise requires a parallel effort to develop and intensify manufacturing capacity. Many viral vaccines are still produced using decades old processes that are unable to effectively respond to pandemics, and the availability of viral vectored cell and gene therapies is severely constrained by astronomical manufacturing costs. In this thesis, I use HEK-293SF suspension cells as a model to explore the use of genome-wide CRISPR/Cas9 screening as a tool to intensity the production of two viral biologics: influenza vaccines and AAV vectors.

Influenza are non-enveloped, negative stranded RNA viruses of the family *Orthomyxoviridae* that cause seasonal endemics and occasional pandemics. The burden of disease from influenza is felt worldwide, with nearly 400,000 deaths annually from seasonal endemics alone[3,4]. Of the four genera of influenza, influenza A is responsible for the majority of severe disease during seasonal endemics and all influenza pandemics. Lacking a truly effective antiviral, the principle public health intervention to blunt the morbidity and mortality of influenza is vaccination. Due to a high frequency of mutation and changes in circulating strains, vaccines must be reformulated and administered annually, with over 1.5 billion doses required[5].

Upwards of 80% of influenza vaccines are still manufactured using embryonated chicken eggs, a technology that has remained largely unchanged since initial development in the 1940s[6]. While economical, long production timelines using this method frequently cause strain mismatch and genetic drift of the vaccine from the wildtype strain, resulting in efficacies in the range of 10-60%[7]. Long lead times also impact the ability of egg-based manufacturing processes to

effectively respond to pandemics. During the most recent influenza pandemic caused by the 2009 "Swine flu", 123,000 to 203,000 people died and ~10% of the global population was infected in the nine months before a vaccine was widely available[8,9]. With recent studies highlighting the potential for highly pathogenic avian influenza to cause a much deadlier pandemic, there is an imperative to develop alternative vaccine production capabilities[10].

Cultivating influenza using cell lines instead of eggs removes a major limiting factor of vaccine production capacity in that it is not constrained by the availability of billions of pathogen-free, fertilized chicken eggs. Cell-based methods also allow the use of reverse genetics (i.e. producing live virus from the transfection of DNA into cells) to accelerate viral seed stock production, condensing the early stages of vaccine production from two months to two weeks[11]. While promising, cell-based influenza vaccines are currently something of a specialty product and make up a fraction of the overall doses produced. This is in part due to the lower volumetric yield (and therefore higher manufacturing costs) of cell-based vaccines compared to egg-based methods[12].

The widespread adoption of many promising gene therapy vectors is also limited by manufacturing issues. AAV are small, non-enveloped, single stranded DNA viruses of the family *Parvoviridae*[13]. Originally identified as contaminants in Adenovirus cultures, AAV require coinfection of a cell with a "helper virus" in order to initiate their replication[13]. When no helper virus is present, they have evolved to persist in the cell integrated into the host AAVS1 locus or (in the case of AAV vectors) as an extragenomic episome[14]. This inherent ability to persist in the cell as an episome, combined with a lack of apparent pathogenicity and ability to transduce a wide range of tissues, has contributed to the popularity of AAV as a viral vector platform[15]. As of 2021, there are 250 ongoing clinical trials utilizing AAV vectors to treat a

range of inherited disorders[16]. While AAV has many attractive qualities as a viral vector, it does suffer from relatively poor transduction efficiency *in vivo*, necessitating doses as high as 10^{13} TU/kg to reach therapeutic transgene levels[17]. These high dose requirements, and associated manufacturing costs, are a contributing factor to the astronomical cost of AAV vectored therapies. The first approved AAV gene therapy, a treatment for lipoprotein lipase deficiency, was infamously dubbed the "million dollar drug"[17].

The bulk of AAV produced for both clinical and research activities are generated via multi-plasmid transfection of mammalian cells[18,19]. Low yields in transfection-based AAV vector production platforms are a long-standing issue, and have spurred the development of numerous alternatives. Plasmid-free systems such as baculovirus expression vectors, herpesvirus vectors, and more recent self-silencing adenoviral systems boast significantly increased yields and are far more amenable to scaleup[20,21]. There have also been attempts both within academia and industry to develop a stable producer cell line, though little of the latter work is published[19]. While these platforms may supplant transfection for late-stage and approved therapies, the unparalleled speed and simplicity of transfection-based manufacturing means it is likely to remain a mainstay of AAV vector production for early clinical and research applications in the foreseeable future.

Both influenza vaccines and AAV vectors are commonly grown in variants of the HEK-293 cell line. The HEK-293 cell line can be adapted to serum-free suspension culture, is highly transfectable, and is amenable to the replication of a wide range of viruses[11]. Process development over the past 20 years to optimize the production of these two viral biologics has focused on media composition, optimization of transfection or infection conditions, the use of fed-batch and perfusion processes, etc. However, the advent of CRISPR genome editing

technology has allowed for a relatively unexplored avenue of process development: genetic engineering of the host cell line.

CRISPR has the potential to modify host innate antiviral pathways and alter metabolism to create a more permissive environment for viral vaccine and vector production. However, determining targets for genetic modification is non-trivial. Thousands of host proteins are known to be involved in the replication of influenza, either as direct binding partners of viral proteins or indirectly as regulators of host metabolism and innate immunity[22]. Determining targets from existing literature is further complicated by the fact that immortalized cell lines cells used to produce viral biologics such as HEK-293, Vero, and Madin-Darby canine kidney (MDCK) cells are often highly divergent from primary cells in terms of their response to viral infection and assembly[12,23].

An alternative approach is to use pooled genome-wide CRISPR screening libraries to identify host targets for genome engineering. Such libraries, consisting of a pool of sgRNA expression constructs delivered by lentiviral vectors, can simultaneously assess the impact of a given genetic perturbation (knockout, overexpression, etc.) on a particular phenotype for every gene in the human proteome. Genome-wide CRISPR screening libraries have been used extensively in basic research applications explore the genetic contributions to various phenotypes[24]. With some adaptation, it should be possible to use Genome-wide CRISPR screening libraries to identify host cell engineering targets with the greatest impact on viral yield, and subsequently use this information to create high yield cell lines. Here, I explore this technique as a process development tool to intensify the cell-based production of A/PuertoRico/08/1934 (PR/8) influenza and AAV2 vectors in the HEK-293SF cell line. Specific aims include:

1. Design and conduct a pooled genome-wide CRISPR/Cas9 knockout screen for influenza "restriction factors" that inhibit influenza replication.

2. Apply the results of the screen to create high yield knockout cell lines in HEK-293SF for cell-based influenza vaccine production.

3. Explore whether pooled genome-wide CRISPR screening can be applied to intensify the production of other viral biologics such as AAV2 vectors.

2.0 LITERATURE REVIEW

2.1 Influenza taxonomy and host range

Influenza are enveloped viruses with multipartite, negative stranded ssRNA genomes (type V Baltimore classification system) of the family *Orthomyxoviridae*, along with Quaranjaviruses, Isaviruses and Thogotoviruses[25-28]. Influenza viruses are divided into four genera: *Alphainfluenzavirus*, *Betainfluenzavirus*, *Gammainfluenzavirus*, and *Deltainfluenzavirus* (also called influenza type A, B, C, and D, respectively)[28,29]. Based on homology studies, last common ancestor of all influenza viruses is estimated to have existed 8000 years ago before diverging into two clades that would be the common ancestors of influenza A/B and influenza C/D, respectively. The influenza A/B clade is readily distinguished from the influenza C/D clade based on the number of genomic segments, with influenza A virus (IAV) and influenza B virus (IBV) each carrying eight unique genomic segments, while influenza C virus (ICV) and influenza D virus (IDV) each have seven[28,29].

Influenza viruses within a given genus are further classified in terms of lineages-sequence homology to well characterized isolates. Classification of IAV viruses also incorporates the subtype of the two surface glycoproteins hemagglutinin (HA) and neuraminidase (NA), which are major determinants for IAV host range, virulence, and vaccine efficacy. The standard naming convention for describing influenza isolates is depicted in Figure 1 below.

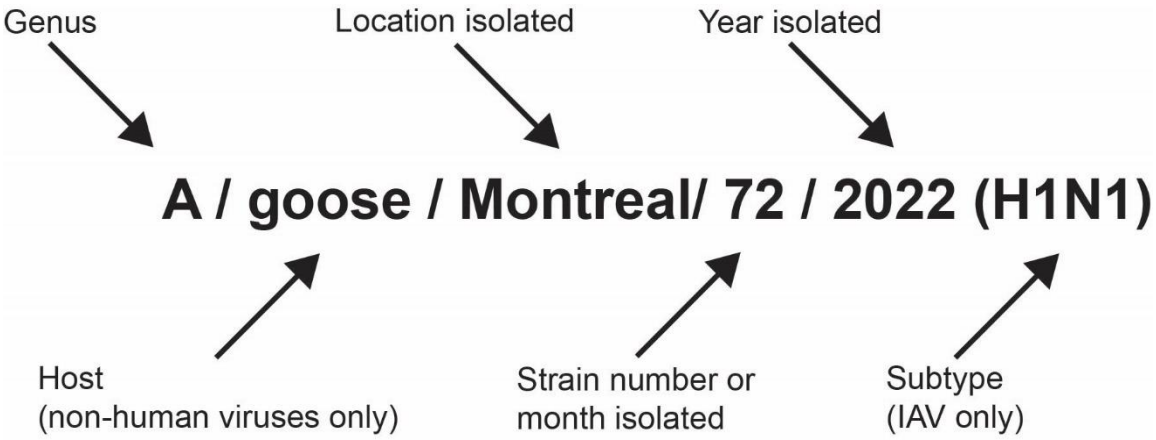


Figure 1. Naming convention for influenza isolates

All influenza viruses encode functional analogs of HA and NA, but the diversity of these two proteins is far higher in IAV which enables a significantly expanded host range. Of the 18 known subtypes of HA and the 11 known subtypes of NA, nearly all are present in wild birds, particularly ducks, gulls, and shorebirds[30]. These animals are infected at a high prevalence, exceeding 20% during major migrations[30]. A notable exception to this is three highly divergent subtypes of HA (H17-19) and two NA subtypes (N10-11) that thus far have been detected only in bats[31]. In addition to its primary reservoir of waterfowl, IAV is also capable of sustained transmission in pigs, horses, dogs, cats, aquatic mammals, rodents, and humans in a subtype specific manner[30]. Only H1N1, H3N2, and the now extinct H2N2 subtypes have shown circulation within the human population in the last century[32]. However, several subtypes have shown the potential for cross-species transmission and eventual adaptation to humans including H5Nx, H7Nx, H9N2, H6N2, H5N5, and H5N6 [33].

Influenza B viruses form two main lineages: B/Victoria and B/Yamagata[34]. In contrast to IAV, IBV primarily infects humans. Lineage determination of ICV and IDV are less clear, though for different reasons. Comprehensive homology analysis of ICV isolates collected since 1947 indicates that two main lineages exist: C/Mississippi/80-like and C/Yamagata/81-like[35]. However, these lineages readily reassort with one another at a much higher rate than is seen with IBV or IAV, and a virus often possess genomic segments derived from multiple different lineages[35]. IDV was only recently discovered in 2011, and at present lacks a comprehensive catalog of sequenced isolates from which to conduct a robust homology analysis[35].

Seasonal influenza endemics and their associated morbidity and mortality are almost exclusively caused by IAV and to a lesser extent IBV. IAV is also responsible for all influenza pandemics and significant damage to human agriculture (see section 2.7). In contrast, ICV

primarily infects young children, with up to 98% of children >12 years of age seropositive[36,37]. The disease is typically mild, and generally only requires intervention when significant comorbidities are present, with prematurely born children particularly vulnerable[37]. The primary host of IDV appears to be cattle and smaller ruminants with zoonotic transmission to humans that does not cause apparent disease[29]. For these reasons, ICV and IDV are considered to represent a minor threat to public health and are not currently targets for vaccination. Thus, the rest of this thesis will thus focus primarily on IAV with some discussion of IBV.

2.2 Influenza A genome

The IAV genome is composed of eight segments of negative sense ssRNA that total roughly 13.5kb. They are commonly referenced either ordinally in terms of decreasing length, or by their primary protein products[28]. All segments are flanked by highly conserved semi-complementary sequences that allow recognition by the viral polymerase, act as promotors for mRNA transcription and genome replication, and mediate the tertiary structure of viral RNA (vRNA)[38]

2.2.1 PB2/Segment 1

Segment 1 is the largest genomic segment of IAV, measuring 2.3kb in length[28]. The principle transcript of segment 1 is the Polymerase basic 2 (PB2), a ~750aa protein that primarily functions as a subunit in the viral RNA dependent RNA polymerase (RdRp); localizing to the nucleus via a C-terminal nuclear localization signal (NLS) and forming a heterotrimer with the viral polymerase acidic (PA) and polymerase basic 1 (PB1) proteins to facilitate viral transcription. Within this complex PB2 recognizes and binds the 5’ cap structure on host pre-

mRNA, a necessary function in the "cap-snatching" pathway used by influenza to inhibit host translation and prime transcription of influenza mRNA[39].

In addition to its core role as a component of the influenza RdRp, the PB2 protein also accumulates at mitochondrial matrix via an N-terminal localization signal where it has been shown to inhibit the signalling of Mitochondrial antiviral-signaling protein (MAVS), which have import roles in signal transduction for several antiviral pathways (chiefly the retinoic acid-inducible gene I (RIG-I) viral RNA sensing pathway) [40]. PB2 localization to the mitochondria is associated with decreased induction of downstream antiviral effectors such as type I interferon (IFN)[40].

Certain seasonal lineages also produce a splice variant of PB2, termed PB2-S1. PB2-S1 features a 250aa deletion that eliminates cap-binding functionality and localizes to the mitochondria[41]. Though splice donor/acceptor sites are highly conserved in certain lineages, the function of this splice variant is not clear, and deletion does not slow viral replication *in vitro* or impact virulence *in vivo*[41].

2.2.2 PB1/Segment 2

The PB1 segment of IAV is in many ways similar to the PB2 segment, being of similar length and encoding a comparably sized principle transcript (the PB1 protein) that localizes to the nucleus to form part of the viral RdRp[28]. Within the RdRp PB1 functions as the catalyst for RNA polymerization during mRNA transcription and genomic replication[39].

Segment 2 of IAV may also (in a strain dependent manner) encode an 87aa accessory protein called PB1-F2 in the +1 reading frame of PB1[42]. The function of PB1-F2 appears to be host-specific. In mammalian adapted influenza, PB1-F2 primarily localizes to the mitochondria

where it both inhibits activations of MAVS and alters the permeability of the mitochondrial membrane in a manner that ultimately induces apoptosis[43]. Cytosolic PB1-F2 inhibits the assembly and activation of the inflammasome, which has roles in cytokine processing and signaling during viral infection[44]. Conversely, the PB1-F2 of H5N1 avian influenza localizes to the nucleus, where it optimizes viral transcription[45]. This differential localization of PB1-F2 between mammalian and avian adapted influenza has been suggested to contribute to the greater degree of detrimental innate immune activation (i.e. cytokine storm) observed during human infection with avian strains[44].

2.2.3 PA/Segment 3

The principle transcript of IAV segment 3 is the 716aa PA protein which, together with PB1 and PB2, forms the heterotrimeric viral RdRp[28]. Within the RdRp, following recognition and binding of the host cap structure by PB1, PA cleaves the 5' CAP and leading 10-13nt from nascent host mRNA. These capped fragments are then used to prime viral mRNA transcription[39].

Also encoded on segment 3 is PA-X. PA-X is transcribed via ribosomal frameshift during the transcription of PA, and consists of the PA endonuclease domain and the so-called "X-ORF"[46]. X-ORF domain interacts with host splicing machinery during the processing of host RNA pol II transcripts and allows the endonuclease domain to degrade these transcripts. Influenza RNA is transcribed by the viral RdRp and is thus spared degradation[47]. This suppression of host transcripts both liberates host translational machinery to translate viral mRNA and prevents the expression of inducible host innate immune factors[47]. Unlike many of

the accessory proteins discussed in this section, PA-X expression appears to be conserved in all IAV, indicating a critical role in facilitating the viral lifecycle[48].

2.2.4 HA/Segment 4

Segment 4 encodes only one protein: the ~565aa HA[28]. HA is perhaps the best studied of any influenza protein due to its role in facilitating viral attachment and entry and, by extension, plays a critical role in cell tropism, host range, and virulence[49]. HA is also the most abundant protein on the surface of influenza virions by a 4:1 ratio, and as a result the bulk of the humoral immune response to influenza infection in humans (both in terms of total and neutralizing antibody titer) is directed against it[50]. The importance of HA in immunity is underscored by the fact that active HA titer is the only antigen required to be measured during the formulation of influenza vaccines[51].

Active HA exists as a homotrimeric type-1 fusion protein consisting of a highly variable head domain, a more conserved stalk domain, and finally the transmembrane and cytoplasmic domains the anchor the complex to the viral envelope and matrix, respectively[52]. Full length HA (called HA0) must undergo proteolytic cleavage into the HA1 and HA2 subunits to become fusion competent[28]. Highly pathogenic avian influenza IVAs contain a polybasic cleavage site and are cleaved by furin within the trans-golgi network (TGN)[53]. By contrast, human-adapted IAVs contain a monobasic cleavage site and are cleaved by one of several proteases found either on the surface of the cell membrane or within the airway itself. Where and how this cleavage event occurs is an important aspect of influenza host adaptation and virulence, since the expression of host proteases varies depending on species, cell type, and location in the airway[54].

The variable head domain of HA recognizes and binds the influenza receptor; terminal sialic acid residues on host glycoproteins and glycolipids. Depending on species and anatomical location, terminal sialic acid residues vary in how they are bound to the penultimate galactose sugar in the carbohydrate chain[55]. Different subtypes and lineages of HA have adapted to preferentially bind different conformations of sialic acid. Avian influenza strains, for example, preferentially binds α(2,3) linkages. Humans primarily display α(2,6) linked sialic acid in the upper airway epithelium. However, α(2,6) linkages are found deep within human lungs[56]. This is thought to partially explain why avian influenza strains are poorly transmissible to humans but are associated with high mortality when infection occurs. In this way, the HA head domain affinity for different sialic acid linkages is an important determinant of influenza transmissibility, host range, virulence, and viral evolution. The majority of neutralizing antibodies against influenza virus also target the HA head domain and function simply by blocking the interaction between HA and sialic acid[50]. While effective, the rapid evolution of this domain compared to the stalk and transmembrane domains mean immunity is transient[57].

The stalk domain of HA contains three α-helices within a hydrophobic pocket that facilitate membrane fusion. In order for this process to occur the stalk domain of influenza is required to undergo major pH-dependent conformational changes to expose these hydrophobic α-helices[56]. Interestingly, antibody binding to the stalk domain can inhibit these conformational changes and disrupt the viral lifecycle[58]. This subset of neutralizing antibodies directed towards the HA stalk domain, though rare compared to those directed against the head domain, are of considerable interest from a vaccination perspective since the stalk domain is highly conserved among different IAV subtypes[59].

2.2.5 NP/Segment 5

Genomic segment 5 encodes a single protein: the 498aa nucleoprotein (NP)[60]. Broadly speaking, NP is a structural protein that functions to protect and shuttle influenza vRNA to different subcellular locations. NP contains a basic patch of residues that associate with the phosphate backbone of vRNA, at a ratio of one monomer of NP per 24bp of RNA[61]. NP also features a flexible "tail" region that allows it to bind other NP monomers, allowing oligomerization into a chain that covers the entirety of the vRNA backbone, protecting the viral genome from degradation and recognition of dsRNA structures by the host immune system[61]. Individual genomic segments of NP coated vRNA also associate with the viral RdRp to form viral ribonucleoproteins (vRNP) complexes[62]. vRNP complexes are modular segments of viral genome and associated polymerase that facilitate genomic packaging and transport, largely through the interaction of NP with viral proteins or host nuclear import/export machinery[62]. Interaction with the viral (M1) protein, encoded on segment 7, is particularly important as this regulates the activity of NP's nuclear export signals (NES) and NLS signals[63,64]. NP also interacts with a number of host factors, particularly importins that facilitate the traffic of vRNP complexes across the nuclear membrane and acetylases that kinases that appear to regulate the timing of NP oligomerization and vRNP complex formation through modification of NP residues[65,66].

It's notable that NP is one of the most abundantly translated influenza protein, and so is common target for antibody-based assays to detect or quantify influenza infection within the cell[28].

2.2.6 NA/Segment 6

Segment 6 encodes the protein NA, a 470aa sialidase[28]. NA makes approximately 20% of the glycoprotein complement on the surface of mature IAV virions, with HA making up the remainder. NA's primary function is to cleave sialic acid residues from the surface of the host cells. While it might seem contrary for a virus to remove the entry receptor of host cells, NA activity is vital for efficient viral propagation as it balances the opposing activity of HA and prevents aggregation of newly formed virions on the surface of infected host cells[67]. In addition to its role in facilitating viral egress, NA also prevents entrapment of nascent virions on mucins that are rich in sialic acid residues and are excreted into the respiratory tract as a response to IAV infection[68]. Due to its critical role in facilitating efficient viral replication NA is a target for a number of influenza antivirals including Oseltamivir, the most commonly used influenza antiviral[69].

Catalytically active NA exists as a monotetramer that cosmetically resembles HA in structure; it consists of a head domain containing the catalytic site, a stalk domain, and a transmembrane and cytoplasmic domain that anchors the protein in the viral envelope[70]. Different NA subtypes display a preference for specific linkages between sialic acid and galactose in a similar manner to HA, necessitating coevolution between the two in order to maintain fitness in a specific host. In this way NA is also an important determinant of host range and virulence[70].

Despite being underrepresented on the viral envelope compared to HA, NA has been shown to be equally as immunogenic[50]. Antibodies that target NA have been shown to inhibit influenza infection both *in vitro* and *in vivo,* and recent epidemiological studies suggest NA

immunity can strongly impact the severity of influenza infection in humans[58]. Depending on manufacturing method, influenza vaccines may also contain NA in addition to HA, but it is not always quantified and the overall impact of NA on vaccine efficacy is an active area of research.

2.2.7 M/Segment 7

Segment 7 encodes two structural proteins that are transcribed through alternative splicing: the M1 matrix protein and the matric 2 (M2) ion channel[28]. M1 is a highly conserved and abundantly transcribed protein measuring 252aa in length[71]. High transcription and conservation mean it is a common target for PCR-based detection and quantification of influenza. As the name implies, polymers of M1 form and give structure to the viral core. They directly bind vRNPs (both through NP and direct interaction with vRNA), the viral envelope, and the cytoplasmic tails of HA and NA[71]. Binding of these proteins is not merely structural, but also plays a key role in regulating subcellular localization of these complexes. M1 binding upregulates NES signals of NP, promoting export of vRNP and M1 from the nucleus and aggregation at the cellular membrane[56]. Interaction of M1 with the cytoplasmic tails of HA and NA embedded in the cell membrane causes vRNP to localize to patches of the cell membrane rich in these proteins[56]. Direct binding of the cell membrane itself then triggers the polymerization of M1 and drives viral budding[72]. Indeed, the expression of M1 alone is sufficient to induce budding and produce virus like particles (VLPs). When expressed in conjunction with HA, these VLP's are capable of entering new cells and triggering an immune response to HA, and are currently under investigation as an alternative to current influenza vaccines[73]. M1 polymerization triggered through binding of sterols at the cell membrane, as well as binding to vRNPs, can be reversed under low pH conditions[72]. Depolymerization allows viral uncoating within endosomes during viral entry, while unbinding of NP allows the NLS of NP

become dominant and facilitate nuclear import of vRNPs[56]. In addition to its structural functions, M1 may also inhibit this innate immune response by blocking the classical complement pathway[74].

M2, while not as abundantly transcribed, in nonetheless vital to the viral lifecycle. Active M2 exists as a homotetramer that functions as a pH-activated proton channel. After translation M2 is targeted towards the cell membrane along with HA and NA via the TGN[75]. In strains where HA is proteolytically cleaved to a fusion competent form within the TGN, M2 is neutralizes the pH of the TGN to prevent premature pH-dependent conformational changes in HA[76]. Once incorporated into the envelope of mature virions, M2 is activated during endosome acidification, transporting protons across the envelope to allow acidification of the viral core and subsequent depolymerization of M1 and release of vRNPs[75].

2.2.8 NS/Segment 8

Segment 8 encodes two proteins through alternative splicing, the 230aa non-structural protein 1 (NS1) and the 121aa non-structural protein 2 (NS2)[28]. Following the discovery that NS2 is incorporated into mature virions, it was given the alternative name of nuclear export protein (NEP)[77]. NS1 is a non-structural protein that functions to create a more permissive environment for viral replication by modifying the activity of various host factors, while NEP augments vRNP shuttling and vRNA transcription/replication[77,78]. NS1 is not strictly necessary for IAV replication but vastly enhances viral fitness, particularly within *in vivo* systems, and is highly conserved across all IAV subtypes[28].

NS1 is active as a homodimer, though it may adopt one of any three distinct structural conformations, capable of binding both RNA and DNA[79]. Note that this section will not attempt

to provide an exhaustive description of the host-virus interactions mediated by NS1 (the interactome of which spans some 252 host proteins), instead focussing on well characterized interactions[22]. One of the major functions of NS1 is to prevent the induction of type 1 IFN by disrupting the RIG-I signalling cascade. NS1 binds and inhibits RIG-I, as well as the ubiquitin ligases TRIM25 and Riplet involved in RIG-I signal transduction[80,81]. NS1 also binds dsRNA competitively inhibit recognition by RIG-I and other host dsRNA sensors such as protein kinase R (PKR) and RNase L[78]. More generally, NS1 inhibits the induction of antiviral responses by disrupting host gene expression; NS1 binds dsDNA in a non-specific manner and has been shown to inhibit binding of host promotors[82]. In addition, host mRNA maturation is inhibited by SN1 binding of CPSF30, preventing assembly of the cleavage and polyadenylation specificity factor (CPSF) complex which is required for most mRNA polyadenylation and results in pre-mRNA accumulating in the nucleus[83]. This buildup of pre-mRNA may also function to provide substrate for the viral cap-snatching mechanism and may explain why NS1 also functions to block mRNA export through interference with nuclear export factors such as Rae1, NXF1, and Nup98[84].

In addition to preventing the induction of innate antiviral responses, NS1 also directly inhibits downstream effectors. PKR is induced by interferon and activated by dsRNA, shutting down CAP-dependent translation through the phosphorylation of eIF2a. Binding of NS1 to PKR has been shown to prevent eIF2a phosphorylation[85]. NS1 has also been shown to inhibit other antiviral proteins such as ISG15 and RNAse L[78]. Apoptosis, the endpoint of many antiviral pathways, is also modulated by NS1 through interactions with p53, HSP90, and components of the pro-survival PI3k-Akt pathway[86]. NS1 has been shown to both prevent and promote apoptosis in a strain and time dependent manner. This is thought to reflect evolutionary trade

offs between delaying early apoptosis to allow for viral replication, versus induction of apoptosis to minimize inflammatory responses[86].

The best studied function of NS2/NEP is that of an adaptor between M1 and host nuclear export factor Crm1[87]. In this role it facilitates the export of vRNP complexes during virion packaging. Distinct from this role, NEP expression has also been demonstrated to regulate the ratio of influenza vRNA and various positive stranded replicative intermediates[77].

2.3 Influenza A virion structure

IAV virions are pleomorphic, and have been observed to assume roughly spherical, filamentous, or intermediate morphologies. Strain, host species, and environmental factors such as pH and temperature have all been identified as morphology determinants. Spherical forms are in the order of 100nm in diameter, while filamentous forms may approach 20μm in length[56]. The outer surface of the viral envelope is studded with roughly 500 copies of HA and NA at a 4:1 ratio, with a much smaller proportion of M2 ion channels[28]. HA, NA, and M2 are anchored not only to the lipid envelope via transmembrane domains, but also to the viral matrix of M1 that polymerizes against the inner surface of the envelope via cytoplasmic domains[88]. NMR studies indicate that M1 can adopt either parallel or antiparallel polymerization geometry, consistent with the variable morphology of the virus[71]. The viral genome, packaged as vRNP, is also bound to the M1 matrix.

vRNPs themselves are highly ordered. All segments of IAV vRNA are flanked by conserved semi-complementary 5’ and 3’ UTR sequences that allow the segments to form of helical hairpins[38]. The minor groove of this helix is filled with NP. The paired ends of each segment are bound to a copy of the heterotrimeric RdRp via the PB2 subunit that allows

transcription of viral mRNA upon infection[61]. The individual vRNP complexes are arranged in a so-called 7 + 1 configuration with one segment at the center and seven others radiating out from the central unit[89].

2.4 Influenza A lifecycle

The IAV lifecycle beings with the attachment of the virus to an appropriate sialic acid residue via HA. Appropriate is the key word here; heavily sialated mucins, endosomes, as well as sialic acid resides on the glycan chains of HA and NA themselves mean the environment is rich in unproductive attachments[70]. The sialidase activity of NA is thus absolutely vital to prevent viral aggregation and/or entrapment of virions[68]. The host cell itself as displays a dense and highly diverse blanket of sialated resides on its surface. HA subtypes will generally have a preference for either α(2,3) linked or α(2,6) linked terminal sialic acid residues, though this dichotomy is not absolute[90].

To what degree the specific glycolipids or glycoproteins that influenza attaches to play a role in cell entry is not entirely understood. Entry can take place through clathrin-mediated endocytosis or macropinocytosis[91,92]. That fact that clathrin appears to aggregate on the inner leaflet of the cell membrane after IAV attachment (rather than the virus sliding into preformed clathrin pits) implies that host proteins are involved in signal transduction to the cytosol[93]. Recent work by Sieben *et al.* using super-resolution microscopy showed influenza entry is localized to high-density overlapping clusters of sialic acid residues and epidermal growth factor receptor, the latter of which is activated by influenza binding[94].

Following binding and entry, influenza is then trafficked to the endosome. Acidification of the endosome triggers conformational changes to HA, translocating the fusion domain of the

protein from hydrophobic pockets to the endosomal membrane and fusing it with the viral envelope[52]. Concurrently, low pH activates the M2 ion channels, allowing the flow of protons into the viral core[75]. This change in pH results in depolymerization of M1, as well as a release of vRNPs from the matrix[72]. Influenza is somewhat unusual among RNA viruses in that it replicates in the nucleus. Trafficking to the nucleus is promoted by the dissociation of NP from M1, exposing NLS signals on NP and resulting in the entire vRNP complex being translocated to the nucleus in an α/β importin-dependent manner[56]. Imaging studies indicate the total time from attachment to nuclear import of vRNP is in the order of one hour[95].

Once within the nucleus, the RdRp begins mRNA transcription and genomic replication. mRNA transcription takes place using a cap snatching mechanism wherein the 5' cap of nascent host mRNA is bound by PB1 and cleaved by PA, a process aided by association of the RdRp with the cellular RNA PolII. The cap, along with the 10-13 downstream nucleotides, are used to prime transcription of mRNA[39]. Genomic replication, on the other hand, begins with the production of a cRNA intermediate initiated by the *de novo* generation of a complementary dinucleotide within the RdRp[96]. A similar *de novo* initiation mechanism is then used to generate vRNA from the cRNA template[96]. The transcription of mRNA is initially favored over vRNA production for several reasons. Primed initiation is much more efficient than *de novo* initiation. Additionally, cRNA must be stabilized by NP and the RdRp into a cRNP complex before it can be used as a template for vRNA production[56,96]. mRNA, polyadenylated through a recursive stuttering mechanism by the RdRp, is exported either to the cytoplasm (PB2, PB1, PA, M1, NP, NS1, NEP, and accessory proteins), or the ER (HA, NA, and M2)[28]. Proteins translated in the cytoplasm are for the most part imported back into the nucleus while proteins translated in the ER are trafficked to the membrane via the TGN[56].

In later stages of infection, the activity of the RdRp becomes biased towards the production of vRNA. The mechanisms underpinning this switch are not entirely clear but have been shown to be impacted by the relative concentrations of host mRNA and RdRp, conformational changes in the RdRp, and the accumulation of short (22-27 nt) positive stranded viral transcripts called svRNA[97]. The binding of M1 to NP promotes nuclear export of vRNP complexes, using NEP as an adapter between M1 and host nuclear export factors[56]. Interaction between M1 and the cytoplasmic tails of HA and NA embedded in the host cell membrane allow vRNP to localize to regions of the cell membrane rich in these proteins[88]. During viral packaging, genomic segments are arranged in a "7 + 1" configuration wherein seven vRNP radiate out from a central vRNP[89]. This 7 + 1 geometry is critical for achieving the correct stoichiometry of genomic segments within each viral particle and is mediated by a complex series of RNA-RNA, RNA-protein, and protein-protein interactions[98,99]. RNA interactions primarily take place between sequences near the 5' and 3' ends of each vRNA, unique to each segment, sometimes called the influenza packaging signal[98]. Interestingly, even in the case of influenza C and D which only possess 7 unique genomic segments this 7 + 1 configuration is conserved, with one segment being duplicated[100]. NEP is also present in purified viral particles, presumably bound to M1, but its functional significance within mature virions is unclear[77].

The interaction between M1 and sterols at the cell membrane promotes M1 oligomerization and viral budding[72]. Following budding, the sialidase activity of NA is critical to cleaving sialic acid residues on both the host cell and HA and NA themselves, promoting viral egress and preventing aggregation[60].

2.5 Host innate antiviral immunity against influenza viruses

The interferon system is perhaps the best studied innate antiviral mechanism. Upon expression, IFN-α/β is exported into the extracellular space whereupon it immediately binds the IFNAR receptor via autocrine signaling, as well as the receptors of neighboring cells via paracrine signaling[101]. IFN-α/β signal transduction utilizes the JAK1/STAT pathway to trigger the phosphorylation of IRF9. IRF9 subsequently acts as a transcription factor for hundreds of ISGs. ISGs accomplish a wide-range of antiviral effects within the cell, including the induction of apoptosis, arrest of translation, modification of cell-cycle[101]. One of the more important ISG effectors is PKR, which acts as a double stranded RNA sensing pathway to shutdown cap dependent translation within the cell[85]. The ISGs MX1 and IFITM have also shown activity against influenza. IFITM blocks influenza entry by inhibiting membrane fusion, while MX1 appears to block RNP formation and viral transcription[101].

RIG-I is a cytoplasmic sensor of RNA featuring a 5' triphosphate group. Although eukaryotic RNA contains a 5' triphosphate group when initially synthesized, it is removed in subsequent processing. However, 5' triphosphate groups are a common feature of viral RNA. Upon recognition of its ligand RIG-I undergoes a series of conformational changes and is subsequently ubiquitinated by TRIM25 and Riplet, translocating to the mitochondria and interacting with MAVS to recruit TBK1 kinase[81]. TBK1 subsequently phosphorylates IRF3 which translocates to the nucleus and acts as a transcription factor to stimulate the transcription of a number of antiviral genes including IFN-α/β[102].

TLR3 and TLR7 which are present within endosomes, and thus are an important component of innate immune sensing for viruses that enter the cell via endocytosis[103]. TLR3 recognizes double stranded RNA that is formed at the 3' and 5' ends of influenza genomic

segments due to self-complementarity and is exposed during virion uncoating within endosomes. Upon activation, TLR3 phosphorylates IRF3 via TRIF which then translocates to the nucleus and acts as a transcription factor for IFN-α/β. TLR7 recognizes U and G rich single stranded RNA sequences presented by influenza and other RNA viruses such as VSV[103]. Signal transduction of TLR7 occurs via the MyD88. Upon activation it activates the transcription factors IRF3, IRF7, and NF-κB. IRF3 and IRF7 stimulate the transcription of IFN-α/β, while NF-κB triggers the transcription of pro-inflammatory cytokines[103].

2.6 Influenza A evolution

Influenza A virus evolution is driven by two distinct mechanisms, antigenic drift, and antigenic shift. Antigenic drift refers to the gradual accumulation of mutations within the influenza genome due to the relatively low (10^{-3}-10^{-4}) fidelity of the RdRp[28]. This high mutation rate has severe consequences for virus viability, with only an estimated 1 in 1000 particles being infectious[32]. However, rapid mutations allow the virus to quickly adapt to avoid recognition by the host immune system. Not surprisingly, genetic variability is concentrated in regions encoding proteins that are common epitope for the adaptive immune system, such as the HA head domain[50]. It is due to this drift that influenza vaccines must be reformulated annually. Notably though, immune escape by antigenic drift is not complete and immunological memory from previous strains confers partial protection and limits both the transmissibility and virulence of seasonal viruses[50].

Conversely, antigenic shift refers to the generation of novel reassortants resulting from superinfection of the same cell with different influenza A subtypes. Reassortment between human influenza viruses and those that circulate within other species can result in viruses for which there is little or no pre-existing immunity in humans, occasionally triggering

pandemics[104]. Reassortment to generate antigenically distinct viruses is not sufficient to cause a pandemic, the virus must also undergo a series of mutations to efficiently interact with new isoforms of necessary host factors and adapt to new physiological conditions[30]. Though virtually all influenza proteins undergo some form of host adaptation, mutations to PB2 and HA have been identified as being particularly important[30]. Influenza A viruses adapted to birds generally display a preference for α2-3 sialic acid linkages while human adapted viruses display a specificity for α2-6 sialic linkages, and thus adaptation to this new receptor isoform is critical for jumping the species barrier[105]. Pigs possess a mixture of α2-3 and α2-6 linkages within their airways and are therefore ideal vessels to support reassortment and adaptation between avian and human influenza virus strains. In this sense, close proximity of birds, pigs, and humans is thought to accelerate the emergence of novel influenza subtypes, and by extension that human agriculture is a major driver of influenza virus evolution[30,105]. Adaptation of HA to different physiological conditions is also a major barrier to cross-species transmission. The human airway has a lower pH than many avian species, leading to premature HA fusion domain translocation in maladapted avian HA subtypes[30]. Other aspects of HA species adaptation include optimization of HA proteolytic cleavage and balance of HA and NA ratio[30,67]. Several highly conserved point mutations to PB2 have also been shown to be critical to human adaptation of influenza viruses. These mutations optimize interaction with human specific isoforms of host factors such as α importin, ANP32A and DDX17[30].

Within the last century there have been four major influenza pandemics. The most severe was the 1918 Spanish influenza pandemic which resulted in the deaths of 50-100 million people worldwide from a novel H1N1 subtype[106]. The virus responsible for this pandemic was sequenced and reconstructed via reverse genetics using bodies frozen in permafrost[107]. Though

there is no definitive consensus as to the origin of the virus, some studies have suggested it resulted from reassortment or direct adaptation of avian strain[108,109]. Following the pandemic, the 1918 H1N1 continued to circulate as seasonal influenza until 1957 when reassortment between this virus and duck H2N2 resulted in the so-called Asian Flu pandemic, resulting in an estimated 1-4 million deaths worldwide[110]. At this point H2N2 replaced 1918 H1N1 as the dominant seasonal strain. Then in 1968 a reassortment between seasonal H2N2 and an H3 subtype in ducks resulted in the Hong Kong Flu pandemic and the deaths of a further 1-4 million people worldwide[110]. Once again this resulted in the supplanting of the previous dominant seasonal strain as H2N2 became extinct and H3N2 became the dominant circulating seasonal strain. In 1977, an H1N1 subtype caused the Russian Flu pandemic resulting in 700,000 deaths[109]. Interestingly, the H1N1 subtype responsible was virtually identical to H1N1 subtypes that had circulated prior to the emergence of the H2N2 subtype in 1957, leading to speculation the pandemic may have resulted from accidental release of archived samples[111]. From this point until 2009, H3N2 derived from the Hong Kong pandemic and H1N1 from the Russian pandemic co-circulated as the dominant seasonal strains. The 2009 “Swine Flu” pandemic is thought to have arisen from multiple reassortment between the circulating H3N2 strain, classical and Eurasian Swine strains, and several avian strains[112]. This novel H1N1 subtype, often called H1N1pdm09 to distinguish it from the Russian Flu H1N1 subtype that it supplanted, currently cocirculates with H3N2 as the dominant seasonal strains.

2.7 Impact of influenza on human health

The impact of seasonal influenza endemics on society takes on a number of facets such as human health, agriculture, and economic impacts. Since 2017 several studies have attempted to estimate the total number of deaths worldwide directly attributable to respiratory complications

arising from seasonal influenza, with estimates ranging from 99,000-650,000 deaths annually[3,113]. The most recent study conducted in 2019 estimates an average of 398,000 deaths annually[3,4]. In 2017 in Canada influenza and associated pneumonia were the eighth leading cause of death[114]. These deaths are disproportionately skewed towards the elderly and/or those with underlying chronic conditions, with the incidence of death from influenza 18-244 per 100,000 in those 75 years and older compared to a rate of 0.1-6.4 among adults aged 18-65[113]. 80% of adults admitted to hospital also had some form of underlying chronic condition, the most common of which was asthma (42%)[115].

The true impact of influenza on human health is likely far higher than simple estimates of death by respiratory complications would suggest. The weeks following influenza infection are associated with a 6-10 fold increase in the incidence of acute myocardial infarction and a 3-8 fold increase in the incidence of stroke[116]. Indeed, studies examining the incidence of cardiovascular disease between vaccinated and unvaccinated populations have noted that the risk reduction obtained from influenza vaccination is comparable to that of statin therapy or smoking cessation[117,118]. Recent influenza infection is also a highly significant risk factor for serious complications from diabetes, autoimmune disorders, and age-related neurological disorders and physical decline[119].

Studies conducted in the US indicate influenza has the greatest economic burden of any vaccine preventable disease except for Covid-19, costing 6.3-25.3 billion dollars annually from direct health expenses and work absenteeism alone[120]. Similarly, in Canada hospitalization with influenza is estimated to have an average per person cost of $14,612, rising to $39,477 when care in the ICU is required[121].

While seasonal influenza endemics are caused by both IAV and IBV, pandemics are caused solely influenza A. Pandemics occur sporadically, and their severity is unpredictable. The mortality of the 2009 Swine flu pandemic was within the upper estimates for an unusually severe seasonal strain[8,9]. By contrast, upper estimates for the death toll of the 1918 Spanish flu are comparable to the combined loss of life from every armed conflict of the 20th century[106]. The intensification of human agriculture and increasing population density is predicted to increase opportunities for novel subtype recombination and emergence of pandemic strains[30]. Perhaps more concerning, gain of function experiments on existing highly pathogenic (mortality of 20-50%) avian H5N1 subtypes have shown the potential for sustained human transmission with a handful of point mutations[10]. Capacity building to combat seasonal influenza and emerging pandemics is thus a vital public health imperative.

2.8 Current influenza vaccines

Influenza vaccines have been the primary public health intervention to blunt the impact of seasonal and pandemic influenza since their development in the 1940s. The efficacy of influenza vaccines, which as previously mentioned must be reformulated annually to adjust for antigenic drift and changes in dominantly circulating strains, has fluctuated between 10-60% since 2005[7]. While far from perfect, in the US alone during the 2017-2018 flu season influenza vaccines were estimated to have prevented 7.1 million illnesses and over 8,000 deaths[122]. The majority of currently available vaccines are quadrivalent, incorporating two influenza A viruses, one each from the H3N2 and H1N1pdm09 subtypes, and two influenza B viruses, one each from the B/Victoria and B/Yamagata lineages[123]. Trivalent formulations incorporating a single influenza B lineage are also sometimes used. These vaccines fall into three broad categories: Subunit inactivated, split virion inactivated, or live attenuated.

Subunit vaccines contain purified HA and sometimes NA protein (though only HA content is used as a release criteria) and are delivered by intramuscular injection[51]. Subunit vaccines may be produced from whole virions or recombinantly. In the case where subunit vaccines are derived from whole virions, vaccines contain both HA and NA and a dose of 15µg of HA per strain per dose is typically used[124]. Formulations with MF59 as an adjuvant are available for vaccination of the elderly or those with an otherwise suboptimal immune response[124]. Recombinantly produced subunit vaccines contain engineered HA "rosettes" that mimic the structure of the active HA trimer[125]. These vaccines do not contain NA but contain a much higher (45µg) dose of HA[124].

Split virion inactivated vaccines are produced by the inactivation of live virus with formalin or β-propriolactone, followed by disruption of the virion structure with detergents. The envelope and vRNA are largely purified away, but a full complement of viral proteins are retained for use in the vaccine (though only HA is quantified). Similar to subunit vaccines, 15µg of HA per strain per dose is typically used, with 60µg high dose formulations available for vaccination of people >65 years of age[124].

Live-attenuated vaccines are generated from reassortment of wildtype influenza and a cold-adapted, temperature sensitive vaccine strain such as A/Leningrad/134/47/57/H2N2[126]. The resultant reassortant maintains the antigenic qualities of the wildtype strain but requires lower temperatures (~32C) for efficient replication. This restricts the virus to the upper airway where it is delivered via an intranasal spray [124]. As these vaccines do contain replication competent virus they are contraindicated for the very young, very old, and those with depressed immune systems.

2.9 Manufacturing of influenza vaccines

The development of each annual vaccine follows a one-year timeline beginning in January for the Northern hemisphere. The Global Influenza Surveillance Network operated by the WHO selects strains that are predicted to be dominant in the next influenza season during early February and March. Prediction of circulating strains is based on established global migration patterns of influenza, with new strains typically arising in SE Asia and spreading globally in a seasonal manner[127]. For all vaccines except recombinant HA vaccines, wildtype strains must first be reassorted with a high yield vaccine strain before bulk amplification. Internal gene segments are derived from the vaccine strain such as PR/8 or a temperature sensitive variant, while the HA and NA are derived from the wildtype strain[128]. This process is carried put primarily to increase vaccine yield, but also attenuates the pathogenicity of the wildtype virus and makes the manufacturing process safer. Even in cases where the bulk amplification of the virus is accomplished using cells, reassortant preparation is almost exclusively done using limiting dilution in eggs to isolate clonal viruses. This process is slow, requiring several months and is a major bottleneck for vaccine production[128].

Bulk production of monovalent antigens is then carried out over a 3-4 month process, using either cultured cells or pathogen-free embryonated chicken eggs. Egg production simply involves inoculation of a chicken egg via a small hole in the eggshell. The virus then replicates and accumulates in the allantoic fluid of the egg, with roughly one egg is used for each finished dose of vaccine[129]. Bulk amplification of influenza using cells is possible with any number of cell lines, including MDCK, Vero, Per.C6, and HEK-293[130,131]. Of these, MDCK and Vero have produced licenced vaccines, with only MDCK currently being used[124]. The preference for MDCK cells over Vero is largely due to the successful adaptation of MDCK cells to suspension

culture, allowing scalable production in simple stirred-tank bioreactors[132]. Recombinant subunit vaccines, which are not produced from replicating influenza, are generated use a baculovirus expression system and derivatives of the Sf9 insect cell line[125].

Follow bulk amplification, blending of antigens, fill and finish, and vaccine distribution comprise the remainder of the vaccine production timeline. A similar timeline is used to produce vaccines for the Southern hemisphere but beginning in July[128].

2.10 E(gg)xit and the move towards wholly cell-based vaccine manufacturing

The vast majority (85-90%) of seasonal influenza vaccines are currently produced using embryonated chicken eggs, a technology has remained largely unchanged since the 1940s[6]. This is due in part to the difficulty and expense in scaling any other system to meet the annual demand for over a billion vaccine doses. Eggs are also able to produce relatively high yields at a low cost. However, evidence suggests that significant improvements in vaccine efficacy could be made by switching to an entirely cell-based production process. Passaging influenza through eggs induces antigenic drift as the virus adapts to an avian host[6,133,134]. As a result, egg-based vaccines exhibit a 15-20% decrease in protection rate compared to similarly formulated cell-based vaccines[5,134,135]. Cell-based platforms can also drastically reduce production lead time by accelerating seed stock reassortment via reverse genetics[11]. This, in turn, reduces the chance of major changes in circulating strains occurring between initial strain selection and vaccine release, an issue that rendered the 2014-2015 seasonal vaccine largely ineffective[136,137]. Other advantages of cell-based seasonal vaccines include a lack of allergen contamination and better growth of certain strains[5]

While eggs are a passable platform to produce seasonal influenza vaccines, they are completely inadequate to deal with influenza pandemics. The reasons for this are primarily rooted in differences in the epidemiology of seasonal versus pandemic influenza strains. Whereas seasonal influenza follows a predictable migration pattern, pandemic influenza outbreaks may arise at any geographic location[138]. The last two notable pandemics, for instance, were first detected in Mexico and Russia, respectively. The timing of pandemic influenza outbreaks also shows little correlation with the seasonal influenza season. The lack of preexisting host immunity to pandemic strains, combined with rapid international travel, often allows them to spread extremely quickly around the globe[139]. Consequently, production of vaccines during pandemic influenza outbreaks must be ready commence with virtually no warning and at any time of year. Egg-based manufacturing techniques used to meet the needs of seasonal influenza vaccines, reliant on months of prior warning and the timed availability of billions of pathogen-free, embryonated chicken eggs, are simply too slow and logistically inflexible to effectively respond to pandemic outbreaks. This was clearly demonstrated during the the 2009 "Swine flu" pandemic, where 123,000 to 203,000 people died and ~10% of the global population was infected in the nine months before a vaccine was widely available[8,9]. There are also concerns that it may be extremely difficult to develop vaccines against a future pandemic avian influenza strains using egg-based production due to premature death of the chicken embryo[140]. Cell-based vaccine production, by contrast, requires little lead time as cells in culture grow exponentially, enabling the rapid production of substrate to replicate viruses. Further development of cell-based vaccine production platforms is therefore necessary and urgent to prepare for future pandemics.

2.11 Process intensification cell-based influenza vaccine production platforms

Process intensification aims to increase the yield of cell-based influenza production platforms. This has the dual effect of increasing production capacity and lowing the cost per dose of finished vaccine. Current cell-based production platforms generally exhibit 4-10 fold lower volumetric yield than egg-based counterparts, are 40-100% more expensive, and are only able to supply <20% of current vaccine demand[5,6,12]. Process intensification of cell-based production platforms is thus required if cell-derived vaccines are to replace egg-derived vaccines.

Process parameters impacting influenza yield include multiplicity of infection (MOI), cell density, temperature, and media composition. Optimization of these parameters will vary depending on the cell line and culture mode, but some generalizations are still possible. While some viruses such as adenovirus are commonly produced at MOI=10, with influenza this tends to result in rapid apoptosis of cells in culture before appreciable amounts of virus can replicate[141]. Indeed, highest yields are generally obtained using extremely low MOIs (in the range of 10^{-1}-10^{-5})[142]. Like many cell-based virus production platforms, influenza replication in culture is susceptible to the so-called "cell density effect", wherein the cell specific productivity (CSP) and eventually volumetric productivity show an inverse relationship with cell density[143]. The HEK-293SF line, for instance, is capable of achieving cell densities exceeding $10x10^6$ cells/ml in simple batch culture but optimal yields are obtained at a density of just $1x10^6$ cells/ml[11]. The mechanism behind the cell density effect is complex and not entirely understood. The depletion of nutrients such as L-glutamine and glucose has been identified as a contributing factor, and thus most industry processes use a fed batch process to enable better CSP at higher cell densities[144]. The buildup of toxic metabolites such as lactate and ammonia, which also

contribute to the cell density effect, can similarly be ameliorated with perfusion feeding strategies to exchange spent media[145]. Human adapted influenza, as well as the mammalian cells the virus replicates in, grow optimally at 37C. However, volumetric yields in various processes can often be improved by culturing cells at temperatures of around 35C[146]. The somewhat counterintuitive effect stems from the fact that influenza stability is inversely related to temperature, and so low temperatures allow virus accumulated in the media to remain viable until harvest. Studies examining different growth media have also been show in impact vaccine yield. Ideally medias are defined in composition to minimize batch to batch variability. A lack of animal derivatives, particularly fetal bovine serum (FBS) and similar products, is also highly desirable both to minimize the chance of contamination with zoonotic pathogens and to reduce costs[147]. Manufacturers typically don't disclose the precise composition of different medias, making it difficult to draw any more detailed conclusions.

Equally important to the individual process parameters above is the manner in which they are optimized. The most basic method is On Factor At a Time (OFAT). While simple to carry out, OFAT assumes that each parameter of the vaccine production process is entirely independent, which is demonstrably not the true[148]. Alternatively, one might use a factorial optimization processes wherein all possible combinations of each parameter are tested. While this allows precise determination of all higher order interactions between parameters, it quickly becomes unfeasible as the number of parameters increases, particularly in the case where parameters are not binary and instead consist of multiple "levels" (e.g. a range of cell densities from 1-10x10^6 cells/ml). For this reason, fractional factorial or Design Of Experiments (DOE) approaches are popular, allowing the elucidation of main effects and low-order interactions while

running a fraction of the experiments of a full factorial design by exploiting the sparsity of effects principle[149].

2.12 AAV biology

AAV are among the smallest viruses known, with the non-enveloped icosahedral capsid measuring approximately 20nm. The linear 5kb, ssDNA genome encodes only two open reading frames, termed REP and CAP, flanked by inverted terminal repeat (ITR) sequences[13]. CAP, using splice variants and alternative promotors, is translated into structural capsid components. REP is translated into non-structural factors with helicase, DNA binding, and endonuclease activities that facilitate genomic replication and integration [150].

Initial attachment of AAV to the cell surface is mediated by several receptors, but canonically heparan sulphate proteoglycan (HSPG) [151]. Work by Pillay *et al.* strongly suggests that secondary interaction with the poorly characterized transmembrane protein KIAA0319L is essential for AAV entry[152]. AAV is imported into the nucleus as a whole viral particle, followed by virion uncoating and second strand synthesis to generate a transcriptionally active viral genome[153]. Nuclear entry, virion uncoating, and second strand synthesis are all inefficient processes, and might contribute to the low efficiency of AAV transduction[153-155].

The next stage of the viral lifecycle is dependent on whether the host cell is coinfected with a helper virus such as adenovirus or herpesvirus. If no helper virus is present, the virus will remain latent in the cell, persisting as a stable episome or integrated into the host genome [150]. Episomes form through the recombination of the viral ITRs, resulting in stable circularized dsDNA as genomic monomers or concatomers[14]. Genomic integration, mediated by REP proteins, is targeted to the AAVS1 loci on the long arm of chromosome 19 [156].

2.13 AAV vectors

Though a recently developed vector platform relative to adenovirus or lentivirus, AAV have been rapidly adopted by clinicians, with 149 completed or ongoing clinical trials and 5 approved therapies for various forms of cell and gene therapy[157]. Their excellent safety profile, ability to target a wide range of tissues, and stable transgene expression makes AAV vectors particularly useful for applications requiring in vivo administration for the treatment of hereditary disorders[15]. A major drawback of these vectors is low transduction efficiency, often necessitating massive doses (up to 2×10^{13} VG/ kg for systemic administration) to reach therapeutic transgene expression levels. Current manufacturing techniques are unable to efficiently produce AAV vectors in this quantity, with some approved therapies costing over a million USD[17]. Increasing interest in the widespread adoption of AAV vectors from clinicians and industry has only served to further highlight this limitation,

First generation AAV vectors are generated by removal of all elements of the viral genome except the flanking ITR sequences, which are the only elements required *in cis* to ensure transduction in target cells[158]. Vectors are generated by expressing CAP, REP, and with the minimal elements from a helper virus *in trans*. The resultant vector is replication incompetent and has a packaging capacity of approximately 4kb[13]. Since genomic integration is mediated by REP, vectors are unable to incorporate into the genome and persist as extragenomic episomes[14]. Episomes do not co-replicate with the cell, and thus expression typically declines as a function of cell division rate[159].

2.14 Manufacturing of AAV vectors

2.14.1 AAV production by transient transfection of mammalian cell culture

AAV can be generated via transfection with a triple plasmid system composed of a transfer, packaging, and helper plasmids into a number of mammalian cell lines such as HEK-293. The transfer plasmid encodes the transgene cassette flanked by ITR sequences, the packaging plasmid encodes the CAP and REP genes, and the helper plasmid encodes the minimal helpervirus factors (commonly adenovirus E1, E2a. E4, and VA RNA) (Daya 2008). Alternatively, REP and CAP may be split onto separate plasmids to allow for optimization of viral protein ratios, which has shown considerable success in increasing yield [160]. In terms of transfection reagents, polyethylenimine (PEI) is generally preferred as it is relatively inexpensive compared to cationic lipids, but much more robust compared to calcium phosphate methods.

Production of AAV vectors by transient transfection suspension cells is commonly employed to generate vectors for clinical trials, and when properly optimized can generate specific yields of up to 10^{14}VG/L on a 20L bioreactor scale [19]. However, producing AAV by transient transfection is generally not considered viable on an industrial scale due to the spiraling costs associated with transfection reagents and plasmid DNA.

2.14.2 AAV production by mammalian packaging cell lines

In order to avoid the costs associated with transient transfection and facilitate scale-up of AAV production, stable packaging cell lines derived from suspension adapted HeLa cells have been generated[161]. Packaging cell lines incorporate CAP, REP, and the transgene cassette flanked by ITR sequences on one or more plasmids that are retained via drug selection. Cells are

then infected with replication competent adenovirus to supply helper virus proteins and initiate AAV vector production.

Packaging cell lines show volumetric yields comparable to transient transfection at a much lower cost, and have been shown to be scalable in stirred tank bioreactors up to 250L[161]. However, contamination with replication competent adenovirus complicates downstream processing. This can be mitigated somewhat by using cold adapted adenovirus. These viruses express sufficient quantities of helper virus proteins to allow vector production, but are attenuated in their own reproduction[20]. The natural tendency for AAV to repress adenovirus replication presumably also helps to limit adenovirus contamination[162].

2.14.3 AAV production by baculovirus expression vector system (BEVS)

The baculovirus expression vector system (BEVS) is a platform used to produce biologics in insect cells, in this case *Spodoptera frugiperda* (Sf-9) cells derived from the Fall Army Worm[163]. Briefly, separate recombinant baculovirus vectors expressing CAP, REP, and the transgene cassette flanked by ITR sequences are used to infect bioreactor cultures of Sf-9 cells. Expressed proteins and vector genomes then self-assemble into mature AAV vectors[164]. Yields are typically quite high, with optimized processes producing approximately 10^4VG/cell on a 200L scale[165]. Recently advances have simplified the system to the use of one baculovirus that expresses all necessary factors, and also shown that the system can manufacture a range of AAV serotypes[166].

The system offers many advantages; Sf-9 cells can be grown to densities exceeding $2x10^7$ cells/ml in animal component free suspension culture[166]. The system is also quite safe in that no virus with human tropism (save the vector itself) is used during the production process. The use

of a non-human derived cells also all but eliminates the possibility of adventitious agents contaminating the culture. The BEVS platform has been shown to be highly scalable for the production of AAV vectors, as evidenced by its use in the production of Glybera, a recently approved AAV vectored gene therapy for lipoprotein lipase deficiency[167].

2.15 CRISPR/Cas9 genome editing

Clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR associated proteins (Cas) are components of a prokaryotic adaptive antiphage system. Though many of the genetic elements of CRISPR operons were first identified in 1987, it wasn't until 2012 that Doudna and Charpentier demonstrated that they could be adapted as programmable nucleases[168]. Since then, CRISPR has effectively replaced previous genome editing tools such as TALENs and been expanded to a host of applied and research applications. Speaking generally, CRISPR editing tools consist of an effector nuclease enzyme or holoenzyme complexed with RNA that determines nuclease specificity towards either an RNA or DNA target. Many prokaryotes encode variations of CRISPR nucleases including Cas12a, Cas13, and the most commonly used Cas9[169].

As a class II CRISPR/Cas system, the Cas9 nuclease consists of only one subunit that can be readily expressed in a wide range of prokaryotic and eukaryotic cells. In the endogenous Cas9 system, the protein complexes with both an 18-20 bp CRISPR RNA (crRNA) and a longer trans-activating RNA (tracrRNA). The crRNA is unique and complimentary to the target sequence, while tracrRNA acts as a generic binding scaffold for the Cas9 nuclease[168]. In recombinant systems these two RNA species are commonly fused into an sgRNA. Once the Cas9-sgRNA complex is formed it will bind DNA sequences complimentary to the 18-20 bp guide sequence. Following the recognition of a short downstream consensus sequence known as

a protospacer adjacent motif (PAM), Cas9 will cleave the target site and induce a dsDNA break 3-4 bp upstream of the PAM sequence. The requirement for a downstream PAM sequence is an adaptation that evolved to prevent self-cleavage of CRISPR arrays in the endogenous Cas9 system. The PAM sequence varies between species, but the majority of Cas9 systems use the enzyme derived from S. *pyrogenes* which has a PAM sequence of NGG[170].

CRISPR/Cas9 can be used to induce both gene knockins and gene knockouts depending on the DNA repair pathway exploited. For knockouts, no repair template is supplied, and the cell is forced to use the error-prone nonhomologous end joining (NHEJ) pathway. Cas9-sgRNA complex will continue to cleave the target site until the NHEJ pathway introduces an indel at the cleavage site. If these indels occur within an exon of a protein coding gene, this results in a frameshift mutation and loss of gene expression[171]. sgRNAs designed for gene knockout thus normally target an early exon of that gene. The median size of CRISPR/Cas9 indels is 1bp in length, but this is not an absolute and much larger indels can occur at high frequencies depending on the target site[172]. In the event a repair template is cotransfected or otherwise introduced with the Cas9 nuclease the cell will use the homology directed repair (HDR) pathway and incorporate the repair template at the target site[173]. This allows the knockin of virtually any sequence (from SNPs to insertions of thousands of nucleotides) as long as it is flanked by sequences homologous to either side of the cleavage site[174]. This process is similar to homologous recombination techniques that have been used in bacteria for decades but is many orders of magnitude more efficient due to the presence of a dsDNA break proximal to the integration site.

2.16 CRISPR/Cas9 as a genetic screening tool

Genetic screens involve the perturbation of gene expression using a library of constructs to generate a heterologous array or pool of mutants. These mutants then undergo selection or are

assayed to determine the genetic contribution to a given phenotype. CRISPR screens use a library of sgRNA, each targeting a different locus, to create a panel of mutants. Genetic screens have been previously conducted using short hairpin RNA (shRNA) or random mutagenesis but are greatly improved with CRISPR for several reasons. Unlike shRNA, which "knocks down" or reduces the expression of a given gene, CRISPR/Cas9 screens can completely ablate target gene expression with high efficiency, allowing the elucidation of phenotypes that would otherwise be masked by low-level expression[175]. shRNA is also only able to target genes that produce an RNA product, whereas CRISPR can be used to target both transcribed and non-transcribed genetic elements[176]. The use of engineered Cas9 variants also allows for genetic perturbations beyond simple gene knockouts (described further in section 2.17).

CRISPR screens may be carried out in one of two formats: arrayed screens or pooled screens. In an arrayed screen subpopulations of knockouts are physically separated, typically in wells of a microtiter plate. By contrast, in a pooled screen a heterogeneous pool of knockouts is cultured together in one or more culture vessels. The main advantage of an arrayed screen is that they're far more flexible in terms of readout; virtually any physical or chemical characteristic of the cells or media that can be assayed in a high-throughput manner can be used[177]. This includes the yield of excreted metabolites or viral biologics. On the other hand, pooled CRISPR screens must be designed such that the pool undergoes some form of selection to change the abundance of mutant subpopulations[178]. Samples of the pool that have undergone different treatments are then deep sequenced to identify which mutants were enriched/depleted by selection. The reliance on deep sequencing for pool deconvolution means that the cells in a pooled screen must also be barcoded to identify which mutation they carry. This is accomplished by delivering the CRISPR/Cas9 expression construct with a lentiviral vector, which integrates into the cell

genome. The unique crRNA potion of the encoded sgRNA then acts as the barcode[179]. Arrayed screens also commonly use lentiviral vectors as they require little optimization to achieve high transduction efficiencies, but as mutant populations aren't mixed there is no requirement to barcode cells. Thus, arrayed screens may alternatively employ transfection with plasmids or Cas9/sgRNA nucleoproteins if cells are not amenable to lentiviral transduction[180].

Though arrayed screens offer flexibility in terms of readout and CRISPR/Cas9 delivery method, this flexibility comes at a high cost. Arrayed screens can require thousands of microtiter plates and hundreds of thousands of individual manipulations to achieve genome-wide coverage, necessitating the use of automated liquid handling and other high-throughput robotics[177]. The enormous expense and time required to run a genome-wide arrayed screen, exceeding the resources of all but the most well-funded groups, has doubtlessly contributed to efforts in recent years to expand the utility of pooled screens. In addition to lower costs, pooled screens are not without their inherent advantages; the pooling of all cells together in a small number of culture vessels (compared to the thousands required for an arrayed screen) ensures nearly identical culture conditions for all cells in the pool, reducing variability. Pooled screens also offer great flexibility and fine control in terms of how the cells are cultured, as opposed to arrayed screens which can only feasibly be run in microtiter plates. This difference is critical given that the entire field of process development is predicated on the idea that cells behave very differently depending on cell density, oxygen transfer kinetics, batch versus perfusion bioreactor modes, adherent versus suspension culture, etc. Pooled screens are thus uniquely suited to elucidate context-specific genetic contributions to phenotype, highlighting their potential as a valuable tool for process development if their inherent limitations can be overcome.

2.17 Design considerations of pooled CRISPR/Cas9 screens

Careful design of a pooled CRISPR/Cas9 screen is critical to avoiding screening artefacts, minimizing signal to noise ratio, and ensuring repeatability. CRISPR libraries are capable of inducing different genetic perturbations depending on the goals of the screen. The most common perturbation is gene knockout, which employs unmodified Cas9 and the NHEJ mediated frameshift mechanism described in section 2.15. CRISPR libraries are also able to induce overexpression of genes. These activation libraries employ a catalytically dead Cas9 (dCas9) mutant that maintains high affinity for target sequences but is unable to catalyze a dsDNA break. This dCas9 is fused to a strong transcriptional activator such as 4x Herpes Simplex Viral Protein 16 (VP64) that enhances expression of nearby genes[181]. dCas9 is also used in gene repression libraries. In this case rather than a transcriptional activator, dCas9 is fused to a transcriptional repressor and induces localized reduction in transcription[182]. Repression libraries are useful in screens on cell lines that are particularly sensitive to dsDNA breaks. More exotic perturbations such as modification of RNA splicing, small insertions using CRISPR-prime editing, and promotor methylation are also possible with various Cas9 fusion proteins[183-185].

The selection of which model or cell line to run a screen on has important implications for screen design and is often intrinsically linked to library selection. Though it is theoretically possible to construct a pooled CRISPR library for any model with an annotated genome, time and monetary constrains often limit the choice to commercially available libraries for well-studied models such as mouse, human, or E.*coli*[186]. Additionally, the scope of the screen (eg. genome-wide or a subset of genes) is in part dictated by the number of cells that can reasonably sustained in culture. While not an issue for immortalized cell lines, if primary or otherwise

difficult to culture cells are used it may be difficult to maintain adequate library coverage (a concept discussed in detail bellow). While immortalized cell lines are easy to maintain in culture, many of them are aneuploid. Ploidy is proportional to the number of dsDNA breaks that must be generated to knock out a given gene, which can be stressful for the cell and lead to non-specific effects on cell growth[187].

An important concept in pooled screen design is that of sgRNA coverage. Coverage refers to the absolute number of a unique guide in the sgRNA pool, and the total number of cells that must be maintained to elucidate statistically significant changes in the enrichment or depletion of corresponding mutant subpopulations. Pools constructed from validated libraries are assumed to be evenly distributed during initial library transduction, and thus the required number of cells to maintain coverage is simply the desired coverage (generally 200-500x during pool generation and maintenance) multiplied by the number of unique sgRNA in the library[179]. Failing to maintain coverage during the generation and maintenance of the pool can result in population bottlenecks and subsequent loss of mutant subpopulations within the pool[188]. Note that guides can also be lost when the pool is kept in culture over a high number of passages, both due to random population drift and differences in growth rate between various mutants[188]. Untreated control conditions, where the library is expected to be complete and evenly distributed, should be sampled at similar coverage as pool maintenance. Calculation of coverage (or the number of cells to collect) for enriched/depleted conditions a far more complicated question that depends on whether the screen is positive or negative and the signal to noise ratio of the screen. A positive screen where a small number of sgRNA species are expected to be highly enriched over background (e.g. drug selection) can collect comparatively few cells and still have an adequate sample size to quantify fold-changes in sgRNA abundance for these hits.

On the other side of the scale, a negative screen trying to detect the depletion of a few sgRNA on a very noisy background will likely have to sample at higher coverage[179]. The number of cells to sample is also dependent on the tolerance of the screen for false negatives; often the goal of the screen is to identify only the most significant hits, and thus underrepresentation of weak hits is not a concern. Often pilot experiments are necessary to determine sampling parameters, though software tools such as CRISPulator can also assist in this[189].

As previously stated, in pooled screens measurement of a particular phenotype must be paired with enrichment or depletion of cells in the pool that display that phenotype. Not surprisingly, pooled CRISPR screens have been often used elucidate fitness genes within the cell[190-192]. Pooled screens also naturally lend themselves to studies of cytotoxic drug resistance, or any trait that directly impacts cell survival. In an effort to expand the utility of pooled CRISPR screens, many studies have employed more sophisticated enrichment/depletion methods. One of the most common is the use of FACS to enrich populations displaying a fluorescent marker, either expressed from a reporter or applied through fluorophore staining[193]. If a reporter cell line is used, extensive validation is necessary to ensure that reporter expression correlates strongly and specifically with the phenotype of interest. Staining with fluorophores or antibody-fluorophore conjugates offers a more direct method of quantifying a given phenotype and circumvents the need to construct a custom cell line or reporter construct. However, staining of intracellular targets may necessitate cell permeabilization and fixation which can interfere with measurements and complicate extraction of DNA for deep sequencing[193]. While FACS is a powerful tool for pooled sgRNA screening, depending on the screen design the pool to be enriched may consist of hundreds of millions of cells. In cases where sorting this many cells is unfeasible, alternative methods such as antibodies conjugated to magnetic beads may be used[194].

In addition to conventional readouts for pooled screens, consisting of a raw count of the abundance of each sgRNA in the pool, recent studies have demonstrated that more complex readouts are possible. Single cell CRISPR sequencing (scCRISPR-seq) leverages single cell transcriptomics to generate a transcriptome profile for each cell in the pool. sgRNA are also captured during the RNA sequencing step, allowing a given knockout to be linked to each transcriptome profile[195]. Spatial imaging combined with *in situ* sequencing of sgRNA inserts is able to generate a similarly rich data set[196]. These so-called "high content" CRISPR screens offer exciting potential to expand the utility of pooled screens, but currently practical limitations on sequencing capacity and imaging make whole genome coverage an expensive endeavor.

2.18 CRISPR/Cas9 screening as a tool to intensify the production of influenza vaccines and other viral biologics

Deletion of innate immune "restriction factors" that defend the cell against viral infection could create a more permissive environment to produce viral biologics. Genome-wide CRISPR/Cas9 screening offers a means of identifying these factors with the goal of increasing the yield of cell-based influenza vaccines and other viral biologics such as AAV. In theory a review of previously conducted screens should be sufficient for this purpose. However, the use of data from the relevant literature to identify and rank candidate restriction factors is mired by a lack of agreement between the results of independent studies. For instance, three genome-wide studies to identify influenza restriction factors by Heaton *et al.* (2017), Tripathi *et al.* (2015), and Sharon *et al.* (2020) showed overlap of less genes than would be expected by random chance (see Figure 2 below)[1,197,198].

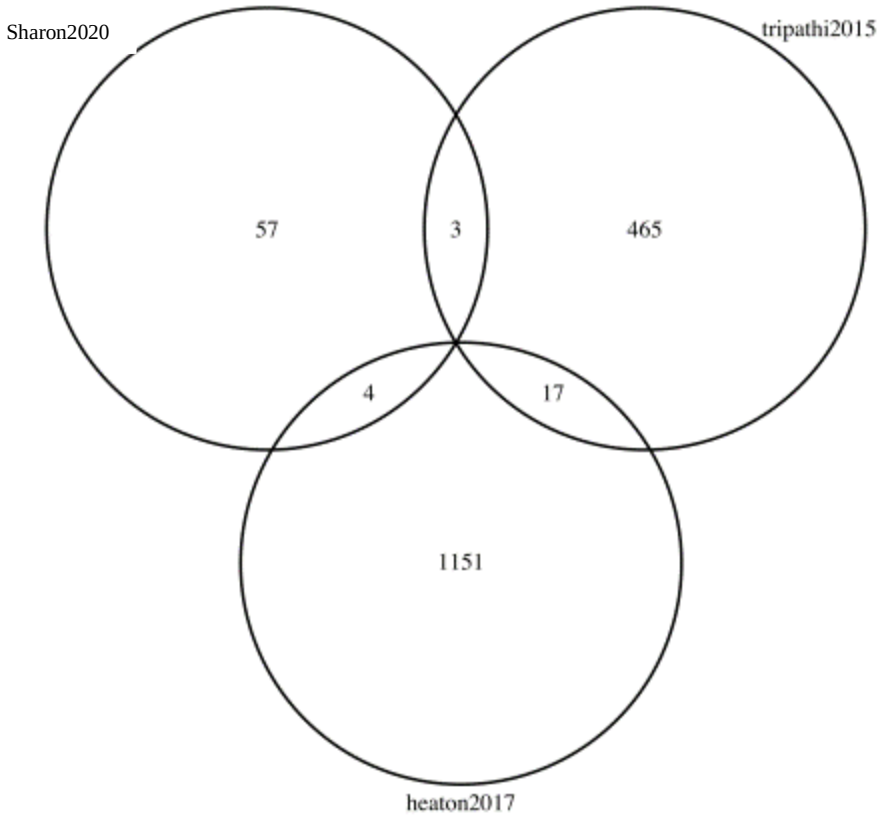


Figure 2. Comparison of overlapping gene hits for genome-wide screens conducted to identify influenza restriction factors. See Sharon *et al.* (2020) supplemental data S1 for detailed analysis[1].

These results reflect the importance of cell line, culture parameters, strain, and other contextual factors in determining the host response to viral infection. This is particularly true for cell lines used to produce viruses such as MDCK, Vero, and HEK-293, which diverge heavily from primary cells and each other in terms of their innate antiviral response[12,23]. The question of which gene knockouts will give the greatest increase in virus yield is then best probed on a case-by-case basis, using a genome-wide screen conducted on the relevant host cell line.

Wu et al. (2017) and van der Sanden et al. (2016) have attempted to use genome-wide screens to identify viral restriction factors in Vero cells and improve polio vaccine yield, with variable success[199-201]. In a retrospective analysis, Hoeksema *et al.* (2018) identified several factors that may explain the difficulties encountered[201], particularly the issue of using human-specific libraries in non-human primate derived cell lines. Both screens also used shRNA-based methodologies to probe for restriction factors. Difficulties can be encountered translating the results of shRNA screens into CRISPR/Cas9 knockout cell lines [201]. A further possible source of error is the arrayed format of previous screens, which as previously mentioned tend to result in high variability[177,178].

Genome-wide screens to improve the yield of AAV have also been conducted. Barnes *et al.* (2021) used a CRISPR/Cas9 activations screen coupled with an iterative enrichment method to identify ITPRIP and SKA2 as AAV host dependency factors. Overexpression of both factors resulted in a 4-fold increase in AAV vector yield. Further investigation showed this was primarily due to an increase in the full to empty ratio of AAV capsids[202].

3.0 MATERIALS AND METHODS

3.1 Cell culture transfection

HEK-293SF cells (RRID accession: CVCL_4V94) were obtained from the National Research Council of Canada [203]. In all cases, cells were maintained in serum-free suspension at a density of 0.5-4.0×10^6 in Hyclone Hycell TransFX-H media (GE Healthcare) supplemented with 4 mM Glutamine and 0.1 % Pluronic F68 (Sigma-Aldrich). Cells were incubated in vented PETG shake flasks (Corning), spun at 110rpm in a 37ºC incubator in a humidified 5 % CO_2 atmosphere. Cell counts and viability were determined by trypan blue exclusion assay using a Vi-CELL XR Cell Viability Analyzer (Beckman Coulter).

All transfections were performed at a cell density of 10^6 cell/ml using linear polyethylenimine with a mean molecular weight of 25000Da (Polysciences) complexed with plasmid DNA at a 1:2 ratio. The final concentration of plasmid DNA in all cases was 1ug/ml.

3.2 Lentiviral vectors

Lentiviral vectors were produced in the HEK-293SF cell line as described previously [204]. The psPAX2 (Addgene #12260) and CMV-VSVG (Addgene # 8454) plasmids were gifts from Didier Trono and Bob Weinberg, respectively [205].

HEK-293SF cells were transduced with the lentiviral vectors by spinfection at 1000 rcf for 45 minutes in media supplemented with 8 µg/mL of polybrene. Cells were then immediately resuspended in normal growth media to eliminate polybrene.

In all cases, infectious titer of lentiviral vectors was determined by ddPCR assay, using a protocol adapted for Baczak *et al.* (2015)[206]. Briefly, HEK-293SF were transduced with serial

10-fold dilutions of lentivirus. After 48 hours, genomic DNA was extracted from cells using the Purelink Genomic DNA mini kit (Thermo Fisher) according to the manufacturer's instructions. ddPCR targeting the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE) sequence of the vector genome was then used to determine the number of integrated vector genomes/cell. A parallel ddPCR assay targeting the albumin gene was used as a normalization control. See Apprendix 3 for primers and thermocycling conditions.

3.3 A/Puerto Rico/8/1934 influenza

PR/8 influenza (NCBI txid:211044) stocks were generated by reverse genetics in HEK-293SF cells. The process was described previously in Milián *et al.* (2017)[11]. The reverse genetics constructs used were a generous gift from Xuguang Li's group at Health Canada, and their construction was described previously in Neumann *et al.* (2005)[207].

To determine infectious particle titer, 10-fold dilutions of virus were used to infect HEK-293SF cells. After a 3-hour incubation, cells were stained for influenza NP expression and the percent of cells expressing influenza NP quantified by flow cytometry. See section 3.11 for details on staining and flow cytometry. Only cultures showing between 2-20% of cells infected were used for quantification to minimize error due to superinfection.

Viral genomes were quantified by extracting RNA from cell-free supernatants using the QIAamp Viral RNA Mini Kit (Qiagen) according to the manufacturer's instructions. RNA was then reverse transcribed with the iScript Select Reverse Transcription Kit (Bio-Rad) according to the manufacturer's instructions and using gene specific primers targeted to influenza genomic segment 7(M). ddPCR assay using the same primers was then used to determine viral genome copy number. See Appendix 1 for primers and thermocycling conditions.

All PR/8 infections were carried out at an MOI of 0.1 at a cell density of 10^6 cells/ml. PR/8 cultures were supplemented with 1 µg/mL of 6-(1-tosylamido-2-phenyl) ethyl chloromethyl ketone (TPCK) trypsin (Sigma-Aldrich) to allow proteolytic activation of HA.

3.4 PR/8$^{GFP\Delta HA}$ reporter influenza

The PR/8$^{GFP\Delta HA}$ virus was a generous gift from Alain Townsend (Oxford University). The cloning and production of this virus was previously described in Powell *et al.* (2012)[208]. Briefly, the coding sequence of the HA gene is removed and replaced with that of GFP. The virus is then propagated in an HA-expressing MDCK line. All PR/8$^{GFP\Delta HA}$ infections were carried out at an MOI of 5.

3.5 AAV Plasmids

For the generation of rAAV2, plasmids pAdDeltaF6 (Addgene #112867), pAAV-RC2 (cell biolabs inc.), and pAAV-CMV-GFP (Addgene #67634) were transfected in a 1:1:1 molar ratio. pAdDeltaF6 was a gift from James M. Wilson and pAAV-CMV-GFP was a gift from Connie Cepko [209]. Stocks of rcAAV2 were generated by equimolar transfection of pAdDeltaF6 and pAV2-Cla. pAV2-Cla, which contains a sequence identical to wildtype AAV2 apart from a point mutation to generate a Cla1 restriction site in the 3' UTR. pAV2-Cla was generously provided by Dr. Thomas Webber.

3.6 rcAAV2 infection

Initial stocks of rcAAV2 were generated via plasmid transfection and titrated by ddPCR. When using rcAAV2 infection as a positive control for capsid assembly, HEK-293SF cells at an initial cell density of 10^6 cell/ml were infected with rcAAV2 stocks and hAd5 and an MOI of 10.

3.7 Knockout pool generation

sgRNA for all knockout pools were randomly selected from corresponding Brunello library guides for that gene[210]. This sgRNA was then cloned into LentiCRISPR.V2 (Addgene #52961) using standard techniques and verified by Sanger sequencing. LentiCRISPR.V2 was a gift from Feng Zhang[24]. Lentiviral vectors produced using this construct were used to infect HEK-293SF at an MOI of 10. Following selection with 2µg/mL Puromycin for 48 hours, CRISPR editing efficiency was assessed using the TIDE webtool (v2.0.1)[211]. Cells were then incubated for a further 10 days to allow knockout phenotypes to manifest and recover the drop in cell viability arising from DNA cleavage. TBK1 knockout was further verified by Western blot using mouse αTBK1 (Santa Cruz, sc9085) at a 1/200 dilution. αβ-Actin (Sigma A1978) at a 1/1000 dilution was used as a loading control.

3.8 Genome-wide CRISPR knockout/activation pool generation

The Human Brunello CRISPR knockout pooled library and Human Calabrese CRISPR activation libraries (set A) were gifts from David Root and John Doench (Addgene #73178 and #92379, respectively)[210]. Lentiviral preparations obtained were used to transduce cells at an MOI of 2, incubated for 48 hours, and then selected with 2 µg/mL Puromycin for 48 hours. Cells were then incubated for a further 10-15 days to allow knockout phenotypes to manifest and recover the drop in cell viability arising from DNA cleavage. Aliquots of cells were then frozen in 10% DMSO. Cells were thawed and subcultured for 48 hours before use in the screen. At all points, a minimum representation of 300 copies/sgRNA was maintained. See Appendix 1 for primers used to generate the amplicon libraries.

3.9 Digital droplet PCR (ddPCR)

All ddPCR assays were carried out on the QX200™ Droplet Digital™ PCR System (Bio-Rad) using the QX200 EvaGreen Digital PCR Supermix (Bio-Rad) according to the manufacturer's instructions. Primers and thermocycling conditions for all ddPCR assays are listed in Appendix 1.

3.10 Microscopy

HEK-293SF cells were seeded at low confluence on a 35mm plate with coverslip (MatTek). 24h post seeding, the cells were transfected for the production of AAV2-GFP. 72h post seeding, the media was removed, and the cells fixed and stained as previously described [212]. AAV capsid monomers were detected with anti-VP1/VP2/VP1 (Progen) labelled with AlexaFluor 700 (Invitrogen), hAd5 E2A by anti-E2A labelled with AlexaFluor 350 (Invitrogen). Anti-E2A was a gift from Arnold J. Levine [213]. Antibodies were incubated at room temperature for 1 hour. Cells were then imaged using an Olympus IX-83 confocal microscope. The images were analyzed using FIJI v1.53 [214].

3.11 Flow cytometry

Flow cytometry to assess GFP expression was carried out on the Accuri C6 (BD Biosciences) instrument. Influenza NP protein expression was assessed by first fixing and permeabilizing cells with the BD Transcription Factor Buffer Set (BD Pharmingen). Infected cells were then identified by staining with a 1:50 dilution of mouse αNP-FITC (Thermo Fisher, clone D67J) for 50 minutes. In cases where GFP expression interfered with the use of the conjugated FITC fluorophore, goat αmouse-PE-Cy5.5 (Thermo Fisher, cat#M32218) was used as a secondary antibody, and any GFP/FITC fluorescence compensated for. To assess the percentage of cells

expressing assembled AAV capsids, cells were fixed and permeabilized as previously described [212]. Flow cytometry was carried out on the BD FACSJazz (BD Biosciences) or the BD LSRFortessa (BD Biosciences). Assembled particles of AAV2 were detected by staining with anti A20R (Progen) labelled with AlexaFluor 594 (Invitrogen). In all cases, data analysis was conducted using FlowJo (v.10).

3.12 FACS

In the case of the Brunello screen, cells were sorted live. Cells were sorted 33-40 hours hpi. Two fractions of cells were collected: the top 10% of GFP/AAV2 expressing cells ("high yield"), and a control fraction consisting of all infected cells (i.e. GFP positive). The number of cells collected was such that a minimum representation of 300 copies/sgRNA was maintained. In the case of the Calabrese screen, cells were fixed and permeabilized as previously described [212]. Assembled particles of AAV2 were detected by staining with anti A20R (Progen) labelled with AlexaFluor 594 (Invitrogen). Cells were sorted 48 hpi. Two fractions of cells were collected: the top 25% of AAV2 capsid expressing cells ("high yield"), and a control fraction consisting of all AAV2 capsid positive cells. A total of 1-2x10^6 cells were collected in each fraction.

In all cases, sorting was carried out on the FACSJazz (BD Biosciences) instrument using 1.5 drop yield mode and an event rate of 7000-9000 events/second.

3.13 Deep sequencing sample preparation and sequencing

Genomic DNA (gDNA) was extracted from cells using the JetQuick Blood and Cell Culture DNA Midiprep Kit (Thermo Fisher). PCR was then used to amplify the sgRNA inserts and append Illumina adaptors and hexamer barcodes to the amplicons. PCR was performed using the Q5 Hot Start High-Fidelity 2X Master Mix (New England Biolabs). Before creating the

amplicon library, ddPCR was used to assay the copy number of sgRNA inserts in extracted genomic DNA. Sufficient genomic DNA template was used to ensure a minimum read depth of 300 per sample. PCR products were then pooled, concentrated by isopropanol precipitation, and gel purified on a 2% agarose gel before sequencing. Gel extraction was carried out with the PureLink Gel Extraction kit (Thermo Fisher). The purified, barcoded amplicon libraries were then pooled and sequenced. Amplicons for the Brunello screen were sequenced on the Illumina HiSeq 4000 (Illumina) as 50bp single ended reads. Amplicons for the Calabrese screen were sequenced on the Illumina Novaseq (Illumina) as 100bp single ended reads. Amplicons for the validation of the MK lines were sequenced on the Illumina MiSeq (Illumina) as 300bp paired ended reads. See Appendix 1 for Illumina adaptor sequences.

3.14 Bioinformatics

Data to assess sequencing quality and read mapping, as well as quantification of the changes in sgRNA abundance between the high yield and control cell populations was carried out using the MAGeCK software suite (v.0.5.9.2) [215-217]. In this analysis, read counts were normalized using a set of 1000 nontargeting control sgRNA's that were provided in the Brunello and Calabrese library. The initial set of 1000 was reduced to 963 after the removal of outliers whose difference in normalized read counts between conditions were outside of the range $[Q1 - 1.5 * IQR, Q3 + 1.5 * IQR]$, where Q1, Q3, and IQR are the first quartile, third quartile, and interquartile range, respectively. To test for significance of sgRNA abundance between conditions, the MAGeCK "tool" test was used with additional parameters --remove-zero and --remove-zero-threshold set to "control" and "30 ", respectively. This removed 1760 sgRNA's that have an average read count in the control condition that is less than 30. All other parameters were left at the default setting. Of note, sgRNA-level p-values were adjusted using the Benjamini-Hochberg procedure, which

controls the False Discovery Rate at level $\alpha = 0.25$. To obtain gene-level p-values from multiple sgRNA's targeting a single gene, version 0.5.9 of the modified Robust Rank Aggregation (RRA) algorithm, named α-RRA, was used [218]. Data analysis of Illumina MiSeq data to validate the MK clones was carried out as previously described using the CRIS.py python package[219]. GO analysis was carried out using the Metascape webtool (v.3.5), and plotted using default parameters[220]. Plots were created using Cytoscape (v.3.7.2)[221]. Protein complex enrichment analysis was carried out using the COMPLEAT webtool (v.1.0)[222]. Genes were submitted as a single list using lfc as a ranking metric.

3.15 Statistical analysis

Except where otherwise stated, the coefficient of correlation and statistically significant differences between two groups of means were determined by Student's t-test using Prism GraphPad (v.6.01). Error bars in figures represent SEM.

4.0 A POOLED GENOME-WIDE CRISPR/CAS9 KNOCKOUT SCREEN TO IDENTIFY AND RANK INFLUENZA HOST RESTRICTION FACTORS IN HEK-293SF

4.1 Screen design

This chapter details the results of a pooled screening strategy using a reporter virus coupled with a FACS based selection method to identify and rank host restriction factors for A/Puerto Rico/8/1934 H1N1 (PR/8) influenza in HEK-293SF cells. The screen components and workflow are described in Figure 3.

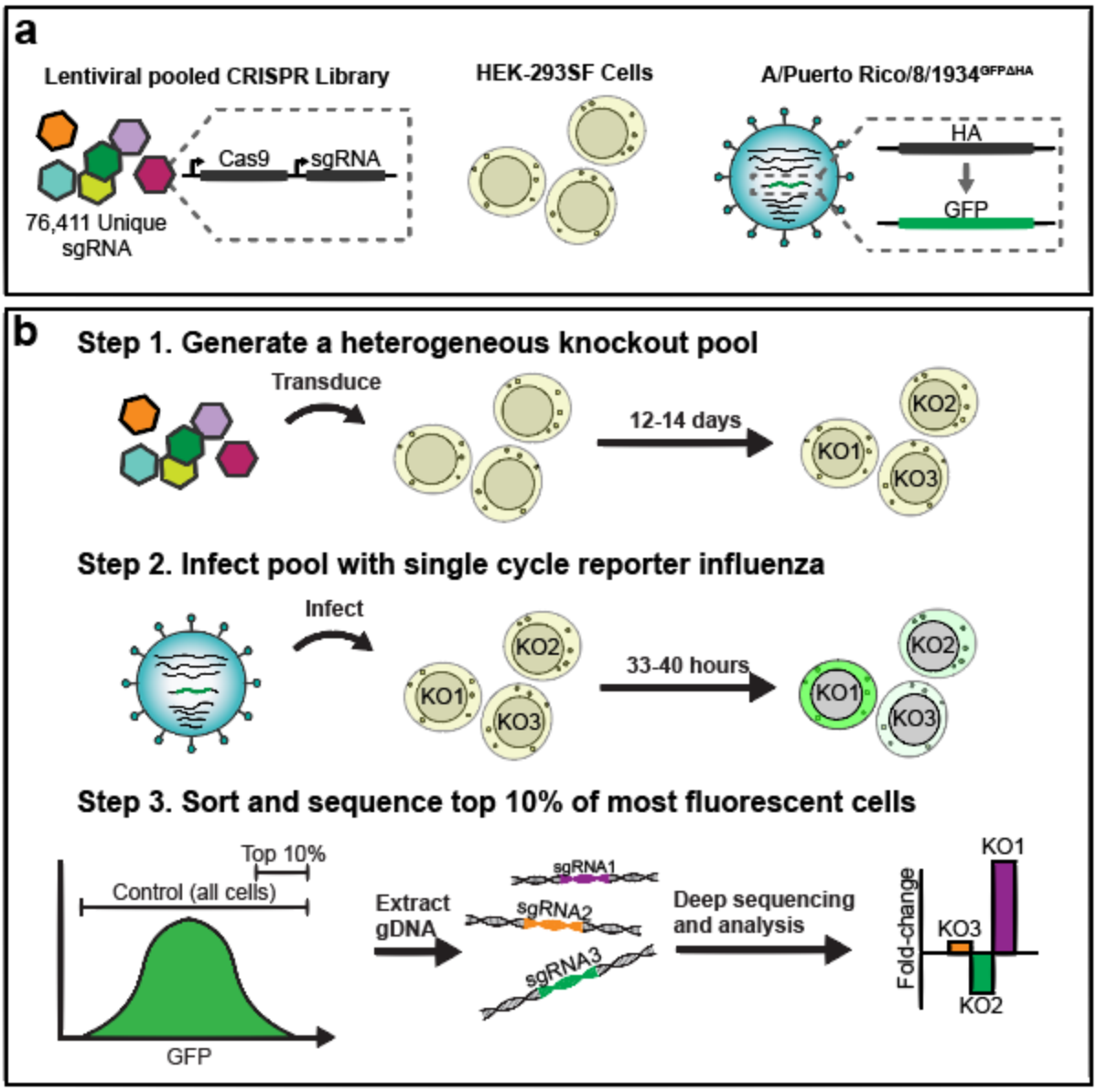


Figure 3. Illustration of pooled screen to identify and rank putative influenza restriction factors

The components of the screen are shown: a) The lentiviral-vectored Brunello Human CRISPR Knockout Pooled Library, consisting of a pool of 76,411 unique sgRNA as well as 1000 non-targeting controls, each with a CRISPR/Cas9 expression cassette packaged in lentiviral vectors; The HEK-293SF cell line; The A/Puerto Rico/8/1934 $^{GFP\Delta HA}$ (PR/8$^{GFP\Delta HA}$) reporter virus, wherein the HA coding sequence on genomic segment 4 of the virus has been replaced with the coding sequence of GFP. b) The workflow of the screen is shown, including library transduction, infection with reporter influenza, and selection of a "high yield" knockout population via FACS. The high yield fraction is composed of the top 10% of all GFP expressing cells. A control population, composed of all GFP-expressing cells, is also collected. Deep sequencing is then used to determine the abundance of integrated sgRNA expression cassettes in the genomic DNA (gDNA) of the two fractions. Fold-change enrichment of sgRNA species in the high yield fraction relative to the control is then used to identify and rank putative restriction factors.

The Brunello Human CRISPR/Cas9 Knockout Pooled Library (Figure 3a) consists of a pool 76,411 unique sgRNA, each with a CRISPR/Cas9 expression cassette, packaged in lentiviral vectors[210]. This library was selected due to redundant (~4x) coverage of all protein coding genes and the inclusion of extensive (n=1000) non-targeting controls, which are useful for normalization during downstream analysis[210]. The HEK-293SF cell line (Figure 3a), a serum-free and suspension adapted variant of the HEK-293 parent line, was selected due to its human origin and the availability of a well-annotated genome to ensure minimal off-target effects[203,223]. The HEK-293SF line is also highly attractive from a bioprocess standpoint and, though currently less popular than the Vero or MDCK lines for influenza vaccine production, has shown considerable potential as a scalable and high yield vaccine production platform[11,203,224,225].

The screen presented in Figure 3 was designed to enable a pooled screening format, using a single-cycle reporter virus and a FACS-based selection step. The A/Puerto Rico/8/1934$^{GFP\Delta HA}$ (PR/8 $^{GFP\Delta HA}$) reporter virus (Figure 3a) can enter cells and replicate wild-type infection kinetics for one infection cycle, but it is unable to produce infectious progeny due to the lack of the HA coding sequence, a critical entry and attachment factor [226]. Furthermore, as we will show in subsequent results, PR/8$^{GFP\Delta HA}$ transcribes GFP in direct proportion to wildtype PR/8 HA transcription, and GFP expression predicts relative wildtype viral titer. Critically, this allows the individual viral yield of each cell in the knockout pool to be approximated via GFP fluorescence intensity. High yield cells in the knockout pool can then be selected via FACS. The relative abundance of various CRISPR/Cas9 cassettes within the genomic DNA of these cells, integrated via the lentiviral vectors, can then be determined with deep sequencing to identify candidate restriction factors (Figure 3b). The pooled screening format enables the entire knockout pool of a given replicate to be contained in a single 3L shaker flask, ensuring identical conditions for all

cells during the screen. The use of a single-cycle infection virus avoids the confounding effects of secondary infection from neighboring cells.

4.2 PR/8$^{GFP\Delta HA}$ closely replicates PR/8 infection kinetics and effects on host cells for one infection cycle

The first step before conducting the screen was to investigate how closely PR/8$^{GFP\Delta HA}$ mirrors the infections kinetics of PR/8 for a single infection cycle. HEK-293SF cells were infected under conditions identical to those used in the screen. Over the next 72 hours, samples of cells were assayed for the percentage of cells infected, cell density, and cell viability. In addition, cellular levels of mRNA for genomic segment 4 (HA/GFP) were measured over 40 hours in the two cultures to determine whether the modified genomic segment 4 in PR/8$^{GFP\Delta HA}$ is transcribed at the same rate as that of wildtype PR/8. Results are shown in Figure 4.

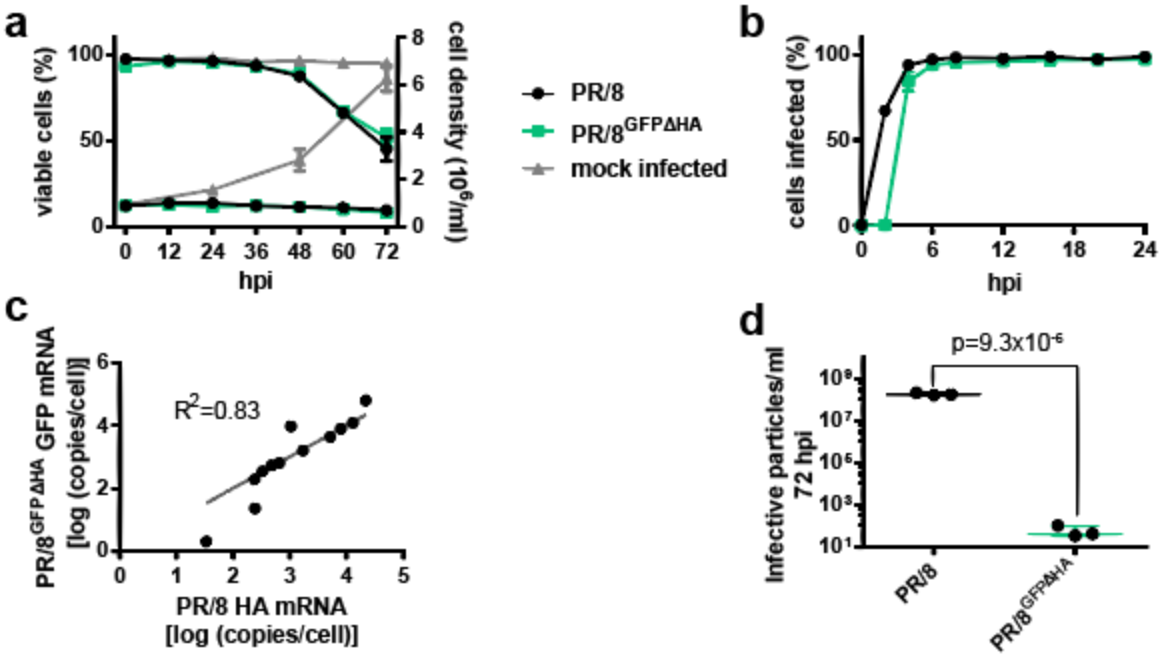


Figure 4. Comparison of PR/8 and PR/8 $^{GFP\Delta HA}$ infection kinetics and effect on cell viability and growth

a) Effect of PR/8 and PR/8$^{GFP\Delta HA}$ on cell viability and growth compared to mock-infected controls. b) Infection kinetics of PR/8 and PR/8$^{GFP\Delta HA}$, infected cells are defined as those expressing the influenza nucleoprotein, as measured by flow cytometry. c) Correlation between expression of PR/8 HA mRNA and PR/8$^{GFP\Delta HA}$ GFP mRNA over the course of a 40 hour infection. Cellular mRNA content was measured by strand-specific ddPCR. d) Comparison of infective viral particle titer in cell-free supernatant (CFS) of PR/8 and PR/8$^{GFP\Delta HA}$ infected cultures 72 hours hpi. In all experiments, datapoints represent the average of n=3 independent replicates run in parallel. Error bars represent standard error of the mean (SEM). Whisker plots show mean, range, and individual data points. R^2 value was determined by linear regression, and p-value was determined using 2-tailed Student's t-test.

Both PR/8$^{GFP\Delta HA}$ and PR/8 infected cultures showed identical effects on host cells, inducing immediate cell cycle arrest and declining viability 40 hours post-infection (hpi) compared to mock infected controls (Figure 4a). In terms of infection kinetics, PR/8$^{GFP\Delta HA}$ cultures showed a 1-2 hour delay in viral entry and expression of the influenza nucleoprotein (NP) as assessed by staining and flow cytometry (Figure 4b). However, within 4 hours, virtually all cells in both cultures were infected (Figure 4b), indicating that PR/8$^{GFP\Delta HA}$ has no significant defects in attachment and entry compared to wildtype. See Sharon *et al.* (2020) supplemental data S2 for flow cytometry gating[1]. There was also a strong linear correlation (R^2=0.83) between the absolute levels of cellular GFP mRNA and HA mRNA expressed over the course of 40 hours in PR/8 $^{GFP\Delta HA}$ and PR/8 cultures (Figure 4c) showing that the modified genomic segment 4 of PR/8$^{GFP\Delta HA}$ is transcribed at a similar rate as that of PR/8 segment 4 and is an accurate predictor of PR/8 transcription levels. Overall, PR/8$^{GFP\Delta HA}$ closely replicates PR/8 infection kinetics and effect on host cells under the conditions tested. Finally, we also sought to verify the "single cycle" nature of PR/8$^{GFP\Delta HA}$ infection. As expected, PR/8$^{GFP\Delta HA}$ cultures 72hpi contain negligible concentrations of infectious viral particles compared to wildtype PR/8 (Figure 4d). The small titer of reporter virus that was observed likely indicates carryover from the initial infection.

4.3 PR/8$^{GFP\Delta HA}$ GFP reporter intensity predicts PR/8 viral yield

Though the results of Figure 4c show that GFP reporter transcription in PR/8$^{GFP\Delta HA}$ infected cells correlates well with wildtype PR/8 transcription, we wanted to assess whether GFP reporter expression was also indicative of viral yield using a positive control wherein a known restriction factor is knocked out. Results are shown in Figure 5.

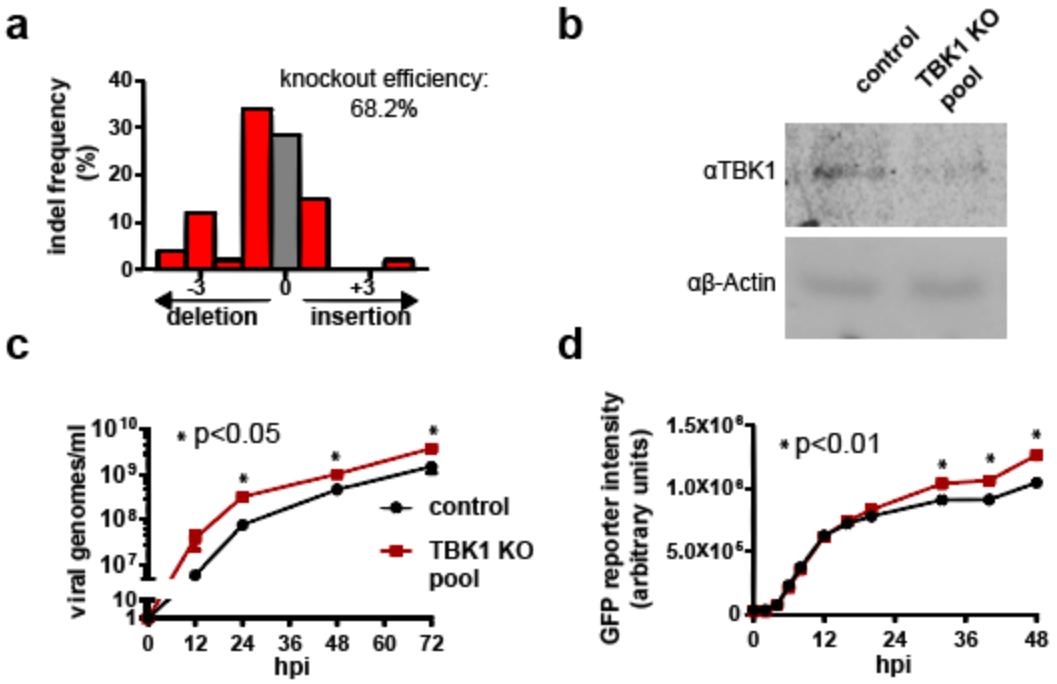


Figure 5. Validation of PR/8$^{GFP\Delta HA}$ GFP intensity as a reporter for increased viral yield using a HEK-293SF$^{\Delta TBK1}$ knockout pool

a) TIDE analysis of gDNA extracted from the HEK-293SF$^{\Delta TBK1}$ knockout pool at the sgRNA cut site. Frameshift from control sequence indicates successful indel at the gRNA target site and disruption of the TBK1 gene. Indel frequency is estimated at 68%. b) Western blot for TBK1 against extracts from the HEK-293SF$^{\Delta TBK1}$ knockout pool and control cells showing a reduction in TBK1 protein levels in the knockout pool. c) Timecourse measurement of wildtype PR/8 viral titers in terms of genomes/mL in HEK-293SF$^{\Delta TBK1}$ and control HEK-293SF CFS. Results show significantly elevated viral titers generated by the HEK-293SF$^{\Delta TBK1}$ knockout pool. d) Timecourse measurement of GFP reporter intensity in HEK-293SF$^{\Delta TBK1}$ and control HEK-293SF cells infected with PR/8$^{GFP\Delta HA}$. Results show significantly elevated GFP reporter intensity in HEK-293SF$^{\Delta TBK1}$ cells. TIDE analysis was conducted using the TIDE webtool. Datapoints of figures c) and d) represent the average of n=3 independent replicates run in parallel. Error bars represent SEM. All p-values were determined using using 2-tailed Student's t-test.

Tank binding kinase 1 (TBK1) is a serine-threonine kinase and a key component of the RIG-I dsRNA sensing pathway[102]. Disruption of TBK1 expression has been shown to increase PR/8 titer in HEK-293[102,227]. In order to examine whether GFP expression in PR/8 $^{GFP\Delta HA}$ infected cells could predict increases in wildtype viral yield, CRISPR/Cas9 was used to knock out TBK1 in HEK-293SF cells, generating a HEK-293SF$^{\Delta TBK1}$ knockout pool. Tracking of Indels by Decomposition (TIDE) analysis (Figure 5a) and Western blot (Figure5b) show successful modification at the target DNA loci and a reduction in TBK1 protein levels in the HEK-293SF$^{\Delta TBK1}$ pool, respectively. Wildtype PR/8 yield was then assessed in these cells. As shown in Figure 5c, cells from the HEK-293SF$^{\Delta TBK1}$ knockout pool showed a significant increase in viral yield compared to controls. Both unmodified cells and those from the HEK-293SF$^{\Delta TBK1}$ knockout pool were then infected with PR/8$^{GFP\Delta HA}$, and the mean GFP fluorescence intensity of infected cells was measured over 48 hours via flow cytometry. The flow cytometry gating strategy can be found in Sharon *et al.* (2020) supplemental data S2[1]. As shown in Figure 5d, the mean GFP intensity of the HEK-293SF$^{\Delta TBK1}$ knockout pool was significantly elevated relative to control cells. This demonstrates that high yield knockout populations can be identified by GFP intensity from PR/8 $^{GFP\Delta HA}$ reporter virus infection, validating its use in the screen.

4.4 Screen controls and quality metrics performed as expected

Three independent replicates of the screen depicted in Figure 1 were run, each generating a control (all infected cells) and high yield fraction (cells in the 90th percentile of GFP expression), totaling six samples. FACS gating strategy for sample collection is shown in Figure 6, while Figure 7 shows assessment of screen replicates in terms of deep sequencing base calling quality, read mapping, and the distribution of these reads within the library.

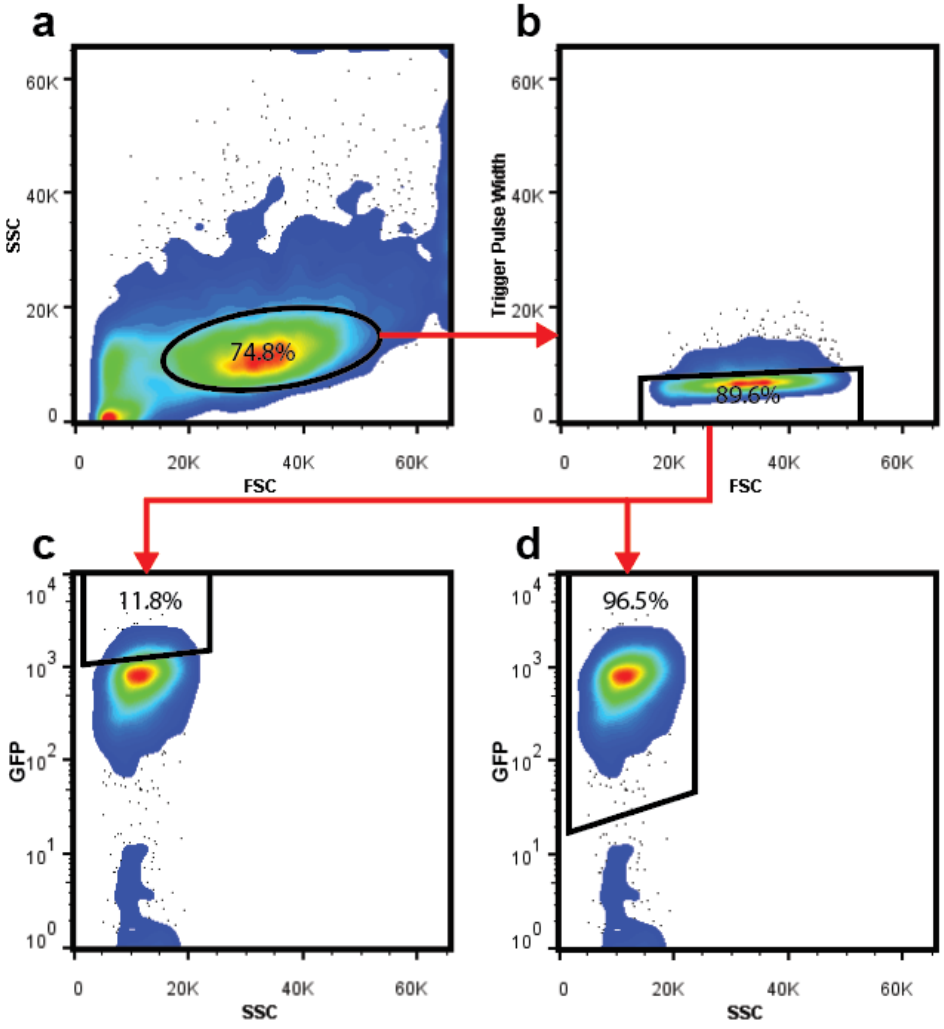


Figure 6. FACS gating strategy for sample collection

a) whole cells are isolated from debris and large clumps followed by b) doublet discrimination based on pulse width. For cell sorting during the screen, two populations of cells were collected: c) a high yield fraction consisting of the top 10% of GFP expressing cells and d) a control fraction consisting of all infected (GFP positive) cells.

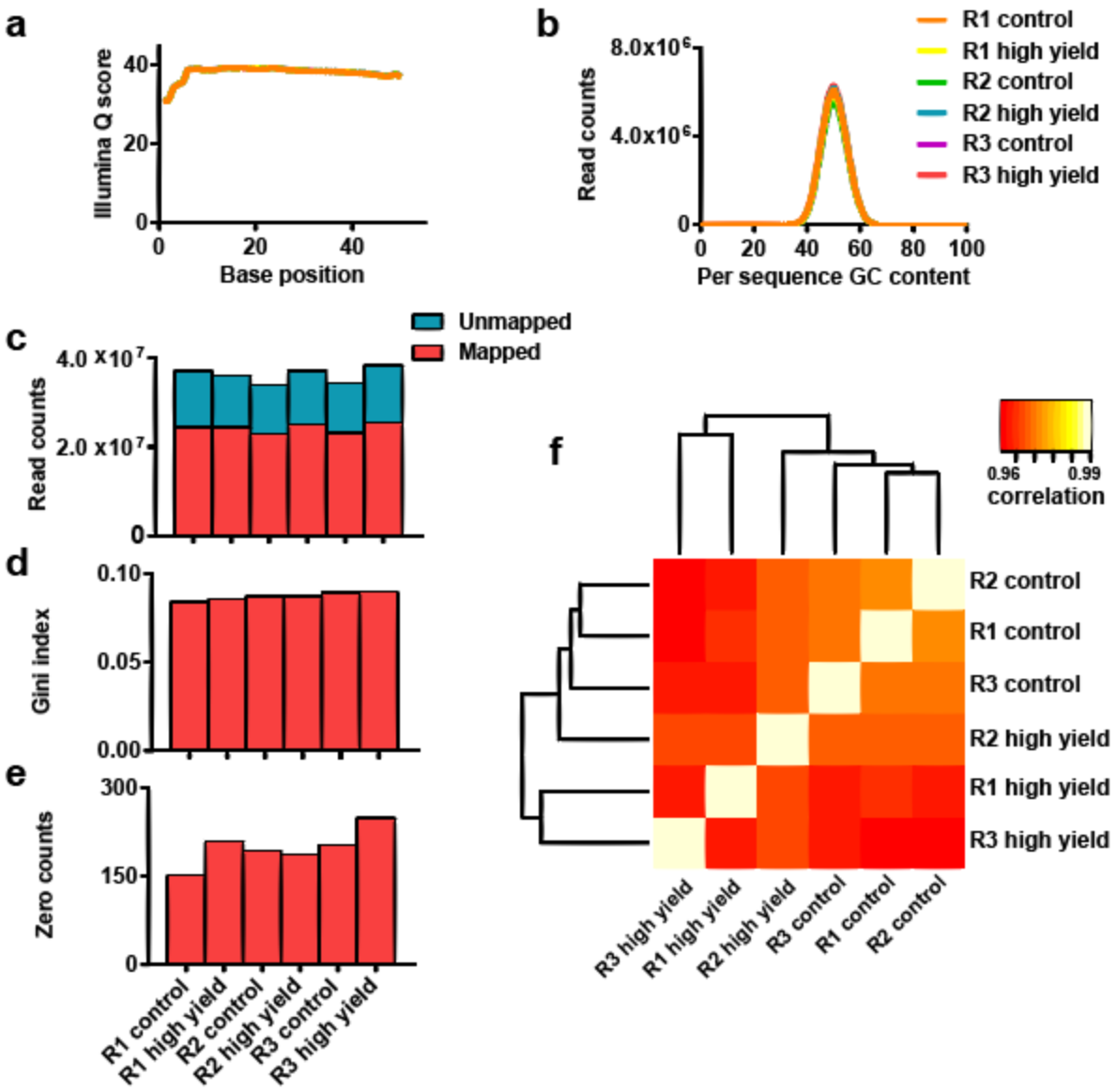


Figure 7. Deep sequencing and read mapping quality metrics

Six samples, consisting of high yield and control fraction for each of the three replicates (abbreviated R1, R2, and R3), were sequenced together as a barcoded pool. a) Average per base Illumina Q score, an aggregate measure of base-calling quality, for each sample. b) Distribution of average GC content per read for each sample. c) Illustration of total reads per sample and reads that were successfully mapped to the library. d) Gini index for each sample, a measure of read distribution across the library. E) Number of zero counts, elements in the library for which no read was mapped, for each sample. F) Heat map showing sample grouping and correlations between samples. Data was generated using the MAGeCK software suite.

For all samples, average per base Illumina Q score, an aggregate measure of base calling quality, was > 30 for the entirety of the 50 bp read (Figure 7a), indicating good base calling during sequencing. GC content per read was consistent between samples and indistinguishable from the theoretical distribution for the library (Figure 7b), indicating that there was no significant bias in the PCR used to generate the amplicon libraries or in the deep sequencing itself. For all samples, > 65 % of reads were mapped to the library (Figure 7c), further indicating satisfactory sequencing quality. The library was also well represented in terms of both average read depth (~300 per sgRNA) (Figure 7c), and a Gini index, a measure of read distribution across the library, below 0.1 for all samples. Zero count sgRNAs, elements in the library for which no read was mapped, comprise <1 % of total library elements (n=76,411) for all samples (Figure 7e). Figure 7f shows correlation and clustering between samples. As expected, the strongest correlations are seen between the three control samples and between the three high yield fractions, respectively.

The library contains an average of four distinct sgRNA targeting towards each protein coding gene. Figure 8a shows the median abundance and enrichment of sgRNA reads for the three replicates.

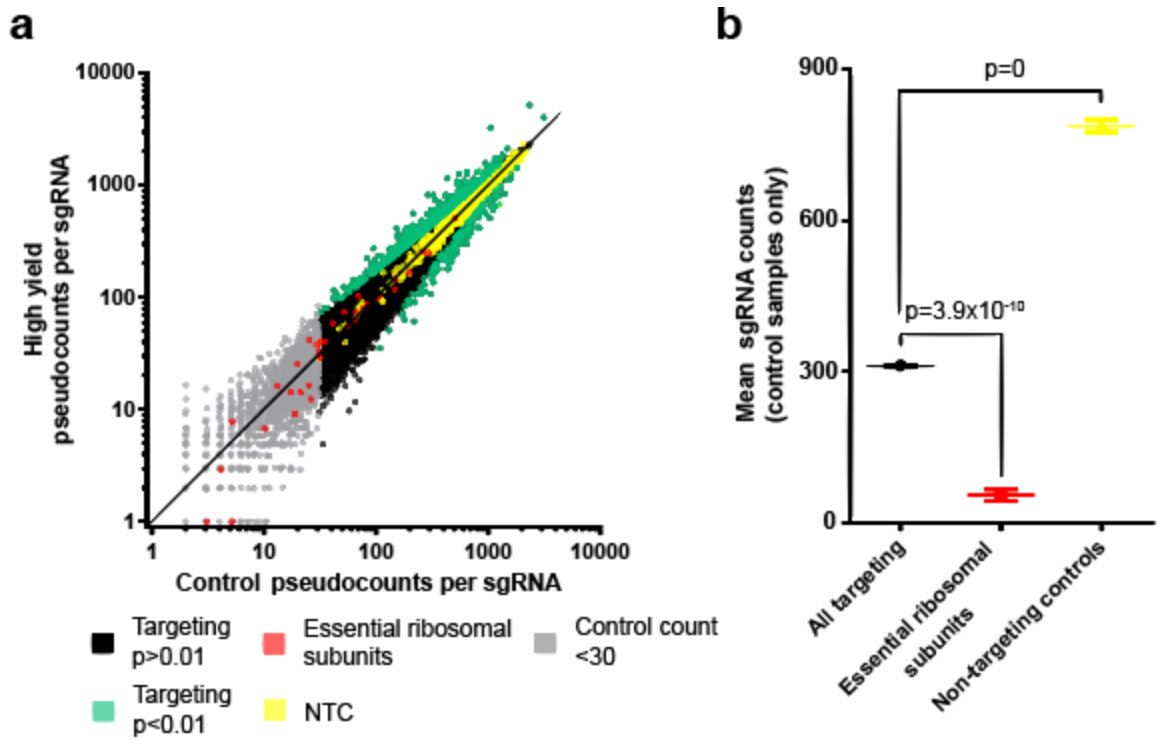


Figure 8. sgRNA abundance, enrichment, and controls

a) Log-Log plot of abundance for each individual sgRNA in high yield and control samples. Values represent the median of three replicates. A pseudocount of +1 was added to all sgRNA to allow plotting of zero count sgRNA. Elevation above or below the line x=y indicates enrichment or depletion of that sgRNA, respectively. Significantly enriched or depleted sgRNA (n=754), non-targeting control (NTC) sgRNA (n=1000), and controls in the form of sgRNA that target essential ribosomal subunits (n=34) are highlighted. sgRNA with a median read count in control samples of <30 (n=1760) are displayed but were excluded from downstream analysis due to insufficient representation. b) Mean abundance in control samples of all targeting sgRNA, sgRNA targeting essential ribosomal subunits, and NTC sgRNA. Significant enrichment/depletion of individual sgRNAs was determined using the MAGeCK software suite. Note that p-values have been adjusted for multiple testing using the Benjamini-Hochberg procedure. Significance values in plot b) were determined using 2-tailed Student's t-test. Whisker plots represent mean and SEM.

A total of n=754 sgRNA were significantly enriched or depleted with a threshold of $p<0.01$ in high yield fractions. As shown in Figure 8b, the mean count of all sgRNA targeting essential ribosomal subunits (as identified by Hart *et al.* (2014)) is significantly lower ($p=3.9x10^{-10}$) than the mean of all other targeting sgRNA in control samples [228]. The depletion of this subpopulation in which essential cell survival factors have been targeted for knockout indicates efficient CRISPR/Cas9 activity in the screen. Non-targeting controls were found to be significantly elevated compared to targeting sgRNA, which is an expected result considering we observe a 24-48 hour growth arrest in HEK-293SF following CRISPR/Cas9 knockout. We speculate this is due to cells resolving the DNA damage that mediates CRISPR/Cas9 activity, as has been reported elsewhere[229].

4.5 Identification and analysis of putative influenza restriction factors in HEK-293SF

A total of n=135 significantly enriched gene knockouts were identified by Robust Rank Aggregation (RRA) analysis using the MAGeCKFlute software suite and a significance threshold of $p<0.01$[215]. See methods section for analysis parameters. An arbitrary threshold for minimal biological significance of $\log_2$ fold-change (lfc) >0.4 was also imposed to refine the results. Results are shown in Figure 9.

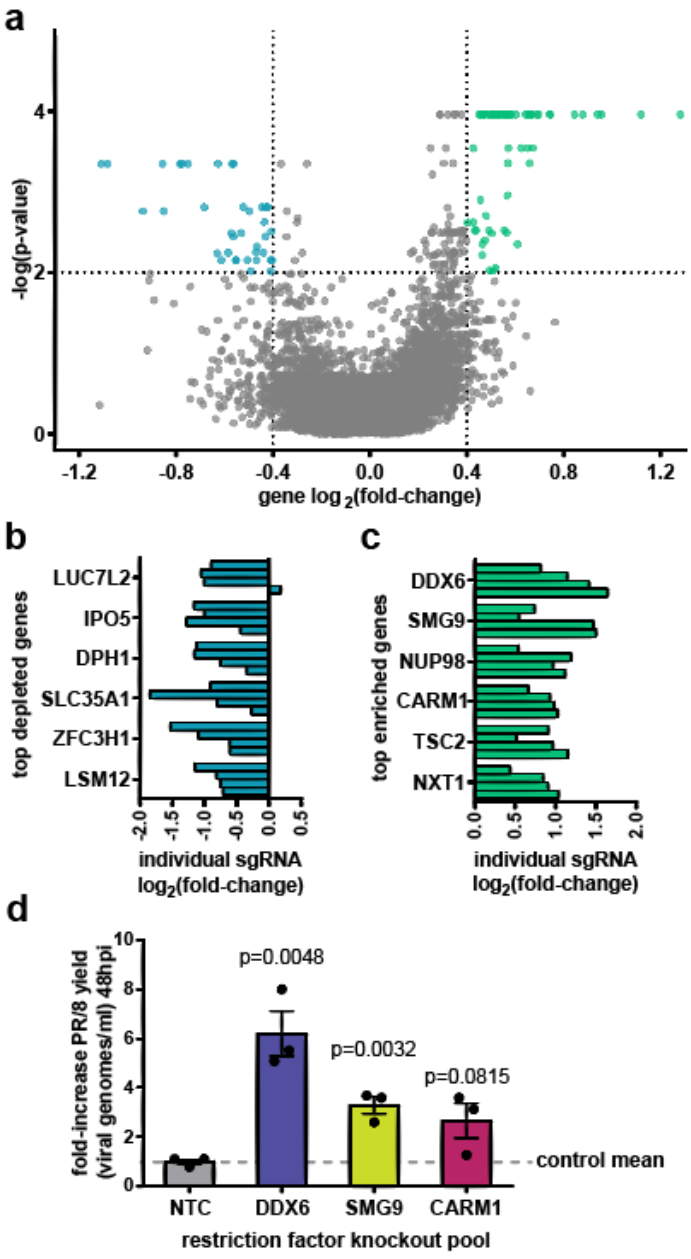


Figure 9. Summary of significantly enriched and depleted genes and hit validation

a) Volcano plot showing the mean $\log_2$ fold enrichment (lfc) and p-value for each gene knockout. Gene "hits" are defined as having $p<0.01$ and $|lfc|>0.4$. Using these criteria, n=64 putative restriction factors were identified, as well as n=37 significantly depleted genes. b) Individual sgRNA lfc for the six most depleted genes in the screen, i.e. knockouts which were detrimental to influenza replications. c) Enrichment of individual sgRNA corresponding to the top six putative restriction factors identified in the screen. d) Validation of top putative restriction factors DDX6, SMG9, and CARM1 using wildtype PR/8 virus at an MOI of 0.1. Results indicate a 6-3 fold increase in viral titer at 48hpi following restriction factor knockout. Significance values in plot d) were determined using 2-tailed Student's t-test, with n=3 independent replicates run in parallel. Error bars represent SEM. Values in panels a), b), and c) were obtained by Robust Rank Aggregation (RRA) analysis with the MAGeCKFlute software suite and plotted in Prism. Note that p-values have been adjusted for multiple testing using the Benjamini-Hochberg procedure. See methods section for analysis parameters.

Overall, n=64 gene knockouts were identified with these criteria, hereafter referred to as putative restriction factors (Figure 9a). Although not a goal of this study, we also identified a subset of genes that were significantly depleted that may be of interest, particularly those looking to identify host targets of antiviral drugs (Figure 9b). However, we caution that the screen lacks adequate controls to distinguish whether these depleted factors exert a specific effect on influenza replication, or merely that their knockout is detrimental to the cell's overall biosynthetic capacity. Top putative restriction factors in the screen are shown in Figure 9c. In addition, because this screen used an indirect readout (GFP fluorescence) to assess viral titer. It was important to validate a portion of the putative restriction factors identified by direct measurement wildtype PR/8 titers to verify the integrity of the dataset. For each of the genes DDX6, SMG9, and CARM1, a knockout pool was generated using a randomly selected sgRNA sequence from the Brunello Library. Editing efficiency was estimated by TIDE analysis as >80% for all pools (see Sharon *et al.* (2020) supplemental data S5[1]). Cells were then infected with PR/8, and fold-change in influenza yield measured compared to cells transduced with a non-targeting control sgRNA (Figure 9d). Results show an increase of viral titer of 3-6 fold over non-targeting controls, verifying the results of the screen. See Sharon *et al.* (2020) supplemental data S4 for full analysis and raw sequencing data[1]

In order to identify common biological pathways, processes, and protein-protein interactions (PPI) within our list of putative restriction factors, a Gene Ontology (GO) analysis was conducted using Metascape (Figure 10)[220].

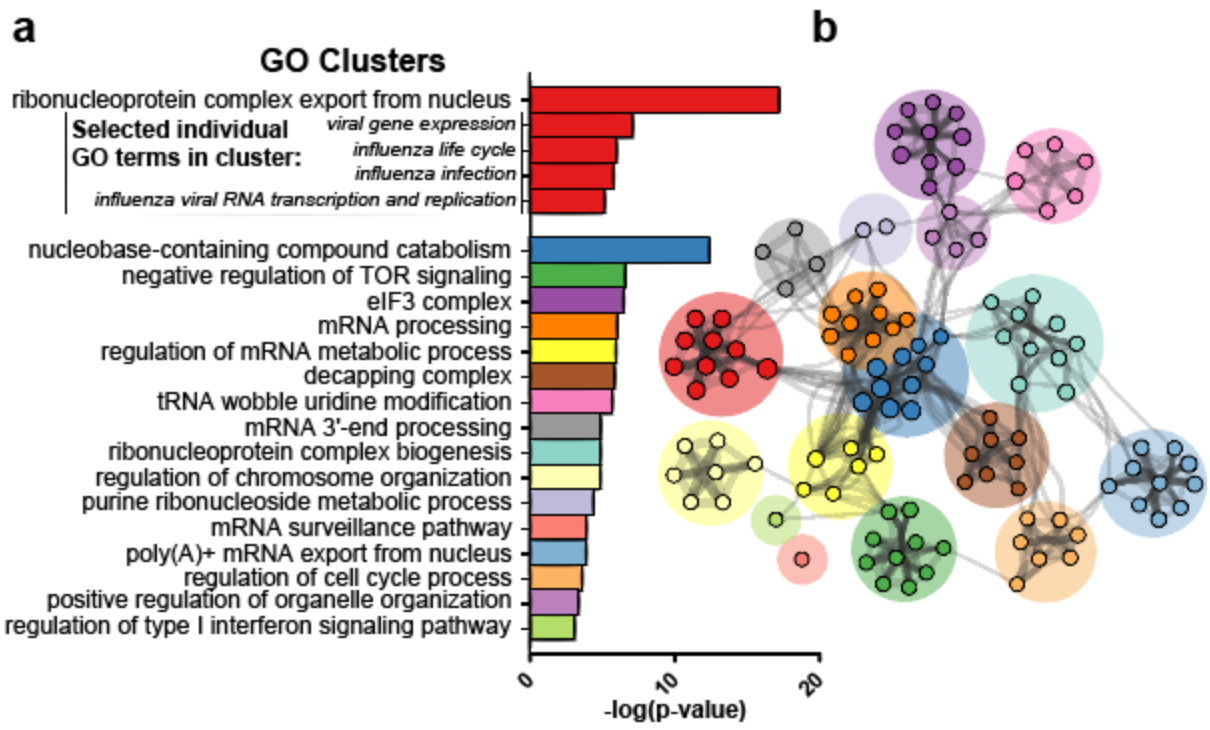


Figure 10. Gene ontology (GO) analysis of putative restriction factors

a) Significantly enriched GO clusters obtained from the 64 putative restriction factors in the screen. In total, n=274 individual GO terms were found to be enriched with a threshold of $p<0.01$. Associated GO terms were clustered, with the name of that cluster represented by the individual GO term with the lowest p-value. Within the first cluster, selected individual GO terms related to influenza have also been displayed. b) Graphical representation of the association between GO clusters. Nodes represent subclusters, with node size corresponding the to the number of genes in that node. Edges represent association between subclusters in terms of biological processes, pathways, or PPI. GO analysis was conducted using the Metascape webtool and plotted in Cytoscape.

A total of n=274 individual GO terms were significantly enriched with a significance threshold of $p<0.01$. As expected, we see highly significant enrichment of terms related to influenza (Figure 10a) including Transport of Virus ($p=2.0x10^{-5}$), Influenza Life Cycle ($p=5.8x10^{-4}$), and Influenza Infection ($p=7.7x10^{-4}$). While enrichment of influenza-related terms serves as a supporting additional validation of the screen results, the majority of the putative restriction factors identified are not associated with these ontology terms, suggesting analysis of other highly enriched GO terms may provide insight into novel antiviral pathways in the cell. Highly significant GO clusters, groups of related GO terms, are shown in Figure 10a. Figure 10b provides a graphical representation of the interactions between subclusters. See Sharon *et al.* (2020) supplemental data S6 for the full list of significantly enriched GO terms[1].

In addition to the GO analysis, we also conducted an analysis for the enrichment of stable protein complexes using COMPLEAT[222]. If an individual restriction factor exerts its antiviral effect as part of a multiprotein complex, it follows that the knockout of any protein in that complex which is required for complex functionality should exert a similar antiviral effect. We would thus expect to see that some of the top ranked genes in the screen are part of the same protein heterocomplexes. This analysis is useful both as an indirect validation of the screen, and to guide subsequent efforts to create high yield knockout cell lines based on the screen results, as the knockout of multiple restriction factors that exert their effect through a common protein complex would be redundant. Results are summarized in Figure 11.

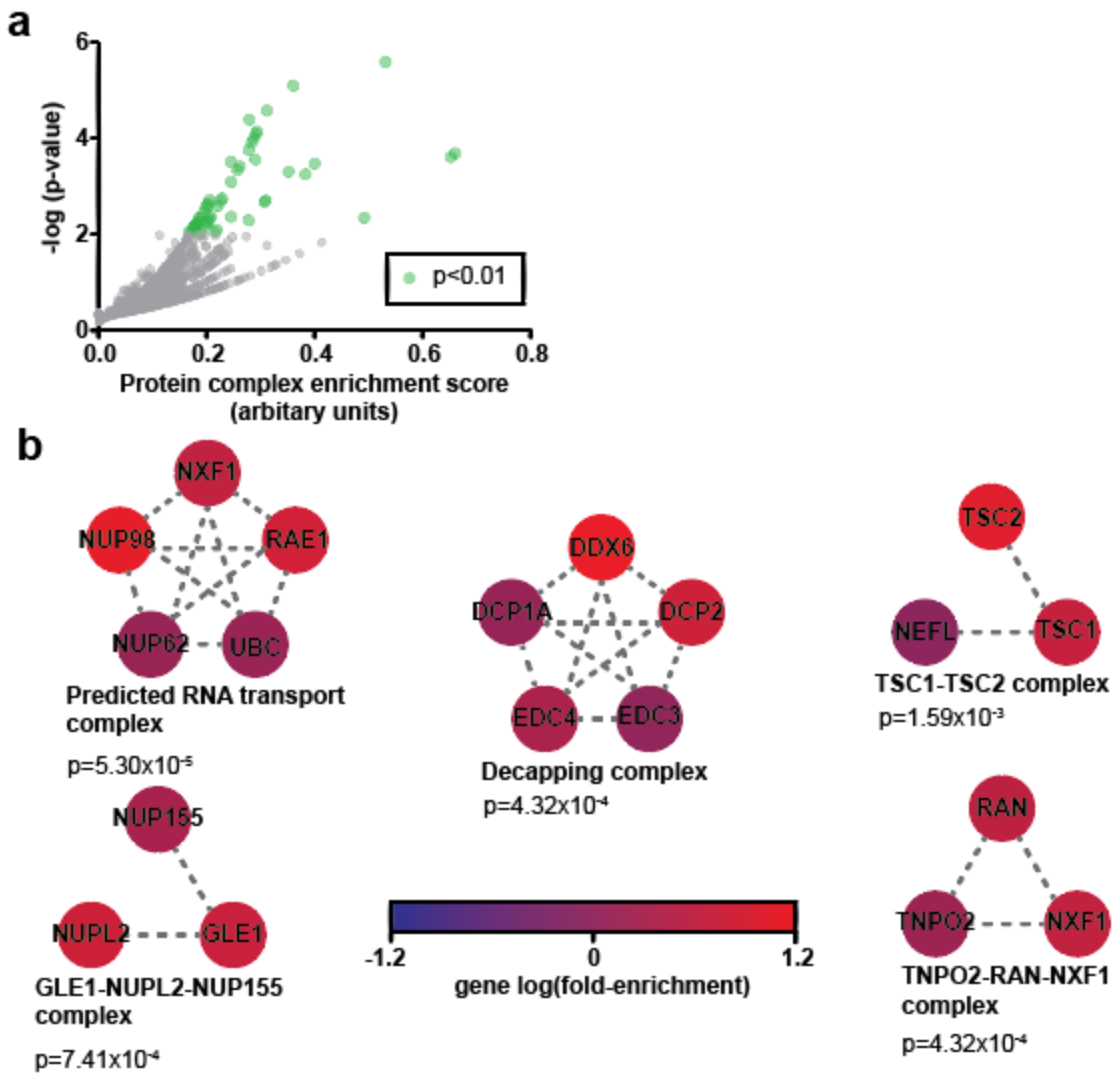


Figure 11. Protein complex enrichment analysis

a) Significance and enrichment score (based on the COMPLEAT aggregate scoring system that factors in complex size, gene lfc, and other factors) of individual protein complexes. n=84 protein complexes were found to be significantly enriched at a threshold of p<0.01. b) The top five enriched protein complexes identified, showing the individual genes corresponding to a given protein and lfc of that gene in the screen, as well as the interactions between proteins in the complex. Both p-values and enrichment scores were generated using the COMPLEAT web tool.

As shown in Figure 11a, n=84 protein complexes (either described in literature or predicted through proteomics studies) were found to be significantly enriched with a threshold of $p<0.01$. The highest scoring complexes as determined by the COMPLEAT algorithm are depicted as well in Figure 11b. See Sharon *et al.* (2020) supplemental data S7 for the full list of significantly enriched complexes[1].

5.0 ATTEMPTS TO DEVELOP A SCALABLE, HIGH YIELD MULTIPLE KNOCKOUT CELL LINE FOR THE CELL-BASED PRODUCTION OF INFLUENZA VACCINES

5.1 Single gene knockout pools guided by screen results show cell-density dependent effects on viral yield

While the screening data is informative in terms of understanding host-virus interactions, the primary purpose of this screen is to direct the creation of high-yield knockout cell lines. Based on screen rank and a review of available literature, n=12 putative restriction factors identified in the screen were selected for follow-up analysis and validation. To this end, separate CRISPR/Cas9 knockout pools were generated for each of these genes to analyze in impact of individual knockouts on cell growth characteristics and influenza yield.

Initial experiments examined cell doubling time and viral yields in single knockouts under infection conditions that were optimized for unmodified HEK-293SF. Results are shown in Figure 11 below.

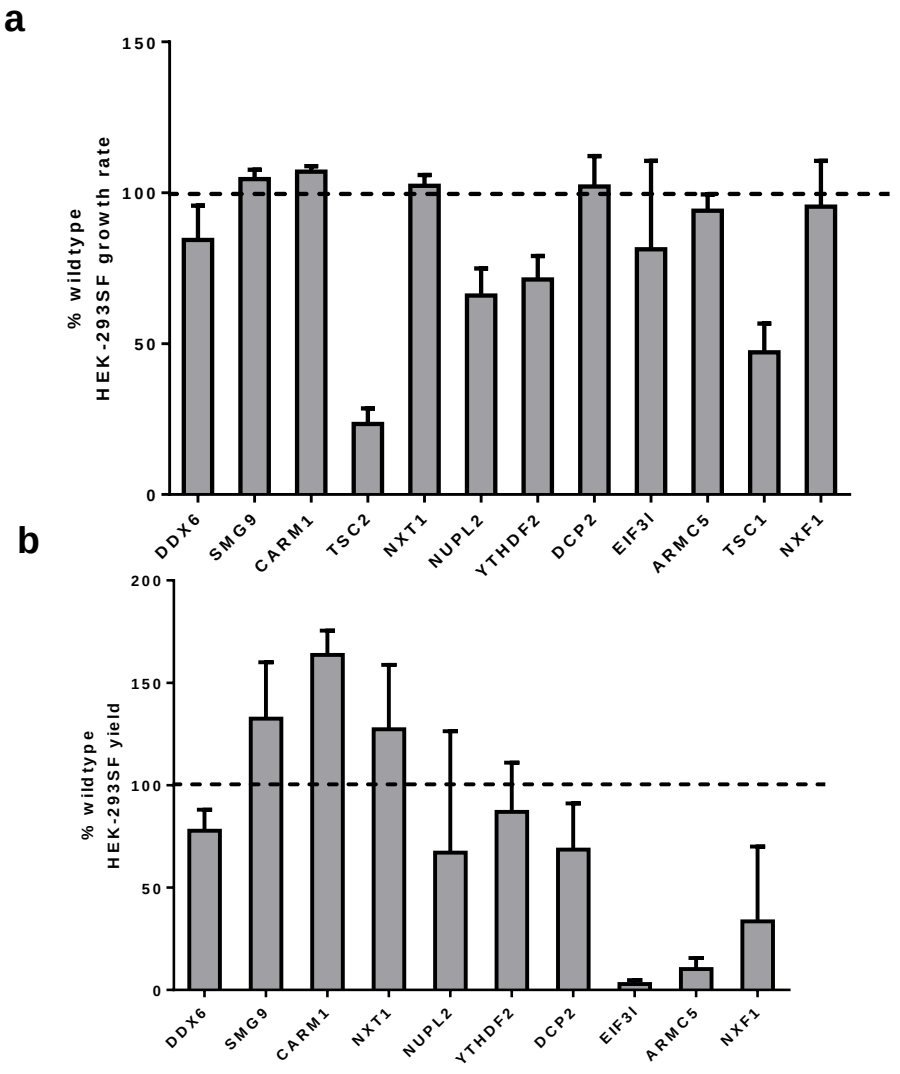


Figure 12. Cell growth rate and PR/8 influenza yield of knockout pools

a) Growth rate of knockout pools compared to HEK-293SF cells transduced with a non-targeting control sgRNA. Growth rate is defined as the average doubling time of the cells when maintained between 0.5-2.0 x 10^6 cells/ml**. b)** PR/8 influenza yield of knockout pools compared to HEK-293SF cells transduced with a non-targeting control sgRNA. Yield is measured as VG/ml in cell-free supernatant at 48 hpi. Cells are infected a MOI=0.001 at a density at 1 x 10^6 cells/ml. Error bars represent SEM of n=3 replicates.

As shown in Figure 12a, TSC1 and TSC2 knockout significantly decreased the growth rate of the cells, and so these pools were not included in subsequent experiments. While unfortunate that these knockouts could not be evaluated further, it was encouraging to see that the knockout of both genes generated similar effects on the cells as the proteins are active as a heterodimer [230].

Increases in viral yield under these conditions were modest, with a 60% increase in yield at best (Figure 12b). As discussed earlier though, many of the restriction factors identified are activated during metabolic stress and play a role in mediating the consequent downregulation of anabolic metabolism. Metabolic stress from buildup of toxic metabolites and nutrient starvation is thought to be a key mechanism behind the cell density effect[231-233]. The effect is conserved across many different cell lines and viruses and is often a major limiting factor in process design for virus production in mammalian cells, necessitating complicated and expensive fed batch or perfusion culture systems to overcome [204,234]. This raised the interesting question as to whether any of the mutations we had generated would allow cells to maintain their specific productivity at higher cell densities. This is particularly pertinent for the HEK-293SF cell line, which exhibits steeply declining specific and volumetric yield of influenza, adenovirus, AAV, and lentivirus when initial infection densities exceed 1 x 10^6 cells/ml[12].

To investigate this, cells were infected at two different densities: 1x10^6 cells/ml and 3x10^6 cells/ml, and yields compared to controls. In addition, a non-exhaustive panel of double knockout pools was generated to investigate whether any restriction factor knockouts exhibited synergy. Selection of double knockouts was based on literature review and identification of restriction factors that participate in parallel (and potentially compensatory) biological processes. Results are shown in Figure 13.

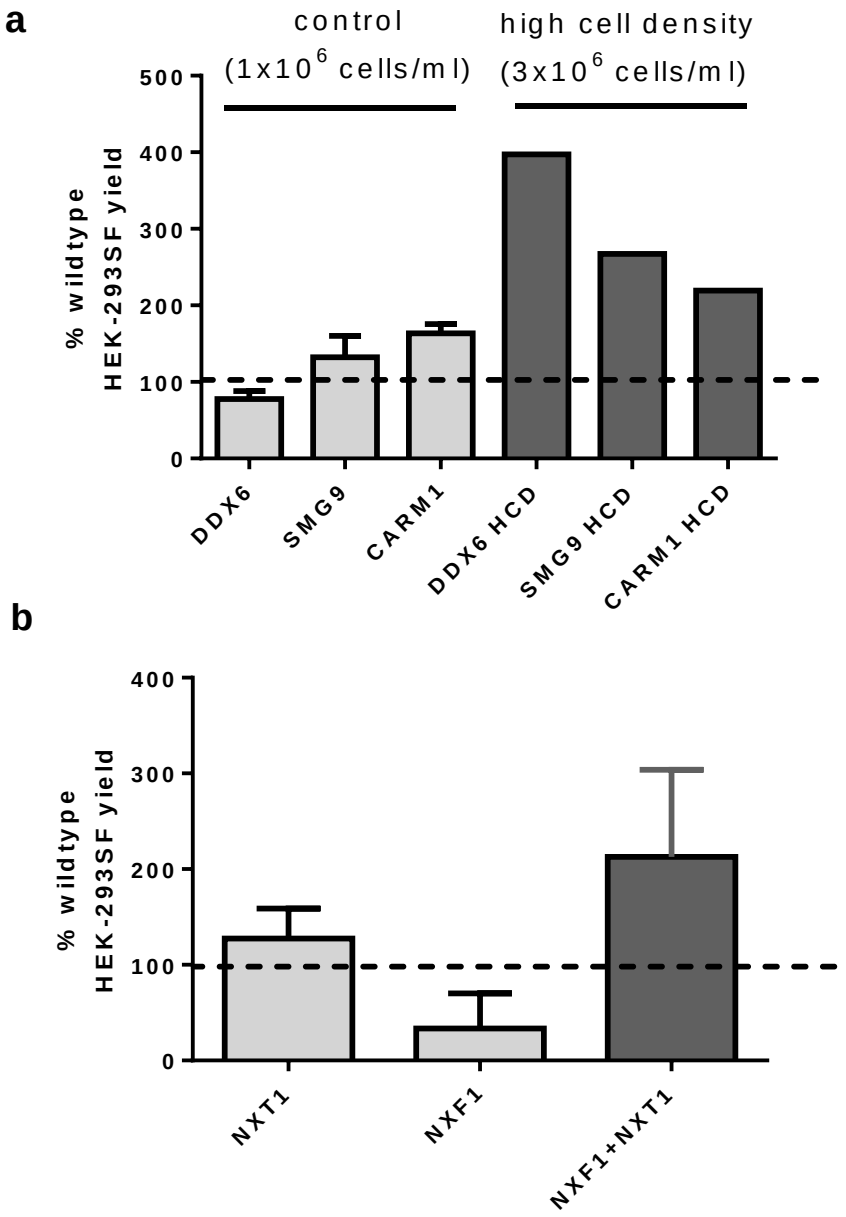


Figure 13. PR/8 influenza yield of selected knockout pools under high cell density culture and double mutants

a) PR/8 influenza yield of knockout pools compared to HEK-293SF cells transduced with a non-targeting control sgRNA. Yield is measured as viral genomes/ml in cell-free supernatant at 48 hpi. Cells are infected a MOI=0.001 at the indicated cell density. Note that yield is relative to a control infected at that same cell density. **b)** PR/8 influenza yield of knockout pools compared to HEK-293SF cells transduced with a non-targeting control sgRNA. Yield is measured as VG/ml in cell-free supernatant at 48 hpi. Where present, error bars represent SEM of n=3 replicates.

As shown in Figure 13a. the DDX6, SMG9, and CARM1 knockout pools exhibited greatly increased (up to 400%) yields over controls at high cell densities. In addition, a synergistic effect between NXF1 and NXT1 was identified that also substantially increased culture productivity (Figure 13b).

5.2 Generation, clonal expansion, and genotyping of multiple knockout clones

Rather than continuing to investigate different culture conditions for single knockout pools, I proceeded directly to clonal isolation of cells featuring simultaneous knockouts of DDX6, SMG9, CARM1, NXT1 and NXF1. Any further data collected from single knockout pools would have been of questionable value for two reasons. Firstly, it was impossible to accurately predict potential interactions between individual knockouts. Secondly, any optimization of cell density or other culture conditions at the 125ml shake-flask scale may not translate to eventual scaleup to a 1L bioreactor.

To this end a pool of HEK-293SF were simultaneously transfected with five CRISPR/Cas9 plasmid constructs to generate the desired mutations. Calculations based on the efficiency of the respective sgRNA indicate that in order to have a >90% chance of generating at least 3 clones featuring complete knockout of every allele, 200 clones must be isolated. HEK-293SF is generally difficult to isolate clonally, and so this task was completed by collaborators at the NRC using the CellCelector™ system. Of the 384 clones that were originally isolated, 160 survived and were designated multiple knockout (MK)1-160. Clone genotyping was accomplished by using multiplex PCR and deep sequencing strategy illustrated in Figure 14.

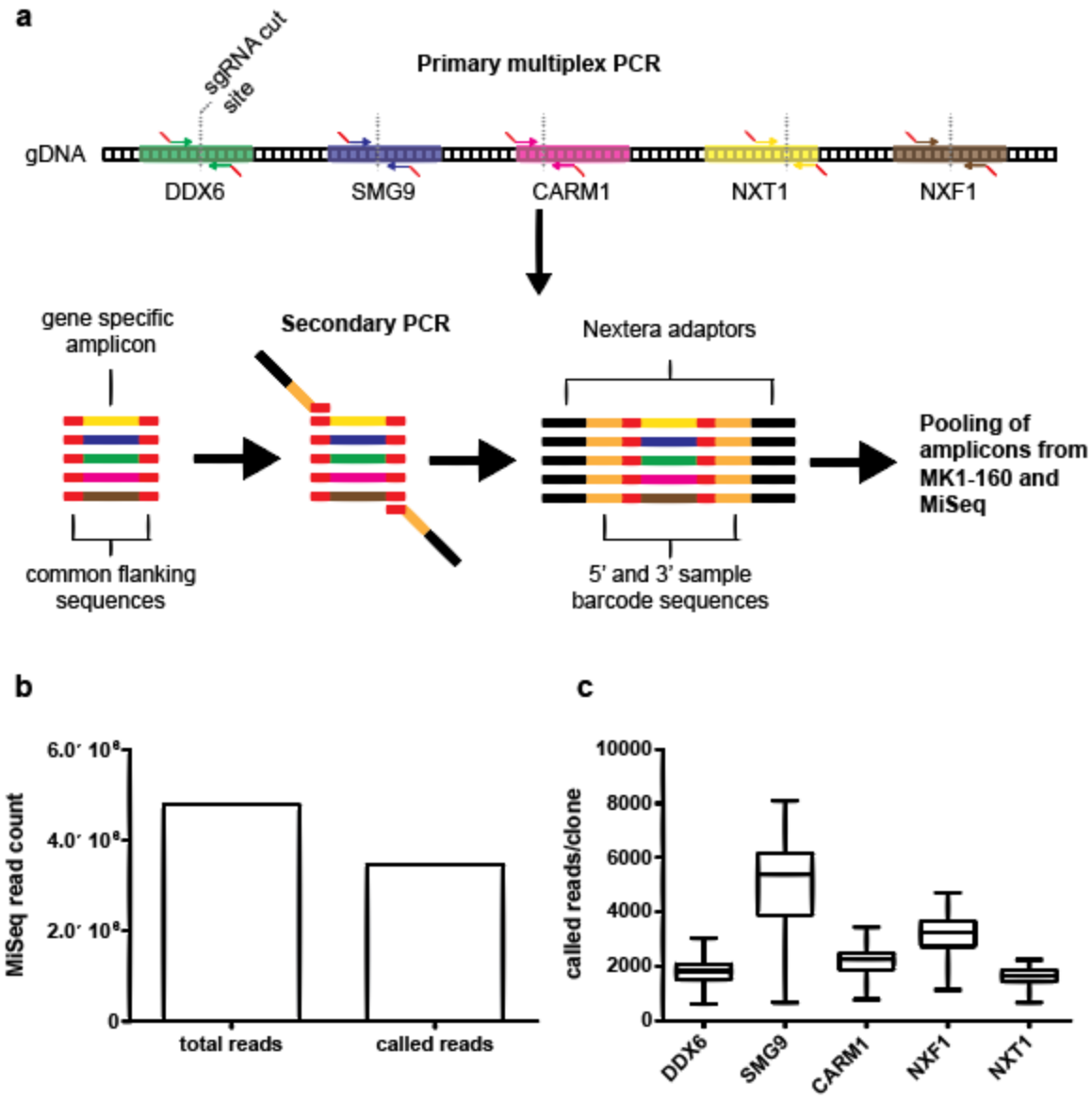


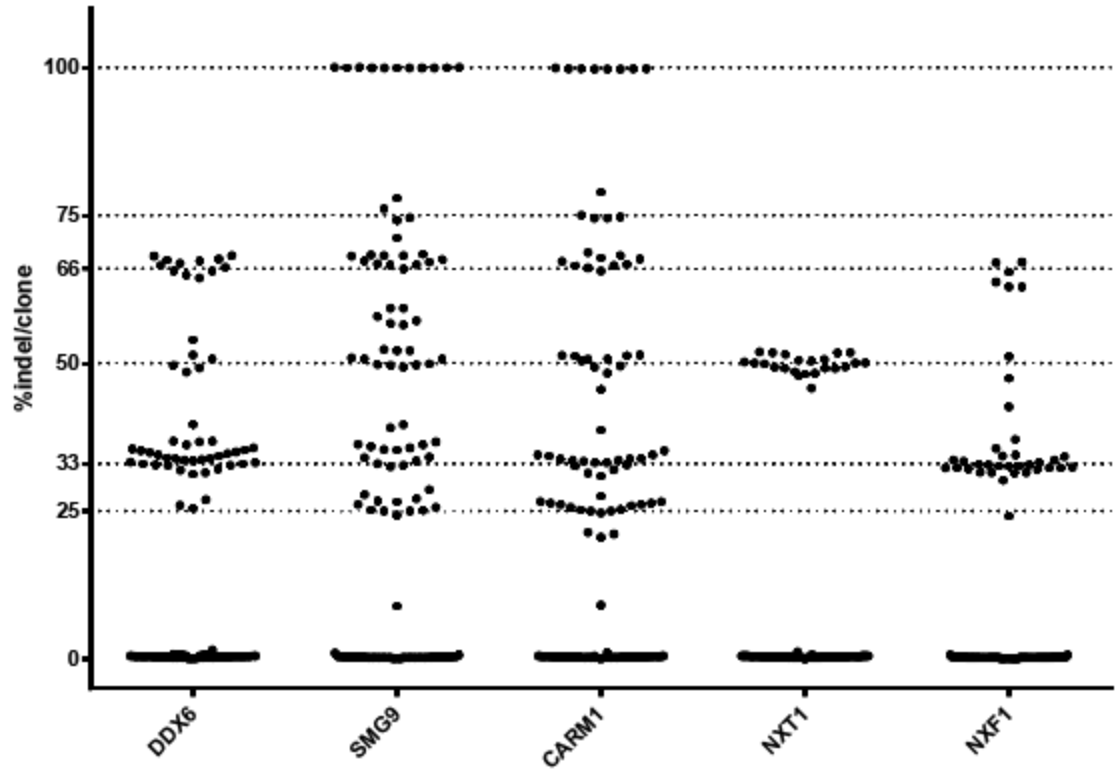
Figure 14. MK clone genotyping by MiSeq amplicon sequencing

a) Multiplex two step PCR strategy to generate amplicons for MiSeq. gDNA from each clone was extracted and the regions flanking the five sgRNA cut sites amplified, barcoded, and appended with Nextera adaptors. Amplicon libraries from the 160 respective MK clones were then pooled prior to MiSeq. b) Comparison of total and uncalled reads for the MiSeq run. c) Called reads per gene for each MK clone. Box plot and whisker plot represents mean, interquartile range, and max/minimum.

The multiplex two step PCR strategy to generate an amplicon pool for MiSeq depicted in Figure 14a was generally successful. Amplicons were ~350bp in length and sequenced as 300bp paired-end reads that were later merged to generate a single contiguous read. Of the $n=4.8x10^6$ reads generated, 72.4% were successfully called and mapped to a particular clone and cut site (Figure 14b). On average, each clone had ~2750 reads per cut site, with no clone having less than 624 reads for any one of the five cut sites (Figure 14c). Considering that the ploidy of the HEK-293SF cell line ranges from diploid to tetraploid, each allele has a minimum mean coverage of >500x and an absolute minimum coverage of 156x, which should be more than sufficient to accurately survey indel frequencies[223].

The MiSeq data analysis pipeline to characterize indels frequencies in clones MK1-160 was modified from the CRIS.py and CRISPresso2 software packages (see methods section for details)[219,235]. The primers to generate the amplicons for sequencing were designed such that the primer binding sites were at least 100bp from the sgRNA cut sites, enabling the detection even of large indels. Results of the analysis are shown in Figure 15 below.

a



b

	DDX6	SMG9	CARM1	NXT1	NXF1
MK30	32.1%	99.9%	31.3%	0.5%	0.1%
MK43	64.4%	100%	49.3%	0.2%	0.2%
MK56	0.4%	76.1%	99.7%	0.4%	33.2%
MK62	54.0%	99.9%	0.2%	0.4%	32.8%

Figure 15. Indel frequencies by gene for MK clones

a) Indel frequencies by gene for all 160 MK clones. An indel frequency of 100% indicates that all MiSeq reads at a particular locus contained an indel, while 0% indicates wildtype. b) Indel frequencies of clones containing the highest summed percentage of indels. These clones were expanded for further characterization

As shown in Figure 15a, the overall indel efficiency in the isolated MK clones was unexpectedly low, ranging from 6.5-25%. Indeed, 46 of the 160 MK clones screened had no detectable indels at all, and only genes SMG9 and CARM1 were completely knocked out in any clone. This is somewhat surprising as the sgRNA sequences used were validated as part of the Brunello library, and transfection with any single sgRNA construct had previously been shown to produce indel efficiencies in excess of 80% (See Sharon *et al.* (2020) supplemental data S5). Despite disappointingly low editing efficiencies, four clones (Figure 15b) were expanded for further characterization. The four expanded clones represent the "most modified" clones based on the summed percentage of indel frequencies for the five gene knockouts.

One interesting point that can be gleaned from Figure 15a is the highly aneuploid nature of the HEK-293 cells line. While the genotyping protocol used doesn't give absolute copy number, only a ratio of modified alleles to unmodified, it can be inferred from the ratios achieved that ploidy varies between diploid to tetraploid. Moreover, it appears that not every chromosome shares the same copy number within the cell.

5.3 Characterization and scaleup of multiple knockout clones

Selected clones were characterized in terms of growth rate and influenza yields in terms of VG/ml at different densities. Results are shown in Figure 16.

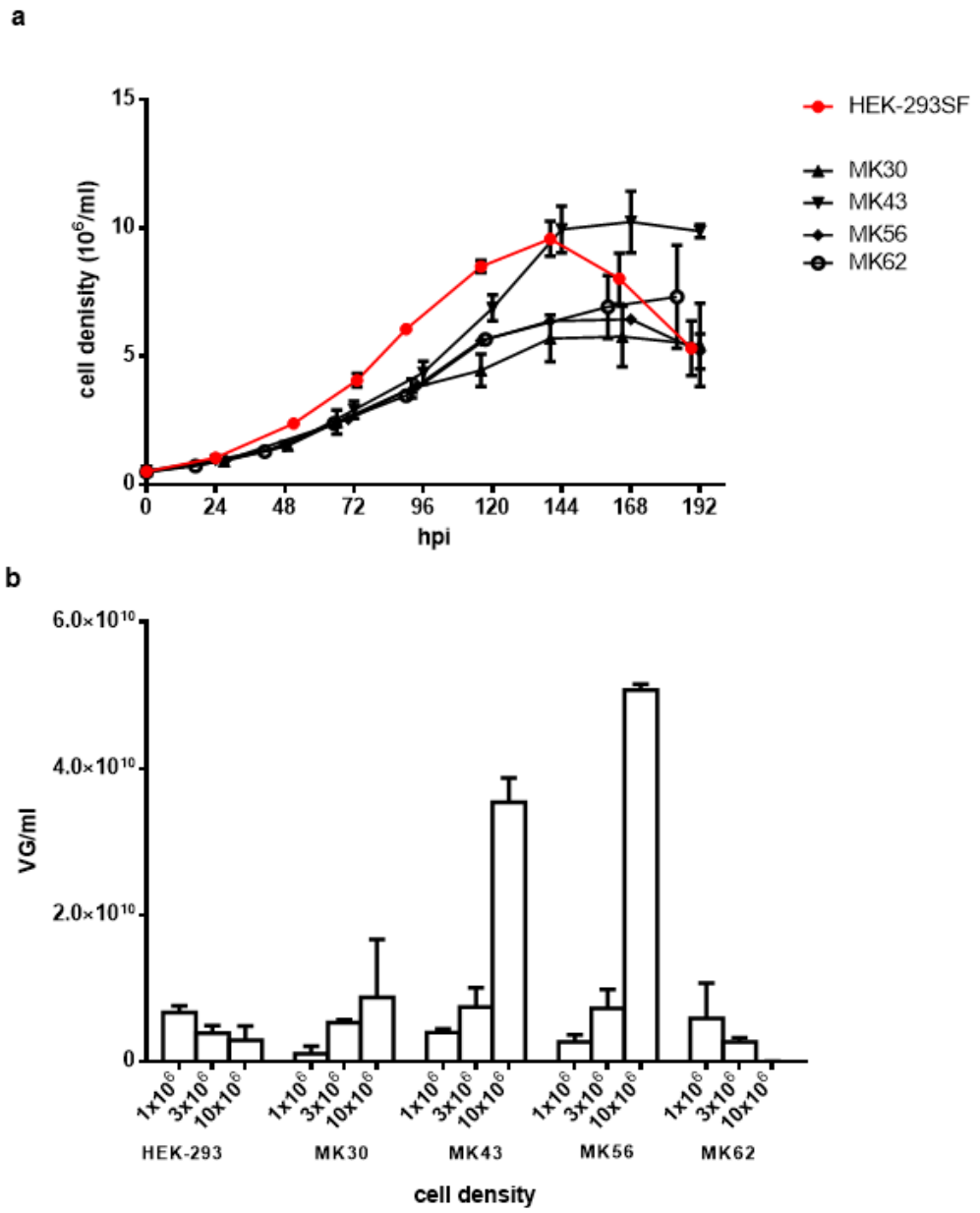


Figure 16. Grow curves and viral yield of selected clones

a) Growth curve of MK clones compared unmodified HEK-293SF cells. Cells were seeded at 1x10^{5}cells/ml and cultured without medium replacement b) PR/8 influenza yield of selected MK clones compared to unmodified HEK-293SF cells at varying cell densities. Yield is measured as VG/ml in cell-free supernatant at 48 hpi. Error bars represent SEM of n=3 replicates.

As shown in Figure 16a, with exception of MK43 all clones achieved a final cell density ~40% lower than that of the parental HEK-293SF line, as well as a markedly slower growth rate. While not ideal, this slower growth rate in the knockout clones was offset by a remarkable insensitivity to the cell density effect when measuring yield in terms of VG/ml (Figure 16b). While the parental HEK-293SF line showed maximal yields at $1x10^6$ cells/ml, the majority of clones achieved maximal yields at the maximum cell density tested ($1x10^7$ cells/ml). Moreover, the maximal yields of clones MK43 and MK56 are nearly 10-fold higher than the maximum yields of the parental cell line. In an effort to confirm these results, the same samples were assayed for total influenza particle yield via a hemagglutination assay. This assay measures total viral particles by testing the ability of the HA protein on virions to agglutinate red blood cells. Critically, this assay relies directly on the content of active HA, the primary measurement of vaccine potency[236]. Results are shown in Figure 17.

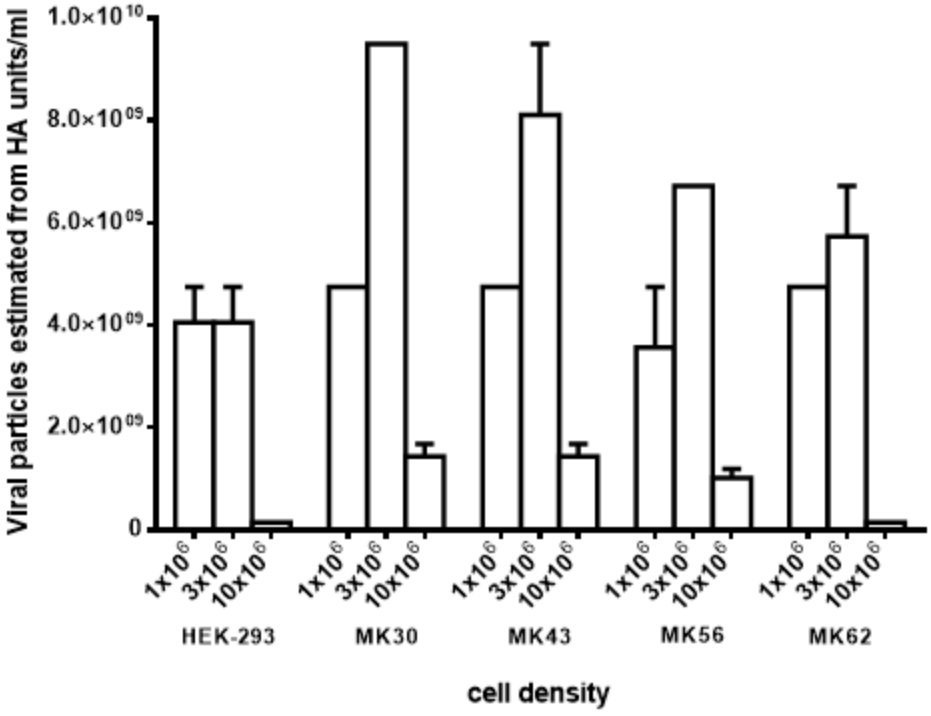


Figure 17. HA yield of selected MK clones

PR/8 influenza yield of selected MK clones compared to unmodified HEK-293SF cells at varying cell densities. Yield is measured as units of active HA and used to estimate viral particles/ml in cell-free supernatant at 48 hpi. Error bars represent SEM of n=2 replicates.

As shown in Figure 17, the results of the hemagglutination assay (which measures active HA/ml) vary considerably from those of the ddPCR based assays (which measure VG/ml), a curious result as these two measures generally correlate quite well[237]. While VG/ml seems to roughly approximate HA content in the parental cell line, in the MK clones this is not the case. Rather than the roughly linear relationship between cell density and yield observed when measuring VG/ml in MK clones, measurement by hemagglutination assay indicates that maximal yields are attained at $3x10^6$ cells/ml. Moreover, rather than a 10-fold increase in maximum yield between MK clones and the parental cell line, the difference is instead roughly 2-fold.

The HEK-293SF cell line is extremely amenable to scaling in bioreactor culture, a chief reason it is used industrially. In an effort to assess how the knockouts induced impacted the cell line's scalability and whether increases in yield were conserved, the best performing clone (MK30) was tested at the 1L bioreactor scale in batch culture. A bioreactor with unmodified HEK-293SF was run in parallel. Generally during bioreactor production of influenza, cells are seeded into a bioreactor and allowed to come up to a predetermined density before infection. During this time cell density and viability were monitored. Unfortunately, the MK30 clone failed to thrive in bioreactor culture, as shown in Figure 18.

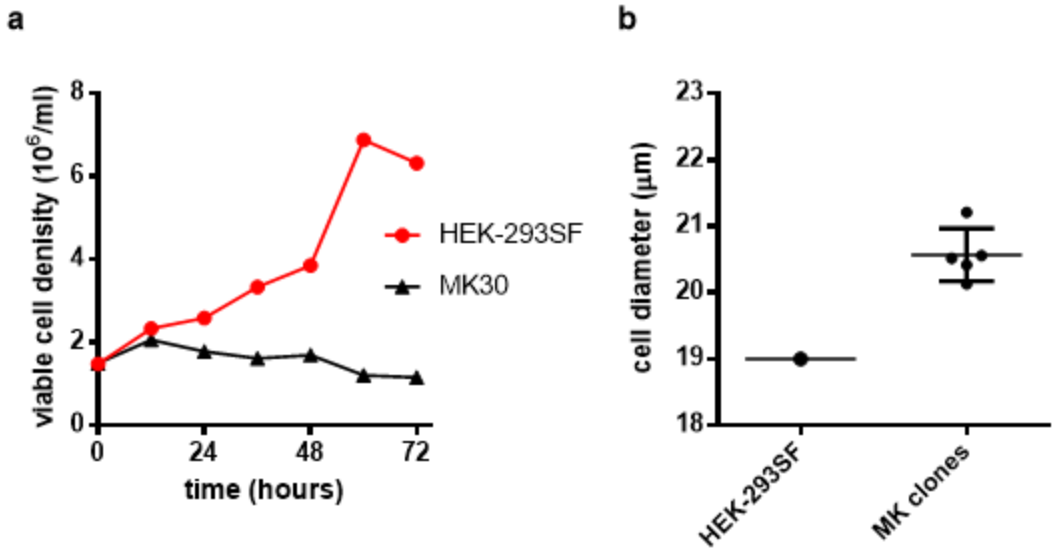


Figure 18. Growth of MK30 in bioreactor culture and average cell diameter of selected MK clones

a) Growth of PR/8 influenza yield of selected MK clones compared to unmodified HEK-293SF cells at varying cell densities. Yield is measured as units of active HA and used to estimate viral particles/ml in cell-free supernatant at 48 hpi, . Error bars represent SEM of n=2 averaged replicates per clone.

Viable cell density and declined for the MK30 after the 12 hour timepoint, whereas the parallel bioreactor containing unmodified HEK-293SF grew as expected (Figure 18a). This precluded any measurement of influenza yields at the 1L bioreactor scale for the MK30 clone. Parameters such as pH and dissolved oxygen were identical between the two bioreactors, confirmed by independent measurement with separately calibrated probes (data not shown). Samples taken from the bioreactor and inoculated into new HEK-293SF cultures also did not cause any loss of viability in the new culture, arguing against microbial contamination. One of the main differences between shake flask culture and bioreactor culture is the level of sheer stress imposed on cells, caused by agitating the culture with an impeller. Measurements show that the average cell diameter of the MK30 clone, as well as other three clones that were expanded, was significantly higher than the HEK-293SF parent line (Figure 18b). Given that larger cells are more susceptible to sheer stress, this may explain the inability of the MK30 clone to survive and grow in bioreactor culture.

6.0 INTRACELLULAR MEASUREMENT OF ASSEMBLED AAV VECTOR CAPSIDS AS A DIRECT REPORTER IN POOLED GENOME-WIDE CRISPR/CAS9 SCREENING

6.1 Application of CRISPR/Cas9 genome-wide screening to intensify transfection-based manufacturing of AAV vectors

Attempts to directly measure cell-specific yield of viral biologics in pooled screens is inherently complicated by the fact that for enveloped viruses such as influenza, virion maturation and egress are simultaneous processes. Since it's impossible to distinguish between viruses originating from different mutant subpopulations once they leave the cell, indirect reporters (such as the GFP reporter used in section 4) are necessary. However, many nonenveloped viruses assemble and aggregate within the cell and egress via lysis. With appropriate timing just before cell lysis, it should be possible to directly measure the yield of mature virions within the cell and avoid the difficulties inherent in using an indirect reporter. This in turn would result in a reduction of false-positives in the genome-wide screen. One non-enveloped virus with significant utility as a viral biologic is AAV.

In addition to viral vaccines, HEK-293SF can also be used to produce AAV vectors. The bulk of AAV vectors used for research, clinical trials, and approved therapies are currently produced by multi-plasmid transfection of mammalian cells [18,19]. Low yields from transfection-based manufacturing platforms currently make it difficult to supply these activities with adequate material. AAV assembles intracellularly and is released from the cell only upon lysis. Additionally, conformation specific antibodies exist that selectively bind fully assembled viral capsids, allowing viral yield to be approximated via staining coupled with flow cytometry. AAV

production by transfection-based manufacturing thus seemed like an ideal platform to explore for process intensification via CRISPR/Cas9 screening.

6.2 Establishing a model for transfection-based manufacturing of AAV

To establish a model for subsequent experiments, a triple plasmid transfection to produce AAV2-GFP was carried out on HEK-293SF cells in suspension, based on an optimized process developed by Chahal et al. [238]. Results are shown in Figure 19.

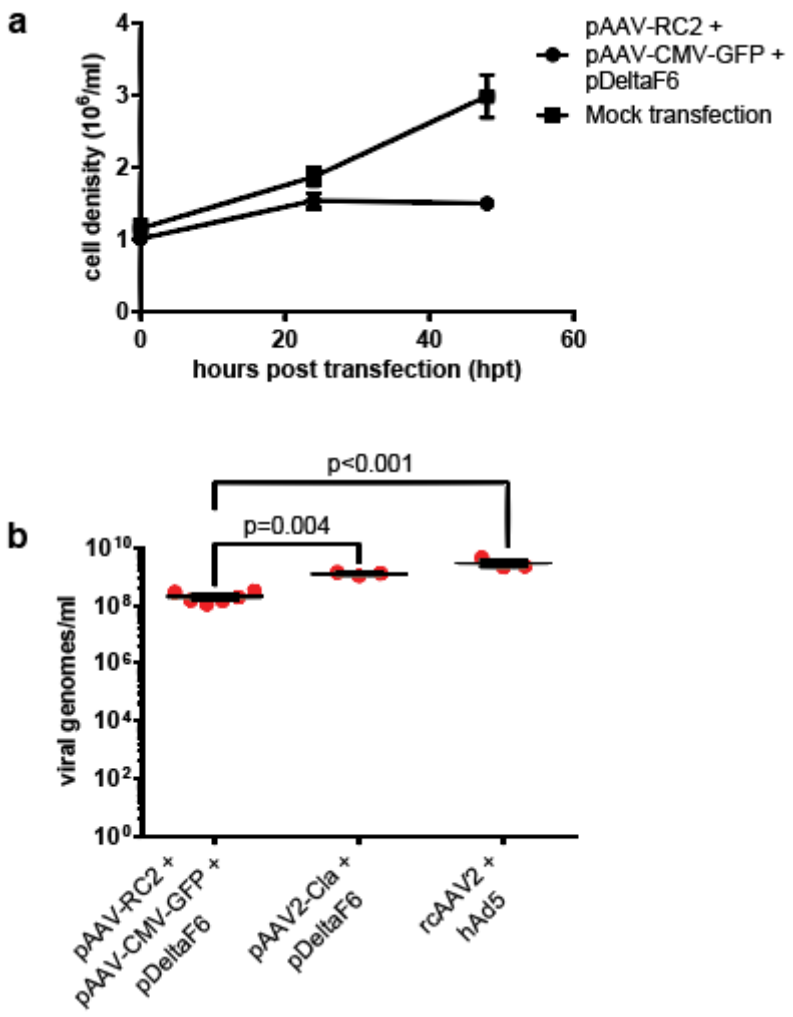


Figure 19. Production of rAAV2-GFP by plasmid transfection

HEK-293SF cells in serum-free suspension were transfected with equimolar amounts of pAdDeltaF6, pAAV-RC2, and pAAV-CMV-GFP to produce AAV2 vectors carrying a GFP transgene. These plasmids correspond to the helper, packaging, and transfer plasmids, respectively. a) Cell density in transfected and mock transfected cells tracked until harvest at 48 hpt. b) Volumetric vector yield in terms of nuclease-resistant vector genomes measured by digital droplet PCR. Also shown are two controls; transfection with equimolar amounts of pAdDeltaF6 + pAAV2-Cla (an infectious clone of wildtype AAV2) to produce replication competent AAV (rcAAC) by transfection, and infection with rcAAV and hAd5 at an MOI of 10 to produce rcAAV. Error bars represent SEM. Significance was determined using a non-parametric t-test with Welch's correction.

As seen in Figure 19a, cell density was relatively stable up to the harvest point at 48 hpt, the previously determined optimal harvest point for this process [238]. The mean vector yield 48 hpt was measured at 2.08 x 10^8 VG/ml, in line with previous studies where AAV vectors were produced by transfection in HEK-293 or derivative cell lines (Figure 19b). Two controls used in subsequent experiments were also assayed for AAV particle yield; transfection with an AAV infectious clone and helper plasmid to generate replication competent AAV (rcAAV) (1.26 x 10^9 VG/ml), as well as cells infected with rcAAV2 and a human adenovirus type 5 (hAd5) helper virus (2.97 x 10^9 VG/ml) (Figure 19b). It should be highlighted that the relatively crude measures of AAV yield used here do not fully capture the differences between rcAAV and AAV vectors; nearly 100% of rcAAV particles are infectious, whereas <1% of AAV vector particles are able to successfully transduce cells. The reasons behind this are poorly understood, but there is evidence that cis-acting sequences within the *REP* and *CAP* genes are necessary for efficient particle maturation[239].

6.3 Only a small fraction of cells produce assembled viral capsids during transfection-based manufacturing of AAV

To determine what proportion of cells in our triple transfection model produce fully assembled vector capsids, transfected cells were stained with a conformation-specific antibody that binds only to assembled viral capsids, allowing the measurement of this subpopulation by flow cytometry [240,241]. The presence of a GFP expression cassette on the transfer plasmid allowed simultaneous assessment of transfection efficiency. Cells infected with rcAAV2 and hAd5 helper virus were used as a positive control for capsid assembly, while cells transfected with an infectious AAV2 clone and helper plasmid were used to determine how capsid assembly

is impacted by transfection and/or the recombination of AAV2 genes. Results are shown in Figure 20.

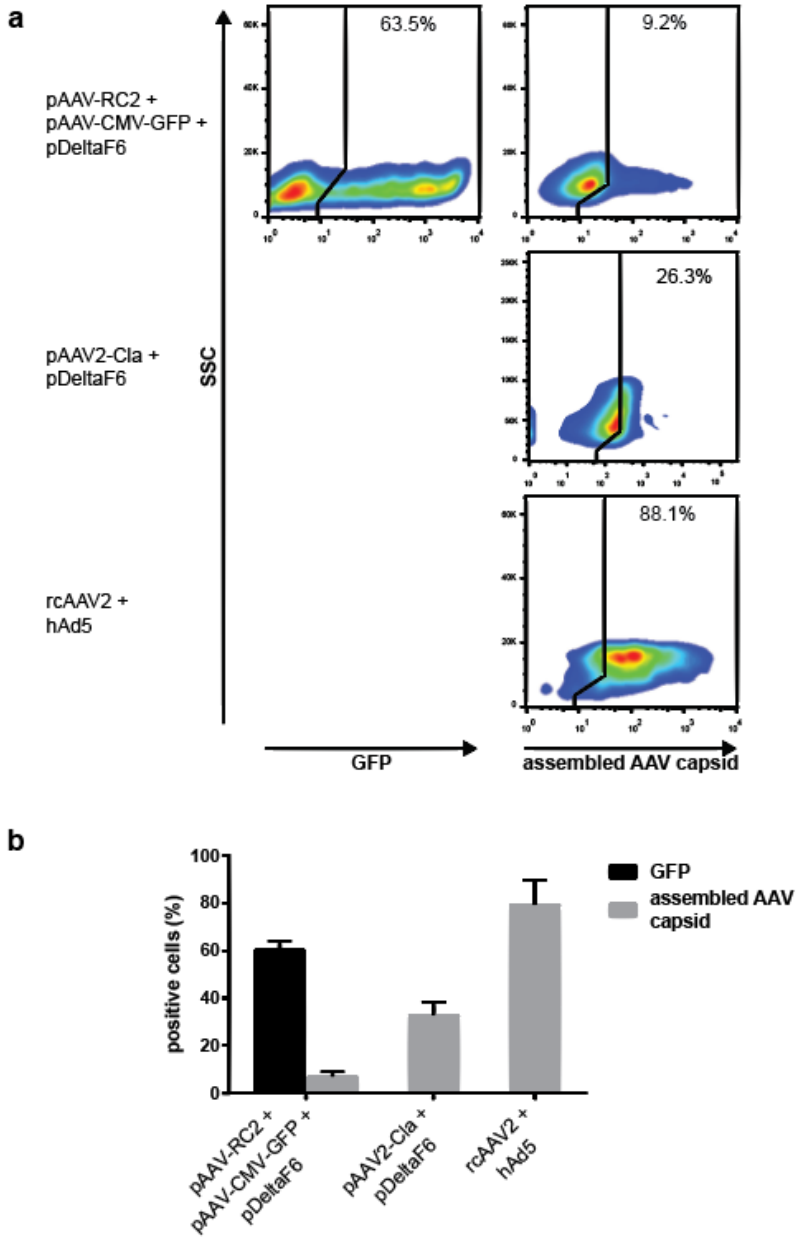


Figure 20. Comparison of transfection efficiency and assembled AAV capsid production in transfected cells producing AAV2-GFP vectors

HEK-293SF cells producing AAV2-GFP vectors by transfection were fixed 48 hpt and stained for assembled AAV capsids. a) Gating for transfection marker (GFP) and assembled AAV capsids following exclusion of debris and singlet gating. Positive gates were set using mock-transfected control set to 1-2% positive (not shown). Also shown are a positive control for capsid assembly consisting of cells infected with rcAAV2 and a hAd5 helper virus, and cells transfected with pAdDeltaF6 + pAAV2-Cla b) Quantification of flow cytometry results from n=3 biological replicates. Error bars represent SEM.

As expected, nearly all cells in the rcAAV2 and helper virus infected control are positive for assembled AAV capsid. However, despite a transfection efficiency of approximately 60%, only a small only a small fraction (*~7%*) cells in our triple transfection model produced measurable amounts of assembled AAV capsid. Interestingly, transfection with an AAV2 infectious clone and helper plasmid improves the proportion of cells positive for assembled AAV capsid by 4-5 fold (Figure 20), commensurate with the differences in particle yield observed in Figure 1b. Evidently the process of transfection is not inherently detrimental to AAV capsid assembly.

6.4 Attempts to utilize intracellular measurement of assembled AAV capsids as a screening reporter

The revelation that only a small fraction of cells produce assembled AAV capsids during transfection-based manufacturing of AAV vectors is interesting and previously unreported in literature. However, it does complicate any genome-wide screen since the population of cells in which we can measure AAV vector yield is now only ~7% of the total pool. If the screen were to enrich for the top 10% of AAV producers (is in the previous screen in section 4), this means the "high yield" sorted fraction consists of only 0.7% of the total pool. Maintaining library representation in the sorted fraction of cells is thus all but impossible, since for a typical library consisting of 70,000 unique elements the number of cells to be sorted would be approximated by:

$$cells\ to\ be\ sorted\ = \frac{(70{,}000\ cells)x\ (coverage\ of\ 200x)}{0.007} = 2.0\ \times 10^9$$

As previously discussed in section 2.17 though, samples in enriched conditions do not necessarily require full coverage. Enriched sgRNA can be highly overrepresented in these samples, and thus depending on signal to noise ratio a much smaller enriched pool may still be

sufficient to show statistically significant fold-changes in sgRNA abundance for hits. Attempts at both genome-wide knockout and genome-wide activation screens, conducted by the NRC and at McGill respectively, were thus undertaken. The knockout screen utilized the Toronto Knockout (TKO) library V3 and the activation screen utilized the Human CRISPRa Calabrese library (sgRNA set A)[242,243]. The FACS gating used in both screens is identical to that employed in Figure 20, sorting the top 25% of AAV-producing cells and collecting enriched fractions of 1-2x10^6 cells. Samples for the Calabrese activation screen conducted at McGill were processed and analyzed in a similar manner to those in section 4.0, except that Illumina NovaSeq rather than Illumina HiSeq was used to sequence the amplicon library. Selected quality control metrics for NovaSeq run and subsequent read mapping are shown in Figure 21.

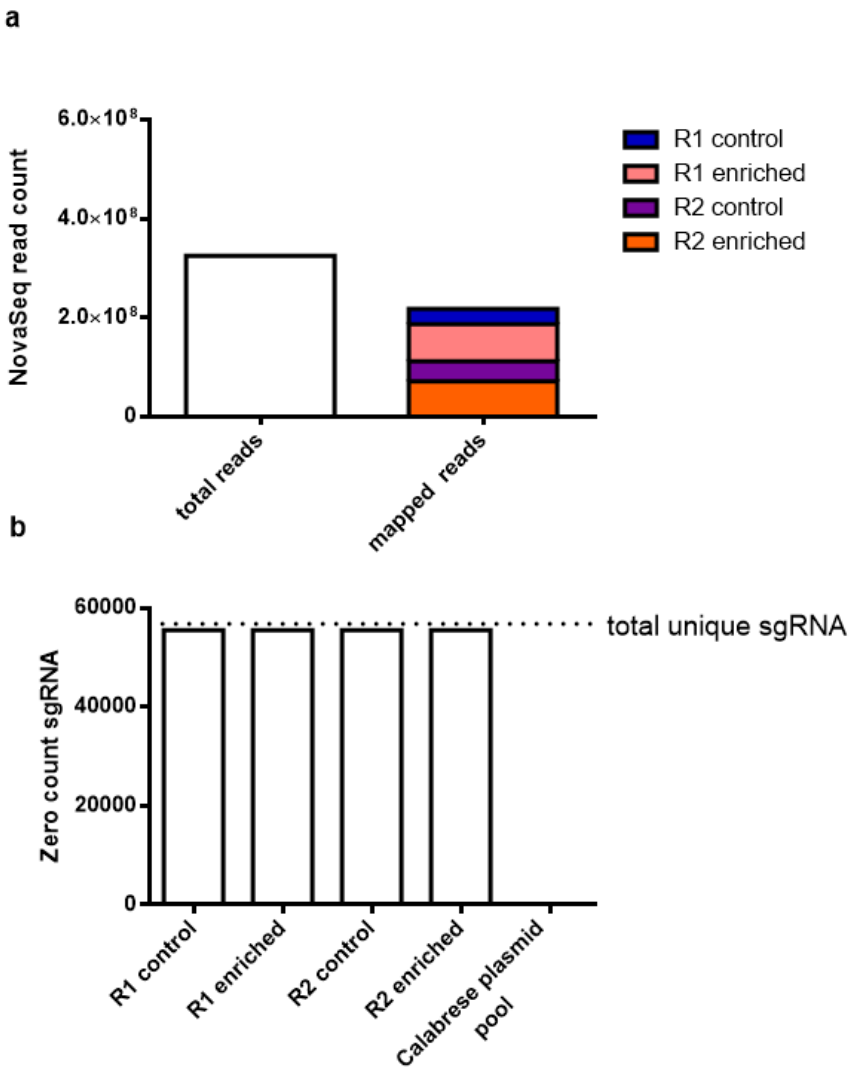


Figure 21. Selected quality control metrics for pooled genome-wide pooled Calabrese activation screen

The screen was carried out in duplicate (with respective replicates labeled R1 and R2), generating two control samples and two enriched samples. a) Comparison of total and mapped reads following NovaSeq of the amplicon library. b) Zero count sgRNA in respective samples following read mapping.

In total the NovaSeq run produced $3.3x10^8$ raw reads, 67% of which were successfully mapped to the Calabrese library index (Figure 21a). Mapped read distribution was unexpectedly skewed towards enriched samples, but the number of reads in the least represented sample (R1 control, $n=3.1x10^7$ reads) still corresponds to a coverage of >500x, significantly overshooting even the upper bound of recommended coverage for pooled screens.

As shown in Figure 21b, there was a loss of guide diversity in unsorted control fractions of the activation screen. Given that the plasmid library used to produce lentiviral vectors for this library displays an expected distribution of guides (Figure 21b), it seems likely that this pool underwent a population bottleneck during initial transduction and drug selection. The knockout screen conducted by collaborators at the NRC, which did not encounter population bottlenecks, was frought with other issues arising from the cell fixation process that was necessary for intracellular antibody staining. The fixation process cross-linked DNA to bound proteins and chemically modified the DNA itself, resulting in extracted DNA yields being reduced by 90% (data not shown). Use of a DNA extraction kit specifically designed for extraction from fixed tissues partially ameliorated this, but DNA recovery was still reduced by ~50%. Fixed cells also displayed a high binding affinity to polypropylene used in laboratory plasticware and pipette tips, and thus the number of cells in the already small sorted fractions was further reduced with every pipetting or transfer step.

In addition to (and perhaps more importantly than) the technical challenges encountered, attempts to optimize and repeat the planned genome-wide screens were highly impacted by the COVID-19 pandemic and the associated reagent shortages, reprioritizing of core facilities to address the pandemic, and reductions in support staff at McGill and the NRC. Ultimately the

decision was made by the NRC, the primary source of funding for this project, to withdraw support due to protracted delays.

6.5 Expression of a transfection marker does not reliably indicate co-expression of AAV proteins or helper virus factors

Attempts at using genome-wide CRISPR/Cas9 screening to intensify transfection-based manufacturing of AAV were unable to yield useable screening data, in part due to the low proportion of cells producing assembled virus capsids. However, this result was in and of itself interesting because it contradicts the prevailing assumption that all successfully transfected calls in transfection-based AAV manufacturing processes are productive[238,244]. This assumption is derived from the fact that commonly used cationic transfection reagents (polyethyleneimine, calcium phosphate, lipofectamine, etc.) coprecipitate heterogeneous plasmid mixtures into larger complexes for transit across the cell membrane, and so in theory all successfully transfected cells contain the genetic elements necessary to produce AAV [245-247]. Thus, the results shown in Figure 19 that indicate that transfection efficiency and the proportion of productive cells in culture vary by 10-fold was surprising and warranted further investigation.

The observation that only ~7% of transfected cells appeared to produce assembled AAV capsids raised the question as to what proportion of cells were expressing the necessary factors for AAV vector production. The three transfected plasmids collectively encode ~17 protein and RNA elements, making measurement of every factor impractical [18,20]. Instead, a subset of proteins from each of the three transfected plasmids were visualized in our triple transfection model and controls via immunofluorescence.

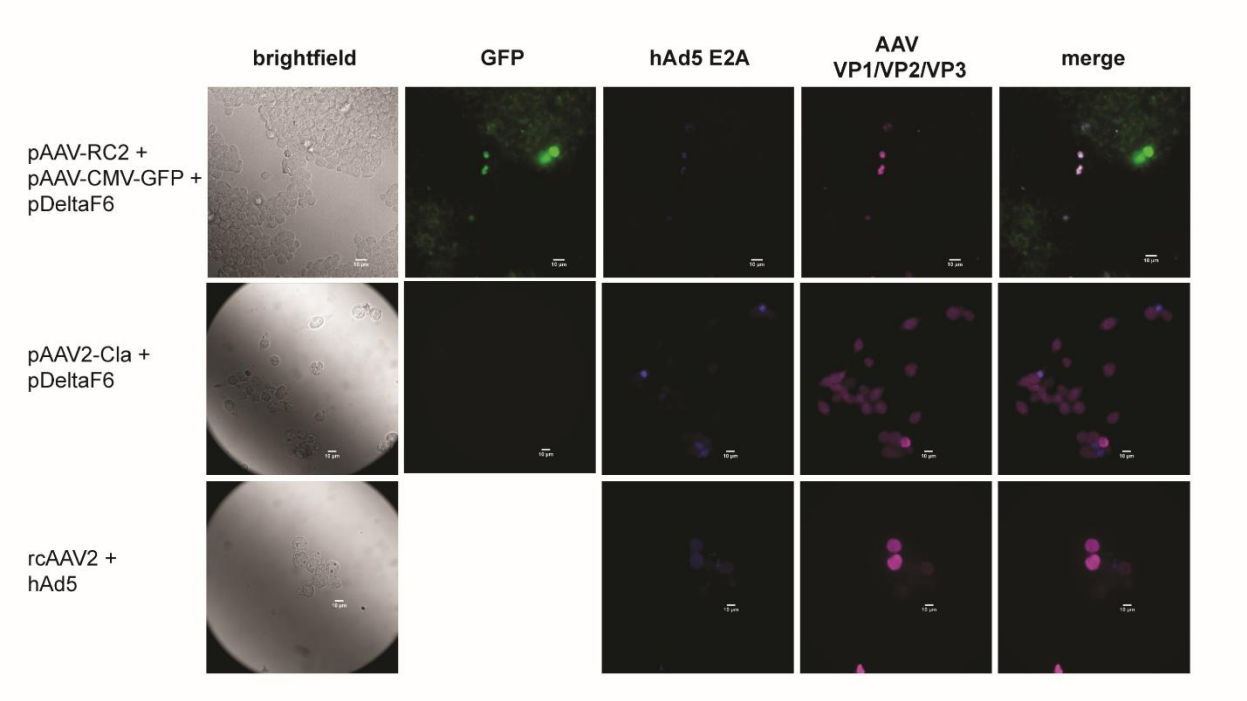


Figure 22. Expression of transfected plasmids during AAV2-GFP vector production

Immunofluorescence of transfected cells producing AAV2-GFP vectors 48 hpt. Images show the expression of GFP from the transfer plasmid, hAd5 E2A from the helper plasmid, and AAV capsid monomers VP1, VP2, and VP3 from the packaging plasmid. Images are taken at 40x magnification. Scale bar is 10µm.

As shown in Figure 22, GFP was broadly expressed in our triple transfection model, consistent with the flow cytometry results in Figure 20. Interestingly though, GFP expression does not seem to reliably indicate coexpression of AAV capsid monomers or the hAd5 E2A helper factor, both of which were expressed at detectable levels in a much smaller subset of cells. This may partially explain the observation in Figure 20b that only ~7% transfected cells contain assembled AAV capsids, despite an apparent transfection efficiency nearly tenfold higher. Transfection with an infectious AAV2 clone and helper plasmid dramatically increased expression of AAV capsid monomers compared to our triple transfection model; a predictable result given that *CAP* copy number is static during vector production, but exponentially increasing in systems with rcAAV. In contrast, hAd5 E2A was weakly expressed in both transfection systems compared to cells infected with rcAAV and hAd5 virus. This is intriguing given that the two transfection systems show substantial differences in particle yield (Figure 19b) and proportion of cells positive for capsid assembly (Figure 20). Speculatively, this could indicate that the threshold for effective helper factor expression is low relative to other factors necessary for AAV vector production, or highlight the importance of cis-acting elements within the REP and CAP genes in capsid assembly.

7.0 DISCUSSION

In this work, I designed and executed a novel pooled, genome-wide CRISPR/Cas9 screening strategy to identify and rank host restriction factors in vaccine-producing cell lines, using the HEK-293SF cell line and PR/8 influenza as a model. The results of this screen were then used to direct the creation of high yield, multiple knockout cell lines for influenza vaccine production. In an attempt to improve upon the indirect reporter system used in the aforementioned screen, I also explored the use of conformation specific antibodies to allow direct measurement intracellular assembled capsids during the transfection-based manufacturing of AAV vectors.

In contrast to previous arrayed screens to identify host restriction factors in cell-based vaccine production systems, the use of a replication defective reporter virus and a pooled screen format allows genome-wide coverage without high-throughput robotics. The coculture of all cells in a given replicate also removes much of the inherent variability seen in large-scale arrayed screens[178]. Using HEK-293SF cells and PR/8 influenza as a model, we identified 64 putative restriction factors that met the threshold for biological ($lfc>0.4$) and statistical ($p<0.01$) significance. The antiviral function of many of these restriction factors is implied from the fact that they are direct or indirect targets of influenza during infection; nuclear pore components like RAE1, NUP98, NXF1, and NXT1 are downregulated by influenza during infection, which has been shown to benefit influenza replication[248]. Similarly, TSC1 and TSC2 negatively regulate the mTORC1 complex, which influenza has been shown to activate during late-stage replication to maximize viral protein production[249,250]. Other factors identified have known antiviral activity against influenza or similar viruses; DDX6 and DPC2 encode components of the decapping

complex, which has been shown to inhibit the expression of Bunyavirus mRNA by competing with the viral "cap snatching" pathway (similar to that utilized by influenza) for the 5' cap of host mRNA[251]. DDX6 has also been shown to interact with an RNA Hairpin in the 39 UTR of the Dengue Virus Genome, as well as augment RIG-I signaling[252,253]. SMG9 and UPF2 encode components of the nonsense mediated RNA decay (NMD) pathway, which normally functions to degrade aberrant transcripts with internal stop codons[254]. Internal stop codons are a common feature of viral transcripts, including those of influenza. Accordingly, NMD has been shown to exert antiviral effects on several RNA viruses[255]. Inhibition of arginine methyltransferase CARM1 activity has been shown to reverse HIV latency[256].

While antiviral function of many putative restriction factors identified in the screen is apparent, n=41 of the factors identified have no association with a GO viral or antiviral process (See Sharon *et al.* (2020) supplemental data S6)[1]. The results of this study therefore serve not only to guide the development of high yield knockout cell lines for influenza vaccine production, but also as a basis for studies into novel host antiviral pathways.

Interestingly, though by no means absent from the results, key mediators of canonical antiviral pathways, such as interferon and RIG-I, had a relatively minor impact on viral yield in this study. One explanation for this is that many vaccine production cell lines such as HEK-293, Vero, and MDCK show impaired antiviral signaling compared to primary cells, hence their inherent utility for virus production[12,23,257]. Another aspect to consider is that *in vivo* a major function of these pathways is to upregulate antigen presentation and recruit specialized immune cells, which is obviously not possible within *in vitro* cell culture systems[101,258]. Far more represented in the results are metabolic factors, particularly negative regulators of nucleic acid anabolism and translation (Figure 10a). Even the highly represented non-canonical "antiviral"

pathways identified, such as the decapping pathway, are activated not just in response to viral infection, but nutrient starvation as well. Together these results imply that in cell-based vaccine production systems, influenza yield is primarily determined by host cell metabolic state, rather than innate antiviral immunity, though there is significant overlap between these pathways.

The screening strategy presented here can in theory be applied to any cell-based vaccine production system. Even within the space of a single virus-host system, the FACS collection parameters can be adjusted depending on the end-use of the data. In this screen, for instance, we used a 1.5 drop yield mode and a high event rate during FACS acquisition, allowing a significant amount (~20%) of cells not in the 90th percentile of GFP expressing cells to contaminate the high yield fraction. In doing so, we sacrificed readout signal dynamic range to collect large numbers of cells and maintain library representation. This allowed us to generate a large list of hits for rich downstream bioinformatics analysis, while potentially compromising our ability to accurately rank those hits. Alternatively, if one was solely concerned with accurately ranking a smaller list of highly significant restriction factors for knockout cell line development, FACS acquisition could be run under more stringent conditions and with a lower fraction of cells collected (e.g. the 99th percentile of GFP expression). It should be noted though that even with the low dynamic range of the screen (± lfc1.2) the ranking of putative restriction factors was remarkably consistent between replicates; DDX6, for instance, was the top ranked gene in every replicate (see Sharon *et al.* (2020) supplemental data S4)[1].

While the screen presented here is a useful tool for identifying and ranking viral restriction factors, it does have limitations. There is no way to differentiate knockouts that induce a generic upregulation of translation from restriction factors that specifically inhibit viral replication, though this could be easily remedied with a stably expressed passive fluorescent

reporter in host cells. Furthermore, only restriction factors with an impact on viral protein expression can be identified, while those impacting viral packaging and egress are not. Finally, the use of an indirect reporter reporter (the translation of GFP inserted into the HA reading frame) to measure a phenotype (active HA in mature virions) introduces the potential for false positive results. Though this can be mitigated somewhat by careful validation of the reporter system, as was conducted in this case, genetic perturbation can have unpredictable effects on the biology of the cell, leading to decoupling of normally correlated metrics.

Attempts to translate the results of the screen into high yield, multiple knockout cells for influenza vaccine production were successful in that the best performing MK clones exhibited a 10-fold increase in the volumetric yield of in terms of VG/mL and a 2-fold increase in active HA yield. This is an interesting result, as VG/mL and the volumetric yield of HA are normally quite tightly correlated[237]. While the measurements of VG/mL may indicate free vRNA in the culture media, this is unlikely given that unprotected RNA is quickly degraded in culture by RNAses. An alternative explanation is that the knockouts created an environment within the cell that was favourable for vRNA replication and viral protein translation, but not necessarily incorporation of additional active HA into mature virions. The latter process requires not only HA translation, but also chaperone-dependent folding of HA monomers, trimerization, several post-translational modifications, and efficient transit through the ER and TGA to reach the cellular membrane[56]. By contrast, the packaging of vRNA into virions requires only the cytosolic expression of viral proteins necessary to form vRNPs, as well as the M1 matrix protein to drive budding. Results obtained are consistent with those reported by Karlas *et al.* (2010), which showed knockout of NUP98 in HEK-293 greatly increased influenza vRNA levels and viral protein translation but

was detrimental to infectious viral titer[259]. This example is especially relevant given that NUP98 was also a top hit in this screen (Figure 9b).

The false positive screening artefacts observed in this study could be reduced by adopting a more direct measure of HA yield as the screen reporter. A recent pooled genome-wide screen for influenza host dependency factors conducted by Li *et al.* (2020) used staining for surface expression of HA on host cells as a basis for pool enrichment by FACS[260]. This reporter not only directly measures host cell HA yield, but also ensures that the HA measured has been successfully processed through the ER and TGA. However, even this strategy does not guarantee that the detected HA is incorporated into budding virions, and theoretically may specifically enrich for mutant subpopulations that inhibit budding and keep surface HA associated with the host cell. Since viral particle maturation and egress are simultaneous processes with influenza and other enveloped viruses, the ideal screening reporter would not be reliant on physical association of a barcoded cell and the virus it produces. One way to accomplish this would be to barcode the HA itself with a library of unique peptide tags that are optimized for deconvolution by mass spectroscopy, such as those reported by Egloff *et al.* (2019)[261]. The speculative library used for such a screen would feature lentiviral vectors that encode expression cassettes for both sgRNA and an associated barcoded HA. Following library transduction, cells would then be infected with HA-deleted influenza (similar to the PR/8$^{GFP\Delta HA}$ reporter virus used in this study) to allow the incorporation of the barcoded HA onto virions[208]. Rather than the relative abundance of barcoded cells in a FACS enriched population, the readout for the screen would be relative abundance of HA-associated peptide tags detected on a pool of influenza viral particles. Though an interesting concept, the library size of this speculative screen would be entirely dependent on the resolution of the mass spectroscopy technique and its

ability to quantitatively measure the abundance of thousands of unique peptide tags in an unbiased manner.

The results of section 5 also highlight the technical challenges inherent in creating multiple knockout cell lines, such as achieving a high knockout efficiency for several genes simultaneously, and the differing ability of the resultant cell lines to thrive in bioreactor culture compared to the parental line. The apparent drop in sgRNA editing efficiency observed in Figure 14 when CRISPR/Cas9 constructs are transfected together (as opposed to individually) could be due to non-specific growth effects caused by multiple dsDNA breaks, which would result in enrichment of unedited cells in the transfected pool. The results of Figure 8b support this, clearly indicate that dsDNA cleavage at only a single loci strongly impacts HEK-293SF growth. Similar non-specific growth effects following dsDNA cleavage by Cas9, proportional to the number of dsDNA breaks induced, have been reported by other authors[229]. A sequential transfection strategy (as opposed to simultaneous transfection of all CRISPR/Cas9 constructs) might therefore improve editing efficiency.

It's also important to note that although in theory any knockouts that impact cell survival will drop out of a pooled CRISPR screen, in practice this is not the case. Parallel screens for cellular fitness genes conducted using RNAi and CRISPR libraries show little overlap in their results, suggesting a large subset of essential gene knockouts do not drop out of CRISPR screens as expected[262]. Indeed, it has been previously reported that homozygous knockouts of genes identified in genome-wide screens aiming to increase polio vaccine yield are lethal[201]. The reasons for this are unclear, but it's speculated that due to differences in mRNA and protein turnover some lethal knockout phenotypes can take a protracted amount of time to manifest, longer than the 9-12 days a pooled screen can be run before library coverage in the pool is lost

due to population drift[262]. This would also explain why lethal knockouts would not have been detected during evaluation of single gene knockout pools in Figure 12a, as these pools were only cultured for ~10 days before evaluation (as opposed to the ~6 weeks required to clonally expand and evaluate the MK clones). It follows then that homozygous knockout of some of screen hits that were incorporated into the MK lines, particularly genes such as NXT1 and NXF1 that impact fundamental processes like shuttling of mRNA across the nuclear membrane, may ultimately be incompatible with cell survival[84,248]. In addition to reducing the apparent editing efficiency of sgRNA that target these particular genes, the coprecipitation of plasmids transfected using cationic transfection reagents (such as the PEIpro used in this study) means that these lethal sgRNA would have likely been co-expressed with any other sgRNA[245-247]. In this way, even one sgRNA that negatively impacts cell survival acts as a "poison pill" during simultaneous transfection of multiple sgRNA expression constructs, resulting in dropout of all cells that were efficiently transfected and reducing the apparent editing efficiency of all sgRNA. This may explain the generalized reduction in sgRNA efficiency observed in Figure 15, and further argues for a sequential, rather than simultaneous, transfection strategy.

Although barcoding to allow pooling of amplicon libraries for deep sequencing is a standard technique, this is generally carried out using singleplex PCR reactions, with a single primer pair and associated amplicon per reaction. By contrast the two-step, multiplex PCR barcoding strategy described in Figure 14 allowed simultaneous barcoding of amplicons generated from different sgRNA target loci in the same reaction. Variations of this technique have been previously described for large-scale genotyping in ecology studies, but to my knowledge this is the first time it has been applied to clone screening in cell line development[263]. Though this technique was applied here to genotype 160 clones with an average ploidy of n=3.2

at 5 different loci (~2560 unique reads), this was a severe underutilization of the sequencing power available. An Illumina NovaSeq run can generate roughly 1.5 billion paired-end reads, and so at a generous coverage of 500x over 1000-fold more clones or sgRNA target loci could have been surveyed in a single sequencing run. Thus, this technique may have utility as a scalable and efficient sample preparation method for parallel screening of extremely large panels of clones harboring multiple genomic edits.

The poor growth of MK30 mutant in bioreactor culture highlights that cell line engineering may also necessitate process redesign to accommodate the divergent biology of modified cell lines. Assuming that poor growth was due to increased susceptibility to sheer stress, as the results of Figure 18b suggest, growth of the MK30 mutant may have been improved by reducing the agitation rate in the bioreactor or switching to an impellerless design (e.g. the WAVE bioreactor)[264]. Alternatively, if one wishes to avoid the need for additional process development, the pooled screen itself can be conducted under the same conditions are the current process (i.e. in the bioreactor). In this way, any mutant subpopulations that fail to thrive under the existing process conditions will be negatively selected for and drop out of the pool.

As discussed previously, the selection of an appropriate screening reporter is critical to obtaining accurate results from a pooled genome-wide CRISPR screen. The results of section 6 serve as a convincing demonstration that at least for viral vectors such as adenovirus and AAV vectors where virions mature intracellularly, direct measurement of assembled viral capsids is possible. This screening reporter represents a significant improvement over the PR/8 $^{GFP\Delta HA}$ reporter virus, as it allows mutations that impact the viral lifecycle downstream of viral protein translation to be detected in the screen. It should be noted though that this is still not a perfect

proxy for vector yield, as AAV vectors are prone to assembling capsids without encapsulating a genome[239].

More broadly, the demonstration that pooled genome-wide screening could be used to intensify the production of not just viral vaccines, but also viral gene therapy vectors, is important to maintaining the relevance of the technology as mRNA vaccines become more prominent. Viral vaccines currently make up the bulk of viral biologics produced. However, the COVID-19 pandemic and the approval of two highly effective mRNA vaccines has marked a paradigm shift in vaccinology. mRNA vaccines allow intracellular expression of viral antigens, and are therefore far more effective at promoting a cell-mediated immune response than inactivated virion or subunit vaccines, which rely almost exclusively on humoral immunity for their protective effect[265]. Notably, cell-mediated immunity has been shown to be considerably more refractory to antigenic drift, facilitating superior protection against novel COVID-19 variants and, theoretically, seasonal influenza endemics[266]. While live-attenuated vaccines are capable of inducing cell-mediated immunity, mRNA vaccines also have several advantages from a manufacturing and regulatory standpoint. mRNA vaccines provide a generic platform for which vaccines for virtually any virus can be manufactured with minimal changes to the production process. This contrasts with cell-based vaccine production, which generally requires selection of an appropriate cell line and viral strain, upstream and downstream process optimization, and development of analytics to characterize the product for each new vaccine. The use of vectored vaccines, such as Oxford's ChAdOX1-nCoV and CanSino's Ad5-nCOV COVID-19 vaccines, does circumvent the need for entirely *de novo* process development, but these are still inherently slower to produce than mRNA vaccines due to the need to clone, grow, and purify the vector from culture[267]. mRNA vaccines are also more attractive from a regulatory

standpoint, with the product being inherently free of replication competent viruses, adventitious agents, or much of the natural variability seen in viral vaccines. Viral vaccines are still far more economical to produce than mRNA vaccines and do not require a -80^{o}C cold chain for transport, a critical difference particularly for vaccinating populations in the developing world, but as mRNA technology matures these advantages are expected to diminish[268]. Considering all the above, it seems likely that mRNA vaccines will eventually replace cell-based manufacturing of vaccines for influenza and many other viral pathogens. Indeed, multiple mRNA vaccines for influenza are already in the development pipeline[269]. While mRNA therapeutics may supplant viral biologics as the dominant vaccine platform, viral biologics remain the platform of choice gene therapy applications due to their ability to confer long-term transgene expression. The last decade has seen a resurgence of interest in viral vectored gene therapies, with 250 ongoing clinical trials utilizing AAV vectors alone as of 2021[16].

The observation that only ~7% of transfected cells appeared to produce assembled AAV capsids during transfection-based manufacturing of AAV2 vectors has significant implications for optimization of this platform. As seen in Figure 19b, the difference in yield between AAV produced by transfection versus those produced by infection with rcAAV and an Ad5 helper virus (roughly 10-fold) is approximately proportional to the difference in productive cells between the two cultures (Figure 20). Taken together, this implies that the reduced yields in transfection-based AAV vector production are not due to lower cell specific productivity, but simply that >90% of the culture biomass is nonproductive.

Interestingly, the proportion of productive cells in culture can be improved 6-fold (with a commensurate increase in yield) when cells are cotransfected with the pAAV2-Cla infectious clone and the pDeltaF6 helper plasmid (Figure 20b). This indicates that the process of

transfection does not inherently reduce the proportion of productive cells in culture, nor is the expression of helper virus factors from a plasmid wholly limiting. It would seem then that the proportion of productive cells during transfection-based manufacturing of AAV2 vectors is instead limited by insufficient expression of REP and CAP from the pRC-AAV2 packaging plasmid. This follows logically from the fact that the copy number of REP and CAP increases in cultures transfected with the pAAV2-Cla infectious clone (as opposed remaining static during the production of replication defective vectors), and is further supported by confocal microscopy data in Figure 20 that shows that despite high transfection efficiency, few cells in the triple transfection model express detectable levels of AAV capsid monomers[20]. This conclusion is supported by DOE optimization studies conducted by Zhou *et al*. (2020), which showed that by increasing the proportion of packaging plasmid (in addition to other optimization steps) AAV vector yields in the order of 10^{10} VG/mL could be obtained for a wide range of AAV serotypes[270].

The results of Figure 20 also demonstrate that the expression of a transfection marker does not necessarily imply a cell is producing AAV vector particles. While transfection efficiency remains an important process development metric, the conformation-specific antibody staining against assembled AAV capsids that was used here may also prove useful in the future development and optimization of transfection-based AAV vector platforms.

8.0 CONCLUSION

The pooled genome-wide CRISPR/Cas9 knockout screen identified n=135 putative influenza restriction factors, n=41 of which were not previously known to impact influenza replication. Subsequent GO and PPI analysis of the genes identified demonstrates that the genetic determinants of influenza yield in the HEK-293SF cell line are primarily metabolic, a striking contrast to the innate immune factors identified in previous screens on other cell lines intended to model human infection. This finding has broader implications for cell line engineering efforts for HEK-293SF and other immortalized cell lines used to produce viral vaccines and gene therapy vectors.

Attempts detailed in section 5.0 to apply screening data towards the creation of high-yield multiple knockout cell lines for influenza vaccine production were highly instructive on several fronts. It was shown that individual single guide RNA (sgRNA) efficiency can be adversely impacted by simultaneous expression of multiple constructs. This finding suggests that multiple knockouts should be induced sequentially, rather than in parallel, when constructing multiple knockout cell lines. Section 5.0 also highlights the dangers of using indirect reporters as a proxy for viral replication in pooled genome-wide screens. The results demonstrate that even with careful validation of the reporter system, the genetic perturbations induced by screening can cause decoupling of normally correlated biological processes (in this case total viral particles and the yield of active HA protein), leading to unpredictable results.

The conformation-specific antibody staining to quantify the percentage of cells that contain assembled viral capsids during the transfection-based production of AAV2 vectors demonstrates proof of concept for direct reporter systems in pooled genome-wide CRISPR

screens on non-enveloped viral biologics. Such screens may reduce screen artefacts that hamper indirect reporter screens. Though I was not able to carry out a screen using this reporter, validation of the reporter system revealed the surprising result that despite achieving transfection efficiencies in excess of 60% and nominal yields for the process, only ~7% of cells in culture produced detectable amounts of assembled AAV capsid during transfection-based manufacturing of AAV2 vectors. Subsequent experiments indicate that this is due to insufficient expression of AAV capsid monomers. This finding has important implications for future efforts to optimize this widely used vector production platform, and suggests that the flow cytometry assay developed here for AAV capsid assembly may prove a useful process development metric for this purpose.

9.0 REFERENCES

1 Sharon *et al.* A pooled genome-wide screening strategy to identify and rank influenza host restriction factors in cell-based vaccine production platforms. *Sci Rep* **10**, 12166, doi:10.1038/s41598-020-68934-y (2020).

2 Dash, S., Sharon, D. M., Mullick, A. & Kamen, A. A. Only a small fraction of cells produce assembled capsids during transfection-based manufacturing of adeno-associated virus vectors. *Biotechnol Bioeng* **119**, 1685-1690, doi:10.1002/bit.28068 (2022).

3 Paget, J. *et al.* Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. *J Glob Health* **9**, 020421, doi:10.7189/jogh.09.020421 (2019).

4 Cozza, V. *et al.* Global Seasonal Influenza Mortality Estimates: A Comparison of 3 Different Approaches. *Am J Epidemiol* **190**, 718-727, doi:10.1093/aje/kwaa196 (2021).

5 Harding, A. T. & Heaton, N. S. Efforts to Improve the Seasonal Influenza Vaccine. *Vaccines (Basel)* **6**, doi:10.3390/vaccines6020019 (2018).

6 Barr, I. G. *et al.* Cell culture-derived influenza vaccines in the severe 2017-2018 epidemic season: a step towards improved influenza vaccine effectiveness. *NPJ Vaccines* **3**, 44, doi:10.1038/s41541-018-0079-z (2018).

7 Bouvier, N. M. The Future of Influenza Vaccines: A Historical and Clinical Perspective. *Vaccines (Basel)* **6**, doi:10.3390/vaccines6030058 (2018).

8 Simonsen, L. *et al.* Global mortality estimates for the 2009 Influenza Pandemic from the GLaMOR project: a modeling study. *PLoS Med* **10**, e1001558, doi:10.1371/journal.pmed.1001558 (2013).

9 Shrestha, S. S. *et al.* Estimating the burden of 2009 pandemic influenza A (H1N1) in the United States (April 2009-April 2010). *Clin Infect Dis* **52 Suppl 1**, S75-82, doi:10.1093/cid/ciq012 (2011).

10 Herfst, S. *et al.* Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534-1541, doi:10.1126/science.1213362 (2012).

11 Milian, E. *et al.* Accelerated mass production of influenza virus seed stocks in HEK-293 suspension cell cultures by reverse genetics. *Vaccine* **35**, 3423-3430, doi:10.1016/j.vaccine.2017.04.065 (2017).

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

13 Daya, S. & Berns, K. I. Gene therapy using adeno-associated virus vectors. *Clin Microbiol Rev* **21**, 583-593, doi:10.1128/CMR.00008-08 (2008).

14 Duan, D., Yan, Z., Yue, Y. & Engelhardt, J. F. Structural analysis of adeno-associated virus transduction circular intermediates. *Virology* **261**, 8-14, doi:10.1006/viro.1999.9821 (1999).

15 NIH, N. I. o. H. *clinicaltrials.gov*, <clinicaltrials.gov> (2017).

16 Bulcha, J. T., Wang, Y., Ma, H., Tai, P. W. L. & Gao, G. Viral vector platforms within the gene therapy landscape. *Signal Transduct Target Ther* **6**, 53, doi:10.1038/s41392-021-00487-6 (2021).

17 Morrison, C. $1-million price tag set for Glybera gene therapy. *Nat Biotechnol* **33**, 217-218, doi:10.1038/nbt0315-217 (2015).

18 Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov* **18**, 358-378, doi:10.1038/s41573-019-0012-9 (2019).

19 Clement, N. & Grieger, J. C. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Mol Ther Methods Clin Dev* **3**, 16002, doi:10.1038/mtm.2016.2 (2016).

20 Sharon & Kamen, A. Advancements in the design and scalable production of viral gene transfer vectors. *Biotechnol Bioeng* **115**, 25-40, doi:10.1002/bit.26461 (2018).
21 Cawood, R. *Redefining AAV Manufacturing: The Time Is Now*, <https://www.genengnews.com/magazine/redefining-aav-manufacturing-the-time-is-now/> (2020).
22 Chua, S., Cui, J., Engelberg, D. & Lim, L. H. K. A Review and Meta-Analysis of Influenza Interactome Studies. *Front Microbiol* **13**, 869406, doi:10.3389/fmicb.2022.869406 (2022).
23 Reynolds, A. *et al.* Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA* **12**, 988-993, doi:10.1261/rna.2340906 (2006).
24 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).
25 Ejiri, H. *et al.* Characterization of a novel thogotovirus isolated from Amblyomma testudinarium ticks in Ehime, Japan: A significant phylogenetic relationship to Bourbon virus. *Virus Res* **249**, 57-65, doi:10.1016/j.virusres.2018.03.004 (2018).
26 Sameroff, S. *et al.* Novel quaranjavirus and other viral sequences identified from ticks parasitizing hunted wildlife in Trinidad and Tobago. *Ticks Tick Borne Dis* **12**, 101730, doi:10.1016/j.ttbdis.2021.101730 (2021).
27 Cottet, L., Rivas-Aravena, A., Cortez-San Martin, M., Sandino, A. M. & Spencer, E. Infectious salmon anemia virus--genetics and pathogenesis. *Virus Res* **155**, 10-19, doi:10.1016/j.virusres.2010.10.021 (2011).
28 Bouvier, N. M. & Palese, P. The biology of influenza viruses. *Vaccine* **26 Suppl 4**, D49-53, doi:10.1016/j.vaccine.2008.07.039 (2008).
29 Su, S., Fu, X., Li, G., Kerlin, F. & Veit, M. Novel Influenza D virus: Epidemiology, pathology, evolution and biological characteristics. *Virulence* **8**, 1580-1591, doi:10.1080/21505594.2017.1365216 (2017).
30 Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza A virus species specificity. *Nat Rev Microbiol* **17**, 67-81, doi:10.1038/s41579-018-0115-z (2019).
31 Yang, W., Schountz, T. & Ma, W. Bat Influenza Viruses: Current Status and Perspective. *Viruses* **13**, doi:10.3390/v13040547 (2021).
32 Taubenberger, J. K. & Kash, J. C. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* **7**, 440-451, doi:10.1016/j.chom.2010.05.009 (2010).
33 Mostafa, A., Abdelwhab, E. M., Mettenleiter, T. C. & Pleschka, S. Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview. *Viruses* **10**, doi:10.3390/v10090497 (2018).
34 Dudas, G., Bedford, T., Lycett, S. & Rambaut, A. Reassortment between influenza B lineages and the emergence of a coadapted PB1-PB2-HA gene complex. *Mol Biol Evol* **32**, 162-172, doi:10.1093/molbev/msu287 (2015).
35 Matsuzaki, Y. *et al.* Genetic Lineage and Reassortment of Influenza C Viruses Circulating between 1947 and 2014. *J Virol* **90**, 8251-8265, doi:10.1128/JVI.00969-16 (2016).
36 Matsuzaki, Y. *et al.* A nationwide epidemic of influenza C virus infection in Japan in 2004. *J Clin Microbiol* **45**, 783-788, doi:10.1128/JCM.01555-06 (2007).
37 Sederdahl, B. K. & Williams, J. V. Epidemiology and Clinical Characteristics of Influenza C Virus. *Viruses* **12**, doi:10.3390/v12010089 (2020).
38 Dadonaite, B. *et al.* The structure of the influenza A virus genome. *Nat Microbiol* **4**, 1781-1789, doi:10.1038/s41564-019-0513-7 (2019).
39 Stubbs, T. M. & Te Velthuis, A. J. The RNA-dependent RNA polymerase of the influenza A virus. *Future Virol* **9**, 863-876, doi:10.2217/fvl.14.66 (2014).

40 Graef, K. M. *et al.* The PB2 subunit of the influenza virus RNA polymerase affects virulence by interacting with the mitochondrial antiviral signaling protein and inhibiting expression of beta interferon. *J Virol* **84**, 8433-8445, doi:10.1128/JVI.00879-10 (2010).
41 Yamayoshi, S., Watanabe, M., Goto, H. & Kawaoka, Y. Identification of a Novel Viral Protein Expressed from the PB2 Segment of Influenza A Virus. *J Virol* **90**, 444-456, doi:10.1128/JVI.02175-15 (2016).
42 Gibbs, J. S., Malide, D., Hornung, F., Bennink, J. R. & Yewdell, J. W. The influenza A virus PB1-F2 protein targets the inner mitochondrial membrane via a predicted basic amphipathic helix that disrupts mitochondrial function. *J Virol* **77**, 7214-7224, doi:10.1128/jvi.77.13.7214-7224.2003 (2003).
43 Cheung, P. H. *et al.* Virus subtype-specific suppression of MAVS aggregation and activation by PB1-F2 protein of influenza A (H7N9) virus. *PLoS Pathog* **16**, e1008611, doi:10.1371/journal.ppat.1008611 (2020).
44 Cheung, P. H., Lee, T. T., Chan, C. P. & Jin, D. Y. Influenza A virus PB1-F2 protein: An ambivalent innate immune modulator and virulence factor. *J Leukoc Biol* **107**, 763-771, doi:10.1002/JLB.4MR0320-206R (2020).
45 Mazur, I. *et al.* The proapoptotic influenza A virus protein PB1-F2 regulates viral polymerase activity by interaction with the PB1 protein. *Cell Microbiol* **10**, 1140-1152, doi:10.1111/j.1462-5822.2008.01116.x (2008).
46 Hu, J., Ma, C. & Liu, X. PA-X: a key regulator of influenza A virus pathogenicity and host immune responses. *Med Microbiol Immunol* **207**, 255-269, doi:10.1007/s00430-018-0548-z (2018).
47 Gaucherand, L. *et al.* The Influenza A Virus Endoribonuclease PA-X Usurps Host mRNA Processing Machinery to Limit Host Gene Expression. *Cell Rep* **27**, 776-792 e777, doi:10.1016/j.celrep.2019.03.063 (2019).
48 Shi, M. *et al.* Evolutionary conservation of the PA-X open reading frame in segment 3 of influenza A virus. *J Virol* **86**, 12411-12413, doi:10.1128/JVI.01677-12 (2012).
49 Russell, C. J., Hu, M. & Okda, F. A. Influenza Hemagglutinin Protein Stability, Activation, and Pandemic Risk. *Trends Microbiol* **26**, 841-853, doi:10.1016/j.tim.2018.03.005 (2018).
50 Krammer, F. The human antibody response to influenza A virus infection and vaccination. *Nat Rev Immunol* **19**, 383-397, doi:10.1038/s41577-019-0143-6 (2019).
51 Wei, C. J. *et al.* Next-generation influenza vaccines: opportunities and challenges. *Nat Rev Drug Discov* **19**, 239-252, doi:10.1038/s41573-019-0056-x (2020).
52 Wu, N. C. & Wilson, I. A. Structural Biology of Influenza Hemagglutinin: An Amaranthine Adventure. *Viruses* **12**, doi:10.3390/v12091053 (2020).
53 Nao, N. *et al.* Genetic Predisposition To Acquire a Polybasic Cleavage Site for Highly Pathogenic Avian Influenza Virus Hemagglutinin. *mBio* **8**, doi:10.1128/mBio.02298-16 (2017).
54 Bottcher-Friebertshauser, E., Klenk, H. D. & Garten, W. Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium. *Pathog Dis* **69**, 87-100, doi:10.1111/2049-632X.12053 (2013).
55 Weis, W. *et al.* Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature* **333**, 426-431, doi:10.1038/333426a0 (1988).
56 Dou, D., Revol, R., Ostbye, H., Wang, H. & Daniels, R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Front Immunol* **9**, 1581, doi:10.3389/fimmu.2018.01581 (2018).
57 Petrova, V. N. & Russell, C. A. The evolution of seasonal influenza viruses. *Nat Rev Microbiol* **16**, 60, doi:10.1038/nrmicro.2017.146 (2018).
58 Rajendran, M., Krammer, F. & McMahon, M. The Human Antibody Response to the Influenza Virus Neuraminidase Following Infection or Vaccination. *Vaccines (Basel)* **9**, doi:10.3390/vaccines9080846 (2021).

59 Isakova-Sivak, I. & Rudenko, L. The future of haemagglutinin stalk-based universal influenza vaccines. *Lancet Infect Dis*, doi:10.1016/S1473-3099(22)00056-1 (2022).
60 Muller, H. [Biology of influenza viruses]. *Berl Munch Tierarztl Wochenschr* **119**, 91-100 (2006).
61 Tang, Y. S., Xu, S., Chen, Y. W., Wang, J. H. & Shaw, P. C. Crystal structures of influenza nucleoprotein complexed with nucleic acid provide insights into the mechanism of RNA interaction. *Nucleic Acids Res* **49**, 4144-4154, doi:10.1093/nar/gkab203 (2021).
62 Dawson, W. K., Lazniewski, M. & Plewczynski, D. RNA structure interactions and ribonucleoprotein processes of the influenza A virus. *Brief Funct Genomics* **17**, 402-414, doi:10.1093/bfgp/elx028 (2018).
63 Wang, P., Palese, P. & O'Neill, R. E. The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal. *J Virol* **71**, 1850-1856, doi:10.1128/JVI.71.3.1850-1856.1997 (1997).
64 Cros, J. F., Garcia-Sastre, A. & Palese, P. An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. *Traffic* **6**, 205-213, doi:10.1111/j.1600-0854.2005.00263.x (2005).
65 Mondal, A. *et al.* Influenza virus recruits host protein kinase C to control assembly and activity of its replication machinery. *Elife* **6**, doi:10.7554/eLife.26910 (2017).
66 Hatakeyama, D. *et al.* Influenza A virus nucleoprotein is acetylated by histone acetyltransferases PCAF and GCN5. *J Biol Chem* **293**, 7126-7138, doi:10.1074/jbc.RA117.001683 (2018).
67 Gaymard, A., Le Briand, N., Frobert, E., Lina, B. & Escuret, V. Functional balance between neuraminidase and haemagglutinin in influenza viruses. *Clin Microbiol Infect* **22**, 975-983, doi:10.1016/j.cmi.2016.07.007 (2016).
68 Cohen, M. *et al.* Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virol J* **10**, 321, doi:10.1186/1743-422X-10-321 (2013).
69 Taieb, V. *et al.* Efficacy and safety of baloxavir marboxil versus neuraminidase inhibitors in the treatment of influenza virus infection in high-risk and uncomplicated patients - a Bayesian network meta-analysis. *Curr Med Res Opin* **37**, 225-244, doi:10.1080/03007995.2020.1839400 (2021).
70 McAuley, J. L., Gilbertson, B. P., Trifkovic, S., Brown, L. E. & McKimm-Breschkin, J. L. Influenza Virus Neuraminidase Structure and Functions. *Front Microbiol* **10**, 39, doi:10.3389/fmicb.2019.00039 (2019).
71 Peukes, J. *et al.* The native structure of the assembled matrix protein 1 of influenza A virus. *Nature* **587**, 495-498, doi:10.1038/s41586-020-2696-8 (2020).
72 Mohd-Kipli, F., Claridge, J. K., Habjanic, J., Jiang, A. & Schnell, J. R. Conformational triggers associated with influenza matrix protein 1 polymerization. *J Biol Chem* **296**, 100316, doi:10.1016/j.jbc.2021.100316 (2021).
73 Wu, C. Y. *et al.* Mammalian expression of virus-like particles for advanced mimicry of authentic influenza virus. *PLoS One* **5**, e9784, doi:10.1371/journal.pone.0009784 (2010).
74 Zhang, J. *et al.* Influenza A virus M1 blocks the classical complement pathway through interacting with C1qA. *J Gen Virol* **90**, 2751-2758, doi:10.1099/vir.0.014316-0 (2009).
75 Cady, S. D., Luo, W., Hu, F. & Hong, M. Structure and function of the influenza A M2 proton channel. *Biochemistry* **48**, 7356-7364, doi:10.1021/bi9008837 (2009).
76 Alvarado-Facundo, E. *et al.* Influenza virus M2 protein ion channel activity helps to maintain pandemic 2009 H1N1 virus hemagglutinin fusion competence during transport to the cell surface. *J Virol* **89**, 1975-1985, doi:10.1128/JVI.03253-14 (2015).
77 Paterson, D. & Fodor, E. Emerging roles for the influenza A virus nuclear export protein (NEP). *PLoS Pathog* **8**, e1003019, doi:10.1371/journal.ppat.1003019 (2012).

78 Ji, Z. X., Wang, X. Q. & Liu, X. F. NS1: A Key Protein in the "Game" Between Influenza A Virus and Host in Innate Immunity. *Front Cell Infect Microbiol* **11**, 670177, doi:10.3389/fcimb.2021.670177 (2021).
79 Carrillo, B. *et al.* The influenza A virus protein NS1 displays structural polymorphism. *J Virol* **88**, 4113-4122, doi:10.1128/JVI.03692-13 (2014).
80 Gack, M. U. *et al.* Influenza A virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. *Cell Host Microbe* **5**, 439-449, doi:10.1016/j.chom.2009.04.006 (2009).
81 Rajsbaum, R. *et al.* Species-specific inhibition of RIG-I ubiquitination and IFN induction by the influenza A virus NS1 protein. *PLoS Pathog* **8**, e1003059, doi:10.1371/journal.ppat.1003059 (2012).
82 Anastasina, M. *et al.* Influenza virus NS1 protein binds cellular DNA to block transcription of antiviral genes. *Biochim Biophys Acta* **1859**, 1440-1448, doi:10.1016/j.bbagrm.2016.09.005 (2016).
83 Kuo, R. L. & Krug, R. M. Influenza a virus polymerase is an integral component of the CPSF30-NS1A protein complex in infected cells. *J Virol* **83**, 1611-1616, doi:10.1128/JVI.01491-08 (2009).
84 Zhang, K. *et al.* Structural basis for influenza virus NS1 protein block of mRNA nuclear export. *Nat Microbiol* **4**, 1671-1679, doi:10.1038/s41564-019-0482-x (2019).
85 Li, S., Min, J. Y., Krug, R. M. & Sen, G. C. Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA. *Virology* **349**, 13-21, doi:10.1016/j.virol.2006.01.005 (2006).
86 Atkin-Smith, G. K., Duan, M., Chen, W. & Poon, I. K. H. The induction and consequences of Influenza A virus-induced cell death. *Cell Death Dis* **9**, 1002, doi:10.1038/s41419-018-1035-6 (2018).
87 Elton, D. *et al.* Interaction of the influenza virus nucleoprotein with the cellular CRM1-mediated nuclear export pathway. *J Virol* **75**, 408-419, doi:10.1128/JVI.75.1.408-419.2001 (2001).
88 Kordyukova, L. V. *et al.* The Cytoplasmic Tail of Influenza A Virus Hemagglutinin and Membrane Lipid Composition Change the Mode of M1 Protein Association with the Lipid Bilayer. *Membranes (Basel)* **11**, doi:10.3390/membranes11100772 (2021).
89 Noda, T. *et al.* Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging. *Nat Commun* **9**, 54, doi:10.1038/s41467-017-02517-w (2018).
90 Wang, Y. F. *et al.* Glycan-binding preferences and genetic evolution of human seasonal influenza A(H3N2) viruses during 1999-2007 in Taiwan. *PLoS One* **13**, e0196727, doi:10.1371/journal.pone.0196727 (2018).
91 Sieczkarski, S. B. & Whittaker, G. R. Influenza virus can enter and infect cells in the absence of clathrin-mediated endocytosis. *J Virol* **76**, 10455-10464, doi:10.1128/jvi.76.20.10455-10464.2002 (2002).
92 de Vries, E. *et al.* Dissection of the influenza A virus endocytic routes reveals macropinocytosis as an alternative entry pathway. *PLoS Pathog* **7**, e1001329, doi:10.1371/journal.ppat.1001329 (2011).
93 Rust, M. J., Lakadamyali, M., Zhang, F. & Zhuang, X. Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat Struct Mol Biol* **11**, 567-573, doi:10.1038/nsmb769 (2004).
94 Sieben, C., Sezgin, E., Eggeling, C. & Manley, S. Influenza A viruses use multivalent sialic acid clusters for cell binding and receptor activation. *PLoS Pathog* **16**, e1008656, doi:10.1371/journal.ppat.1008656 (2020).
95 Smith, A. M. & Perelson, A. S. Influenza A virus infection kinetics: quantitative data and models. *Wiley Interdiscip Rev Syst Biol Med* **3**, 429-445, doi:10.1002/wsbm.129 (2011).

96 Fan, H. *et al.* Structures of influenza A virus RNA polymerase offer insight into viral genome replication. *Nature* **573**, 287-290, doi:10.1038/s41586-019-1530-7 (2019).
97 Perez, J. T. *et al.* Influenza A virus-generated small RNAs regulate the switch from transcription to replication. *Proc Natl Acad Sci U S A* **107**, 11525-11530, doi:10.1073/pnas.1001984107 (2010).
98 Li, X., Gu, M., Zheng, Q., Gao, R. & Liu, X. Packaging signal of influenza A virus. *Virol J* **18**, 36, doi:10.1186/s12985-021-01504-4 (2021).
99 Chou, Y. Y. *et al.* One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proc Natl Acad Sci U S A* **109**, 9101-9106, doi:10.1073/pnas.1206069109 (2012).
100 Nakatsu, S. *et al.* Influenza C and D Viruses Package Eight Organized Ribonucleoprotein Complexes. *J Virol* **92**, doi:10.1128/JVI.02084-17 (2018).
101 Le Page, C., Genin, P., Baines, M. G. & Hiscott, J. Interferon activation and innate immunity. *Rev Immunogenet* **2**, 374-386 (2000).
102 Chen, K. R. *et al.* TBK1-associated protein in endolysosomes (TAPE)/CC2D1A is a key regulator linking RIG-I-like receptors to antiviral immunity. *J Biol Chem* **287**, 32216-32221, doi:10.1074/jbc.C112.394346 (2012).
103 Iwasaki, A. & Pillai, P. S. Innate immunity to influenza virus infection. *Nat Rev Immunol* **14**, 315-328, doi:10.1038/nri3665 (2014).
104 Padilla-Quirarte, H. O., Lopez-Guerrero, D. V., Gutierrez-Xicotencatl, L. & Esquivel-Guadarrama, F. Protective Antibodies Against Influenza Proteins. *Front Immunol* **10**, 1677, doi:10.3389/fimmu.2019.01677 (2019).
105 Thompson, A. J. & Paulson, J. C. Adaptation of influenza viruses to human airway receptors. *J Biol Chem* **296**, 100017, doi:10.1074/jbc.REV120.013309 (2021).
106 Johnson, N. P. & Mueller, J. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bull Hist Med* **76**, 105-115, doi:10.1353/bhm.2002.0022 (2002).
107 Lederberg, J. H1N1-influenza as Lazarus: genomic resurrection from the tomb of an unknown. *Proc Natl Acad Sci U S A* **98**, 2115-2116, doi:10.1073/pnas.051000798 (2001).
108 Shao, W., Li, X., Goraya, M. U., Wang, S. & Chen, J. L. Evolution of Influenza A Virus by Mutation and Re-Assortment. *Int J Mol Sci* **18**, doi:10.3390/ijms18081650 (2017).
109 Saunders-Hastings, P. R. & Krewski, D. Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission. *Pathogens* **5**, doi:10.3390/pathogens5040066 (2016).
110 Honigsbaum, M. Revisiting the 1957 and 1968 influenza pandemics. *Lancet* **395**, 1824-1826, doi:10.1016/S0140-6736(20)31201-0 (2020).
111 Rozo, M. & Gronvall, G. K. The Reemergent 1977 H1N1 Strain and the Gain-of-Function Debate. *mBio* **6**, doi:10.1128/mBio.01013-15 (2015).
112 Guo, F. *et al.* Origin and Evolution of H1N1/pdm2009: A Codon Usage Perspective. *Front Microbiol* **11**, 1615, doi:10.3389/fmicb.2020.01615 (2020).
113 Iuliano, A. D. *et al.* Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet* **391**, 1285-1300, doi:10.1016/S0140-6736(17)33293-2 (2018).
114 Thommes, E. W., Kruse, M., Kohli, M., Sharma, R. & Noorduyn, S. G. Review of seasonal influenza in Canada: Burden of disease and the cost-effectiveness of quadrivalent inactivated influenza vaccines. *Hum Vaccin Immunother* **13**, 867-876, doi:10.1080/21645515.2016.1251537 (2017).
115 Dao, C. N. *et al.* Adult hospitalizations for laboratory-positive influenza during the 2005-2006 through 2007-2008 seasons in the United States. *J Infect Dis* **202**, 881-888, doi:10.1086/655904 (2010).

116 Kwong, J. C., Schwartz, K. L. & Campitelli, M. A. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. *N Engl J Med* **378**, 2540-2541, doi:10.1056/NEJMc1805679 (2018).
117 Modin, D. *et al.* Influenza Vaccine in Heart Failure. *Circulation* **139**, 575-586, doi:10.1161/CIRCULATIONAHA.118.036788 (2019).
118 Mohseni, H., Kiran, A., Khorshidi, R. & Rahimi, K. Influenza vaccination and risk of hospitalization in patients with heart failure: a self-controlled case series study. *Eur Heart J* **38**, 326-333, doi:10.1093/eurheartj/ehw411 (2017).
119 Macias, A. E. *et al.* The disease burden of influenza beyond respiratory illness. *Vaccine* **39 Suppl 1**, A6-A14, doi:10.1016/j.vaccine.2020.09.048 (2021).
120 Putri, W., Muscatello, D. J., Stockwell, M. S. & Newall, A. T. Economic burden of seasonal influenza in the United States. *Vaccine* **36**, 3960-3966, doi:10.1016/j.vaccine.2018.05.057 (2018).
121 Ng, C. *et al.* Resource utilization and cost of influenza requiring hospitalization in Canadian adults: A study from the serious outcomes surveillance network of the Canadian Immunization Research Network. *Influenza Other Respir Viruses* **12**, 232-240, doi:10.1111/irv.12521 (2018).
122 Rolfes, M. A. *et al.* Effects of Influenza Vaccination in the United States During the 2017-2018 Influenza Season. *Clin Infect Dis* **69**, 1845-1853, doi:10.1093/cid/ciz075 (2019).
123 Chen, J., Wang, J., Zhang, J. & Ly, H. Advances in Development and Application of Influenza Vaccines. *Front Immunol* **12**, 711997, doi:10.3389/fimmu.2021.711997 (2021).
124 Becker, T., Elbahesh, H., Reperant, L. A., Rimmelzwaan, G. F. & Osterhaus, A. Influenza Vaccines: Successes and Continuing Challenges. *J Infect Dis* **224**, S405-S419, doi:10.1093/infdis/jiab269 (2021).
125 Cox, M. M. & Hollister, J. R. FluBlok, a next generation influenza vaccine manufactured in insect cells. *Biologicals* **37**, 182-189, doi:10.1016/j.biologicals.2009.02.014 (2009).
126 Blanco-Lobo, P., Nogales, A., Rodriguez, L. & Martinez-Sobrido, L. Novel Approaches for The Development of Live Attenuated Influenza Vaccines. *Viruses* **11**, doi:10.3390/v11020190 (2019).
127 Fiore, A. E., Bridges, C. B. & Cox, N. J. Seasonal influenza vaccines. *Curr Top Microbiol Immunol* **333**, 43-82, doi:10.1007/978-3-540-92165-3_3 (2009).
128 Nafziger, A. N. & Pratt, D. S. Seasonal influenza vaccination and technologies. *J Clin Pharmacol* **54**, 719-731, doi:10.1002/jcph.299 (2014).
129 Milian, E. & Kamen, A. A. Current and emerging cell culture manufacturing technologies for influenza vaccines. *Biomed Res Int* **2015**, 504831, doi:10.1155/2015/504831 (2015).
130 Liu, B., Hossain, M. J., Mori, I. & Kimura, Y. Evaluation of a virus derived from MDCK cells infected persistently with influenza A virus as a potential live-attenuated vaccine candidate in the mouse model. *J Med Virol* **80**, 888-894, doi:10.1002/jmv.21148 (2008).
131 Barrett, P. N., Portsmouth, D. & Ehrlich, H. J. Vero cell culture-derived pandemic influenza vaccines: preclinical and clinical development. *Expert Rev Vaccines* **12**, 395-413, doi:10.1586/erv.13.21 (2013).
132 Huang, D. *et al.* Serum-Free Suspension Culture of MDCK Cells for Production of Influenza H1N1 Vaccines. *PLoS One* **10**, e0141686, doi:10.1371/journal.pone.0141686 (2015).
133 Skowronski, D. M. *et al.* Low 2012-13 influenza vaccine effectiveness associated with mutation in the egg-adapted H3N2 vaccine strain not antigenic drift in circulating viruses. *PLoS One* **9**, e92153, doi:10.1371/journal.pone.0092153 (2014).
134 Zost, S. J. *et al.* Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proc Natl Acad Sci U S A* **114**, 12578-12583, doi:10.1073/pnas.1712377114 (2017).

135 Izurieta, H. S. *et al.* Relative effectiveness of cell-cultured and egg-based influenza vaccines among the U.S. elderly, 2017-18. *J Infect Dis*, doi:10.1093/infdis/jiy716 (2018).
136 Agor, J. K. & Ozaltin, O. Y. Models for predicting the evolution of influenza to inform vaccine strain selection. *Hum Vaccin Immunother* **14**, 678-683, doi:10.1080/21645515.2017.1423152 (2018).
137 Zimmerman, R. K. *et al.* 2014-2015 Influenza Vaccine Effectiveness in the United States by Vaccine Type. *Clin Infect Dis* **63**, 1564-1573, doi:10.1093/cid/ciw635 (2016).
138 Bui, C. M., Chughtai, A. A., Adam, D. C. & MacIntyre, C. R. An overview of the epidemiology and emergence of influenza A infection in humans over time. *Arch Public Health* **75**, 15, doi:10.1186/s13690-017-0182-z (2017).
139 Moghadami, M. A Narrative Review of Influenza: A Seasonal and Pandemic Disease. *Iran J Med Sci* **42**, 2-13 (2017).
140 Singh, N., Pandey, A. & Mittal, S. K. Avian influenza pandemic preparedness: developing prepandemic and pandemic vaccines against a moving target. *Expert Rev Mol Med* **12**, e14, doi:10.1017/S1462399410001432 (2010).
141 Le Ru, A. *et al.* Scalable production of influenza virus in HEK-293 cells for efficient vaccine manufacturing. *Vaccine* **28**, 3661-3671, doi:10.1016/j.vaccine.2010.03.029 (2010).
142 Ping, J. *et al.* Development of high-yield influenza A virus vaccine viruses. *Nat Commun* **6**, 8148, doi:10.1038/ncomms9148 (2015).
143 Henry, O., Dormond, E., Perrier, M. & Kamen, A. Insights into adenoviral vector production kinetics in acoustic filter-based perfusion cultures. *Biotechnol Bioeng* **86**, 765-774, doi:10.1002/bit.20074 (2004).
144 Wlaschin, K. F. & Hu, W. S. Fedbatch culture and dynamic nutrient feeding. *Adv Biochem Eng Biotechnol* **101**, 43-74, doi:10.1007/10_015 (2006).
145 Xu, S., Gavin, J., Jiang, R. & Chen, H. Bioreactor productivity and media cost comparison for different intensified cell culture processes. *Biotechnol Prog* **33**, 867-878, doi:10.1002/btpr.2415 (2017).
146 Petiot, E. *et al.* Metabolic and kinetic analyses of influenza production in perfusion HEK293 cell culture. *BMC Biotechnol* **11**, 84, doi:10.1186/1472-6750-11-84 (2011).
147 Merten, O. W. 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)* **111**, 233-257 (2002).
148 Aggarwal, K., Jing, F., Maranga, L. & Liu, J. Bioprocess optimization for cell culture based influenza vaccine production. *Vaccine* **29**, 3320-3328, doi:10.1016/j.vaccine.2011.01.081 (2011).
149 Puente-Massaguer, E., Badiella, L., Gutierrez-Granados, S., Cervera, L. & Godia, F. A statistical approach to improve compound screening in cell culture media. *Eng Life Sci* **19**, 315-327, doi:10.1002/elsc.201800168 (2019).
150 Smith, R. H. Adeno-associated virus integration: virus versus vector. *Gene Ther* **15**, 817-822, doi:10.1038/gt.2008.55 (2008).
151 Summerford, C. & Samulski, R. J. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J Virol* **72**, 1438-1445 (1998).
152 Pillay, S. *et al.* An essential receptor for adeno-associated virus infection. *Nature* **530**, 108-112, doi:10.1038/nature16465 (2016).
153 Nicolson, S. C. & Samulski, R. J. Recombinant adeno-associated virus utilizes host cell nuclear import machinery to enter the nucleus. *J Virol* **88**, 4132-4144, doi:10.1128/JVI.02660-13 (2014).
154 Ferrari, F. K., Samulski, T., Shenk, T. & Samulski, R. J. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J Virol* **70**, 3227-3234 (1996).

155 Thomas, C. E., Storm, T. A., Huang, Z. & Kay, M. A. Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors. *J Virol* **78**, 3110-3122 (2004).
156 Linden, R. M., Ward, P., Giraud, C., Winocour, E. & Berns, K. I. Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci U S A* **93**, 11288-11294 (1996).
157 Kuzmin, D. A. *et al.* The clinical landscape for AAV gene therapies. *Nat Rev Drug Discov* **20**, 173-174, doi:10.1038/d41573-021-00017-7 (2021).
158 Carter, P. J. & Samulski, R. J. Adeno-associated viral vectors as gene delivery vehicles. *Int J Mol Med* **6**, 17-27 (2000).
159 Afione, S. A. *et al.* In vivo model of adeno-associated virus vector persistence and rescue. *J Virol* **70**, 3235-3241 (1996).
160 Emmerling, V. V. *et al.* Rational plasmid design and bioprocess optimization to enhance recombinant adeno-associated virus (AAV) productivity in mammalian cells. *Biotechnol J* **11**, 290-297, doi:10.1002/biot.201500176 (2016).
161 Thorne, B. A., Takeya, R. K. & Peluso, R. W. Manufacturing recombinant adeno-associated viral vectors from producer cell clones. *Hum Gene Ther* **20**, 707-714, doi:10.1089/hum.2009.070 (2009).
162 Timpe, J. M., Verrill, K. C. & Trempe, J. P. Effects of adeno-associated virus on adenovirus replication and gene expression during coinfection. *J Virol* **80**, 7807-7815, doi:10.1128/JVI.00198-06 (2006).
163 Felberbaum, R. S. The baculovirus expression vector system: A commercial manufacturing platform for viral vaccines and gene therapy vectors. *Biotechnol J* **10**, 702-714, doi:10.1002/biot.201400438 (2015).
164 Virag, T., Cecchini, S. & Kotin, R. M. Producing recombinant adeno-associated virus in foster cells: overcoming production limitations using a baculovirus-insect cell expression strategy. *Hum Gene Ther* **20**, 807-817, doi:10.1089/hum.2009.092 (2009).
165 Cecchini, S., Virag, T. & Kotin, R. M. Reproducible high yields of recombinant adeno-associated virus produced using invertebrate cells in 0.02- to 200-liter cultures. *Hum Gene Ther* **22**, 1021-1030, doi:10.1089/hum.2010.250 (2011).
166 Mietzsch, M. *et al.* OneBac: platform for scalable and high-titer production of adeno-associated virus serotype 1-12 vectors for gene therapy. *Hum Gene Ther* **25**, 212-222, doi:10.1089/hum.2013.184 (2014).
167 Ferreira, V., Petry, H. & Salmon, F. Immune Responses to AAV-Vectors, the Glybera Example from Bench to Bedside. *Front Immunol* **5**, 82, doi:10.3389/fimmu.2014.00082 (2014).
168 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
169 Xu, Y. & Li, Z. CRISPR-Cas systems: Overview, innovations and applications in human disease research and gene therapy. *Comput Struct Biotechnol J* **18**, 2401-2415, doi:10.1016/j.csbj.2020.08.031 (2020).
170 Goldberg, G. W. *et al.* Engineered dual selection for directed evolution of SpCas9 PAM specificity. *Nat Commun* **12**, 349, doi:10.1038/s41467-020-20650-x (2021).
171 Asmamaw, M. & Zawdie, B. Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing. *Biologics* **15**, 353-361, doi:10.2147/BTT.S326422 (2021).
172 Chakrabarti, A. M. *et al.* Target-Specific Precision of CRISPR-Mediated Genome Editing. *Mol Cell* **73**, 699-713 e696, doi:10.1016/j.molcel.2018.11.031 (2019).
173 Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766, doi:10.7554/eLife.04766 (2014).

174 Zhang, L. *et al.* Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLoS One* **10**, e0120396, doi:10.1371/journal.pone.0120396 (2015).
175 Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol* **34**, 631-633, doi:10.1038/nbt.3536 (2016).
176 Wright, J. B. & Sanjana, N. E. CRISPR Screens to Discover Functional Noncoding Elements. *Trends Genet* **32**, 526-529, doi:10.1016/j.tig.2016.06.004 (2016).
177 Burgess, D. J. Functional genomics: Shining a light on genetic screen strategies. *Nat Rev Genet* **19**, 6-7, doi:10.1038/nrg.2017.99 (2018).
178 Joana R. Costa, B. E. B., James E. McGee, Adam I. Fogel, Kyle R. Brimacombe,Robin Ketteler. *Genome Editing Using Engineered Nucleases and Their Use in Genomic Screening*, <https://www.ncbi.nlm.nih.gov/books/NBK464635/> (2017).
179 Miles, L. A., Garippa, R. J. & Poirier, J. T. Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens. *FEBS J* **283**, 3170-3180, doi:10.1111/febs.13770 (2016).
180 Yip, B. H. Recent Advances in CRISPR/Cas9 Delivery Strategies. *Biomolecules* **10**, doi:10.3390/biom10060839 (2020).
181 Polstein, L. R. & Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat Chem Biol* **11**, 198-200, doi:10.1038/nchembio.1753 (2015).
182 Gao, Y. *et al.* Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat Methods* **13**, 1043-1049, doi:10.1038/nmeth.4042 (2016).
183 Mou, H. *et al.* CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol* **18**, 108, doi:10.1186/s13059-017-1237-8 (2017).
184 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157, doi:10.1038/s41586-019-1711-4 (2019).
185 Stepper, P. *et al.* Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res* **45**, 1703-1713, doi:10.1093/nar/gkw1112 (2017).
186 Heigwer, F. *et al.* CRISPR library designer (CLD): software for multispecies design of single guide RNA libraries. *Genome Biol* **17**, 55, doi:10.1186/s13059-016-0915-2 (2016).
187 Aguirre, A. J. *et al.* Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov* **6**, 914-929, doi:10.1158/2159-8290.CD-16-0154 (2016).
188 Sack, L. M., Davoli, T., Xu, Q., Li, M. Z. & Elledge, S. J. Sources of Error in Mammalian Genetic Screens. *G3 (Bethesda)* **6**, 2781-2790, doi:10.1534/g3.116.030973 (2016).
189 Nagy, T. & Kampmann, M. CRISPulator: a discrete simulation tool for pooled genetic screens. *BMC Bioinformatics* **18**, 347, doi:10.1186/s12859-017-1759-9 (2017).
190 Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515-1526, doi:10.1016/j.cell.2015.11.015 (2015).
191 Kim, E. & Hart, T. Improved analysis of CRISPR fitness screens and reduced off-target effects with the BAGEL2 gene essentiality classifier. *Genome Med* **13**, 2, doi:10.1186/s13073-020-00809-3 (2021).
192 Thompson, N. A. *et al.* Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nat Commun* **12**, 1302, doi:10.1038/s41467-021-21478-9 (2021).
193 Genolet, O., Ravid Lustig, L. & Schulz, E. G. 1-24 (Springer US).
194 Garcia-Beltran, W. F. *et al.* Open conformers of HLA-F are high-affinity ligands of the activating NK-cell receptor KIR3DS1. *Nat Immunol* **17**, 1067-1074, doi:10.1038/ni.3513 (2016).
195 Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896 e1815, doi:10.1016/j.cell.2016.11.039 (2016).
196 Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799 e717, doi:10.1016/j.cell.2019.09.016 (2019).

197 Heaton, B. E. *et al.* A CRISPR Activation Screen Identifies a Pan-avian Influenza Virus Inhibitory Host Factor. *Cell Rep* **20**, 1503-1512, doi:10.1016/j.celrep.2017.07.060 (2017).
198 Tripathi, S. *et al.* Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe* **18**, 723-735, doi:10.1016/j.chom.2015.11.002 (2015).
199 Wu, W., Orr-Burks, N., Karpilow, J. & Tripp, R. A. Development of improved vaccine cell lines against rotavirus. *Sci Data* **4**, 170021, doi:10.1038/sdata.2017.21 (2017).
200 van der Sanden, S. M. *et al.* Engineering Enhanced Vaccine Cell Lines To Eradicate Vaccine-Preventable Diseases: the Polio End Game. *J Virol* **90**, 1694-1704, doi:10.1128/JVI.01464-15 (2016).
201 Hoeksema, F. *et al.* Enhancing viral vaccine production using engineered knockout vero cell lines - A second look. *Vaccine* **36**, 2093-2103, doi:10.1016/j.vaccine.2018.03.010 (2018).
202 Barnes, C. R. *et al.* Genome-wide activation screens to increase adeno-associated virus production. *Mol Ther Nucleic Acids* **26**, 94-103, doi:10.1016/j.omtn.2021.06.026 (2021).
203 Cote, J., Garnier, A., Massie, B. & Kamen, A. Serum-free production of recombinant proteins and adenoviral vectors by 293SF-3F6 cells. *Biotechnol Bioeng* **59**, 567-575 (1998).
204 Ansorge, S. *et al.* Development of a scalable process for high-yield lentiviral vector production by transient transfection of HEK293 suspension cultures. *J Gene Med* **11**, 868-876, doi:10.1002/jgm.1370 (2009).
205 Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**, 493-501, doi:10.1261/rna.2192803 (2003).
206 Barczak, W., Suchorska, W., Rubis, B. & Kulcenty, K. Universal real-time PCR-based assay for lentiviral titration. *Mol Biotechnol* **57**, 195-200, doi:10.1007/s12033-014-9815-4 (2015).
207 Neumann, G., Fujii, K., Kino, Y. & Kawaoka, Y. An improved reverse genetics system for influenza A virus generation and its implications for vaccine production. *Proc Natl Acad Sci U S A* **102**, 16825-16829, doi:10.1073/pnas.0505587102 (2005).
208 Powell, T. J., Silk, J. D., Sharps, J., Fodor, E. & Townsend, A. R. Pseudotyped influenza A virus as a vaccine for the induction of heterotypic immunity. *J Virol* **86**, 13397-13406, doi:10.1128/JVI.01820-12 (2012).
209 Xiong, W., MacColl Garfinkel, A. E., Li, Y., Benowitz, L. I. & Cepko, C. L. NRF2 promotes neuronal survival in neurodegeneration and acute nerve damage. *J Clin Invest* **125**, 1433-1445, doi:10.1172/JCI79735 (2015).
210 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).
211 Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168, doi:10.1093/nar/gku936 (2014).
212 Kerviel, A. *et al.* Involvement of an Arginine Triplet in M1 Matrix Protein Interaction with Membranes and in M1 Recruitment into Virus-Like Particles of the Influenza A(H1N1)pdm09 Virus. *PLoS One* **11**, e0165421, doi:10.1371/journal.pone.0165421 (2016).
213 Reich, N. C., Sarnow, P., Duprey, E. & Levine, A. J. Monoclonal antibodies which recognize native and denatured forms of the adenovirus DNA-binding protein. *Virology* **128**, 480-484, doi:10.1016/0042-6822(83)90274-x (1983).
214 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
215 Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554, doi:10.1186/s13059-014-0554-4 (2014).
216 Li, W. *et al.* Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* **16**, 281, doi:10.1186/s13059-015-0843-6 (2015).

217 Wang, B. *et al.* Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat Protoc* **14**, 756-780, doi:10.1038/s41596-018-0113-7 (2019).

218 Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573-580, doi:10.1093/bioinformatics/btr709 (2012).

219 Connelly, J. P. & Pruett-Miller, S. M. CRIS.py: A Versatile and High-throughput Analysis Program for CRISPR-based Genome Editing. *Sci Rep* **9**, 4194, doi:10.1038/s41598-019-40896-w (2019).

220 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).

221 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).

222 Vinayagam, A. *et al.* Protein complex-based analysis framework for high-throughput data sets. *Sci Signal* **6**, rs5, doi:10.1126/scisignal.2003629 (2013).

223 Lin, Y. C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun* **5**, 4767, doi:10.1038/ncomms5767 (2014).

224 Venereo-Sanchez, A. *et al.* Characterization of influenza H1N1 Gag virus-like particles and extracellular vesicles co-produced in HEK-293SF. *Vaccine*, doi:10.1016/j.vaccine.2019.07.057 (2019).

225 Shen, C. F. *et al.* Optimization and scale-up of cell culture and purification processes for production of an adenovirus-vectored tuberculosis vaccine candidate. *Vaccine* **34**, 3381-3387, doi:10.1016/j.vaccine.2016.04.090 (2016).

226 Benton, D. J. *et al.* Influenza hemagglutinin membrane anchor. *Proc Natl Acad Sci U S A* **115**, 10112-10117, doi:10.1073/pnas.1810927115 (2018).

227 Wu, X. *et al.* Inhibition of Influenza A Virus Replication by TRIM14 via Its Multifaceted Protein-Protein Interaction With NP. *Front Microbiol* **10**, 344, doi:10.3389/fmicb.2019.00344 (2019).

228 Hart, T., Brown, K. R., Sircoulomb, F., Rottapel, R. & Moffat, J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol Syst Biol* **10**, 733, doi:10.15252/msb.20145216 (2014).

229 Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med* **24**, 927-930, doi:10.1038/s41591-018-0049-z (2018).

230 Yoshida, S. *et al.* Redox regulates mammalian target of rapamycin complex 1 (mTORC1) activity by modulating the TSC1/TSC2-Rheb GTPase pathway. *J Biol Chem* **286**, 32651-32660, doi:10.1074/jbc.M111.238014 (2011).

231 Kamen, A. & Henry, O. Development and optimization of an adenovirus production process. *J Gene Med* **6 Suppl 1**, S184-192, doi:10.1002/jgm.503 (2004).

232 Chen, A. *et al.* Serum-free microcarrier based production of replication deficient influenza vaccine candidate virus lacking NS1 using Vero cells. *BMC Biotechnol* **11**, 81, doi:10.1186/1472-6750-11-81 (2011).

233 Dill, V., Ehret, J., Zimmer, A., Beer, M. & Eschbaumer, M. Cell Density Effects in Different Cell Culture Media and Their Impact on the Propagation of Foot-And-Mouth Disease Virus. *Viruses* **11**, doi:10.3390/v11060511 (2019).

234 Lin, H., Leighty, R. W., Godfrey, S. & Wang, S. B. Principles and approach to developing mammalian cell culture media for high cell density perfusion process leveraging established fed-batch media. *Biotechnol Prog* **33**, 891-901, doi:10.1002/btpr.2472 (2017).

235 Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**, 224-226, doi:10.1038/s41587-019-0032-3 (2019).

236 Killian, M. L. Hemagglutination assay for influenza virus. *Methods Mol Biol* **1161**, 3-9, doi:10.1007/978-1-4939-0758-8_1 (2014).

237 Nakaya, Y., Fukuda, T., Ashiba, H., Yasuura, M. & Fujimaki, M. Quick assessment of influenza a virus infectivity with a long-range reverse-transcription quantitative polymerase chain reaction assay. *BMC Infect Dis* **20**, 585, doi:10.1186/s12879-020-05317-8 (2020).
238 Chahal, P. S., Schulze, E., Tran, R., Montes, J. & Kamen, A. A. Production of adeno-associated virus (AAV) serotypes by transient transfection of HEK293 cell suspension cultures for gene delivery. *J Virol Methods* **196**, 163-173, doi:10.1016/j.jviromet.2013.10.038 (2014).
239 Zeltner, N., Kohlbrenner, E., Clement, N., Weber, T. & Linden, R. M. Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. *Gene Ther* **17**, 872-879, doi:10.1038/gt.2010.27 (2010).
240 Wobus, C. E. *et al.* Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and identification of capsid domains involved in AAV-2-cell interaction and neutralization of AAV-2 infection. *J Virol* **74**, 9281-9293, doi:10.1128/jvi.74.19.9281-9293.2000 (2000).
241 Xiao, W., Warrington, K. H., Jr., Hearing, P., Hughes, J. & Muzyczka, N. Adenovirus-facilitated nuclear translocation of adeno-associated virus type 2. *J Virol* **76**, 11505-11517, doi:10.1128/jvi.76.22.11505-11517.2002 (2002).
242 Mair, B. *et al.* Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Rep* **27**, 599-615 e512, doi:10.1016/j.celrep.2019.02.041 (2019).
243 Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 5416, doi:10.1038/s41467-018-07901-8 (2018).
244 Nguyen, T. N. T. *et al.* Mechanistic model for production of recombinant adeno-associated virus via triple transfection of HEK293 cells. *Mol Ther Methods Clin Dev* **21**, 642-655, doi:10.1016/j.omtm.2021.04.006 (2021).
245 Fus-Kujawa, A. *et al.* An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Front Bioeng Biotechnol* **9**, 701031, doi:10.3389/fbioe.2021.701031 (2021).
246 Erbacher, P. *et al.* Genuine DNA/polyethylenimine (PEI) complexes improve transfection properties and cell survival. *J Drug Target* **12**, 223-236, doi:10.1080/10611860410001723487 (2004).
247 Cardarelli, F. *et al.* The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery. *Sci Rep* **6**, 25879, doi:10.1038/srep25879 (2016).
248 Satterly, N. *et al.* Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proc Natl Acad Sci U S A* **104**, 1853-1858, doi:10.1073/pnas.0610977104 (2007).
249 Huang, J. & Manning, B. D. The TSC1-TSC2 complex: a molecular switchboard controlling cell growth. *Biochem J* **412**, 179-190, doi:10.1042/BJ20080281 (2008).
250 Kuss-Duerkop, S. K. *et al.* Influenza virus differentially activates mTORC1 and mTORC2 signaling to maximize late stage replication. *PLoS Pathog* **13**, e1006635, doi:10.1371/journal.ppat.1006635 (2017).
251 Hopkins, K. C. *et al.* A genome-wide RNAi screen reveals that mRNA decapping restricts bunyaviral replication by limiting the pools of Dcp2-accessible targets for cap-snatching. *Genes Dev* **27**, 1511-1525, doi:10.1101/gad.215384.113 (2013).
252 Choksupmanee, O. *et al.* Specific Interaction of DDX6 with an RNA Hairpin in the 3' UTR of the Dengue Virus Genome Mediates G(1) Phase Arrest. *J Virol* **95**, e0051021, doi:10.1128/JVI.00510-21 (2021).
253 Nunez, R. D. *et al.* The RNA Helicase DDX6 Associates with RIG-I to Augment Induction of Antiviral Signaling. *Int J Mol Sci* **19**, doi:10.3390/ijms19071877 (2018).

254 Fernandez, I. S. *et al.* Characterization of SMG-9, an essential component of the nonsense-mediated mRNA decay SMG1C complex. *Nucleic Acids Res* **39**, 347-358, doi:10.1093/nar/gkq749 (2011).
255 May, J. P., Yuan, X., Sawicki, E. & Simon, A. E. RNA virus evasion of nonsense-mediated decay. *PLoS Pathog* **14**, e1007459, doi:10.1371/journal.ppat.1007459 (2018).
256 Zhang, Z. *et al.* Crosstalk between histone modifications indicates that inhibition of arginine methyltransferase CARM1 activity reverses HIV latency. *Nucleic Acids Res* **45**, 9348-9360, doi:10.1093/nar/gkx550 (2017).
257 Osada, N. *et al.* The genome landscape of the african green monkey kidney-derived vero cell line. *DNA Res* **21**, 673-683, doi:10.1093/dnares/dsu029 (2014).
258 Murray, N. & McMichael, A. Antigen presentation in virus infection. *Curr Opin Immunol* **4**, 401-407, doi:10.1016/s0952-7915(06)80030-0 (1992).
259 Karlas, A. *et al.* Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* **463**, 818-822, doi:10.1038/nature08760 (2010).
260 Li, B. *et al.* Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nat Commun* **11**, 164, doi:10.1038/s41467-019-13965-x (2020).
261 Egloff, P. *et al.* Engineered peptide barcodes for in-depth analyses of binding protein libraries. *Nat Methods* **16**, 421-428, doi:10.1038/s41592-019-0389-8 (2019).
262 Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol* **34**, 634-636, doi:10.1038/nbt.3567 (2016).
263 Shokralla, S. *et al.* Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci Rep* **5**, 9687, doi:10.1038/srep09687 (2015).
264 Kalmbach, A. *et al.* Experimental characterization of flow conditions in 2- and 20-L bioreactors with wave-induced motion. *Biotechnol Prog* **27**, 402-409, doi:10.1002/btpr.516 (2011).
265 Heinz, F. X. & Stiasny, K. Distinguishing features of current COVID-19 vaccines: knowns and unknowns of antigen presentation and modes of action. *NPJ Vaccines* **6**, 104, doi:10.1038/s41541-021-00369-6 (2021).
266 Goel, R. R. *et al.* mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern. *Science* **374**, abm0829, doi:10.1126/science.abm0829 (2021).
267 Mendonca, S. A., Lorincz, R., Boucher, P. & Curiel, D. T. Adenoviral vector vaccine platforms in the SARS-CoV-2 pandemic. *NPJ Vaccines* **6**, 97, doi:10.1038/s41541-021-00356-x (2021).
268 Uddin, M. N. & Roni, M. A. Challenges of Storage and Stability of mRNA-Based COVID-19 Vaccines. *Vaccines (Basel)* **9**, doi:10.3390/vaccines9091033 (2021).
269 Shartouny, J. R. & Lowen, A. C. Message in a bottle: mRNA vaccination for influenza. *J Gen Virol* **103**, doi:10.1099/jgv.0.001765 (2022).
270 Zhao, H. *et al.* Creation of a High-Yield AAV Vector Production Platform in Suspension Cells Using a Design-of-Experiment Approach. *Mol Ther Methods Clin Dev* **18**, 312-320, doi:10.1016/j.omtm.2020.06.004 (2020).

Thesis title: Pooled Genome-wide CRISPR/Cas9 screening as a tool to intensify the yield of cell-based influenza vaccines and other viral biologics

Appendix 1: Primers and oligos

ddPCR assays

For measurement of total influenza and AAV viral particles and in CFS

Influenza A segment 7, forward	GACCRATCCTGTCACCTCTGAC
Influenza A segment 7, reverse	AGGGCATTYTGGACAAAKCGTCTA
AAV2 ITR, forward	GGAACCCCTAGTGATGGAGTT
AAV2 ITR, forward	CGGCCTCAGTGAGCGA

1) 95°C, 10min 2) 95°C, 30s 3) 61°C, 30s 4) 72°C, 30s 5) Back to step 2, 39 cycles 6) 72°C, 10min 7) 4°C, hold

For measurement of integrated lentiviral vector copy number

WPRE, forward	GTCCTTTCCATGGCTGCTC
WPRE, reverse	CCGAAGGGACGTAGCAGA
Albumin, forward	TTTGCAGATGTCAGTGAAAGAGA
Albumin, reverse	TGGGGAGGCTATAGAAAATAAGG

1) 95°C, 10min 2) 95°C, 30s 3) 60°C, 30s 4) Back to step 2, 39 cycles 5) 72°C, 10min 6) 4°C, hold

For measurement of influenza segment 4 mRNA (GFP and HA)

GFP, forward	CTGCTGCCCGACAACCAC
GFP, reverse*	TCACGAACTCCAGCAGGAC
HA, forward	ATCGACTATGAGGAGCTGAGGG
HA, reverse*	GCCGTTACTCCGTTTGTGTTGT
Actin, forward	GTCATACTCCTGCTTGCTGAT
Actin, Reverse	AAAGACCTGTACGCCAACAC

*Reverse transcription was carried out using these primers as gene-specific primers to selectively amplify positive-sense RNA

1) 95°C, 10min 2) 95°C, 30s 3) 60°C, 30s 4) Back to step 2, 39 cycles 5) 72°C, 10min 6) 4°C, hold

Thesis title: Pooled Genome-wide CRISPR/Cas9 screening as a tool to intensify the yield of cell-based influenza vaccines and other viral biologics

Knockout pools

TBK1

sgRNA sequence	TTCCGCGGCCACGGTAATGA
TIDE assay primer, forward	GGCCGTTTTCCAAAATACCGA
TIDE assay primer, reverse	GATGCAGGTCGAGGACCG

1) 95°C, 60s 2) 95°C, 15s 3) 63°C, 15s 4) Back to step 2, 28 cycles 5) 72°C, 5min 6) 4°C, hold

DDX6

sgRNA sequence	AGGTCTAGCCGTTCAAGTAA
TIDE assay primer, forward	TGTTGCAGGGATGAGGTGTC
TIDE assay primer, reverse	CCTGTCTCACTGGAATGCTGT

1) 98°C, 3min 2) 98°C, 30s 3) 63°C, 30s 4) 72°C, 60s 5) Back to step 2, 34 cycles 6) 72°C, 5min 7) 4°C, hold

SMG9

sgRNA sequence	GCTGAAATGAAGGAACGAGG
TIDE assay primer, forward	TCAAAACATGCACTACCCCC
TIDE assay primer, reverse	CCAGTCAGTGCTAACGACAGT

1) 98°C, 3min 2) 98°C, 30s 3) 63°C, 30s 4) 72°C, 60s 5) Back to step 2, 34 cycles 6) 72°C, 5min 7) 4°C, hold

CARM1

sgRNA sequence	TCGCGTCGCCGATGGTGAGG
TIDE assay primer, forward	TTGTGTGGGGCGGGGTA
TIDE assay primer, reverse	GCTCCCTTGCTCACTCTGG

1) 98°C, 3min 2) 98°C, 30s 3) 63°C, 30s 4) 72°C, 60s 5) Back to step 2, 34 cycles 6) 72°C, 5min 7) 4°C, hold *Use high GC buffer

Non-targeting control (NTC)

sgRNA sequence	TTCCGCGGCCACGGTAATGA

Thesis title: Pooled Genome-wide CRISPR/Cas9 screening as a tool to intensify the yield of cell-based influenza vaccines and other viral biologics

Illumina HiSeq amplicon library prep/barcoding

P5_0nt_stagger*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTTTGTGGAAAGGACGAAACACCG
P5_1nt_stagger*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTCTTGTGGAAAGGACGAAACACCG
P5_2nt_stagger*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTGCTTGTGGAAAGGACGAAACACCG
P5_3nt_stagger*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTAGCTTGTGGAAAGGACGAAACACCG
P5_4nt_stagger*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTCAACTTGTGGAAAGGACGAAACACCG
P5_6nt_stagger*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTTGCACCTTGTGGAAAGGACGAAACACCG
P5_7nt_stagger*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTACGCAACTTGTGGAAAGGACGAAACACCG
P5_8nt_stagger*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC GCTCTTCCGATCTGAAGACCCTTGTGGAAAGGACGAAACACCG
P7_Barcode1	CAAGCAGAAGACGGCATACGAGATCGAGTAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCCAATTCCCACTCCTTTCAAGACCT
P7_Barcode2	CAAGCAGAAGACGGCATACGAGATTCTCCGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCCAATTCCCACTCCTTTCAAGACCT
P7_Barcode3	CAAGCAGAAGACGGCATACGAGATAATGAGGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCCAATTCCCACTCCTTTCAAGACCT
P7_Barcode4	CAAGCAGAAGACGGCATACGAGATGGAATCGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCCAATTCCCACTCCTTTCAAGACCT
P7_Barcode5	CAAGCAGAAGACGGCATACGAGATTTCTGAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCCAATTCCCACTCCTTTCAAGACCT
P7_Barcode6	CAAGCAGAAGACGGCATACGAGATACGAATGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCCAATTCCCACTCCTTTCAAGACCT

*The eight P5 primers were pooled for use in barcoding PCR

1) 95ºC, 60s 2) 95ºC, 30s 3) 53ºC, 30s 4) 72ºC, 30s 5) Back to step 2, 28 cycles 6) 72ºC, 10min 7) 4ºC, hold

Illumina MiSeq amplicon library prep/barcoding

DDX6_inner_2_fw	tcgtcggcagcgtcagatgtgtataagagacagTCATTAAGCAGCTCAGGACTGTAA
DDX6_inner_2_rv	gtctcgtgggctcggagatgtgtataagagacagACTCCCTCTGACGGGTAAAACAC
SMG9_inner_2_fw	tcgtcggcagcgtcagatgtgtataagagacagACAGCTCACAGGGTCACACTT

Thesis title: Pooled Genome-wide CRISPR/Cas9 screening as a tool to intensify the yield of cell-based influenza vaccines and other viral biologics

SMG9_inner_2_rv	gtctcgtgggctcggagatgtgtataagagacagACATGGCTTGACTGGAAGCACA
CARM1_inner_2_fw	tcgtcggcagcgtcagatgtgtataagagacagACAGGAGTGCAGGAACGAATGG
CARM1_inner_2_rv	gtctcgtgggctcggagatgtgtataagagacagAGCCCTCATCTAGCCCAAGTC
NXF1_inner_1_fw	tcgtcggcagcgtcagatgtgtataagagacagACTTACTCGGCTAAGCTGCTTCTG
NXF1_inner_1_rv	gtctcgtgggctcggagatgtgtataagagacagTGGGAGCTACTGGGTCCTTG
NXT1_inner_2_fw	tcgtcggcagcgtcagatgtgtataagagacagACTTGTTCCATTGTCAGGGCAGAAT
NXT1_inner_2_rv	gtctcgtgggctcggagatgtgtataagagacagAGTTCTGGTTGAAGTCCCGT
Outer_fw	aatgatacggcgaccaccgagatctacacATCGTAGCtcgtcggcagcgtc
Outer_rv	caagcagaagacggcatacgagatTAGCGAGTgtctcgtgggctcgg

1) 95^{o}C, 60s 2) 95^{o}C, 30s 3) 53^{o}C, 30s 4) 72^{o}C, 30s 5) Back to step 2, 28 cycles 6) 72^{o}C, 10min 7) 4^{o}C, hold