Controls Engineering Approaches To Regulating Immunity During Respiratory Infection

by

Emily E. Ackerman

Bachelor of Science in Chemical Engineering, Rensselaer Polytechnic Institute, 2015

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Emily E. Ackerman

It was defended on

July 16, 2021

and approved by

Robert Parker, PhD, Department of Chemical Engineering

Susan Fullerton, PhD, Department of Chemical Engineering

Zhi-Hong Mao, PhD, Department of Electrical and Computer Engineering

John Alcorn, PhD, Department of Immunology

Dissertation Director: Jason E. Shoemaker, PhD, Department of Chemical Engineering

Controls Engineering Approaches To Regulating Immunity During Respiratory Infection

Emily E. Ackerman, PhD

University of Pittsburgh, 2021

The human immune system is responsible for the detection and elimination of pathogens. The immune response to viral infection is comprised of a complex set of multi-scale interactions between small molecules, proteins, genes, and cells that govern pro-inflammatory and anti-inflammatory processes. Overactive inflammation is a root cause of severe clinical outcomes and can lead to systemic tissue damage and death. Identifying the drivers of dysregulation is key in the development and administration of therapeutic treatments to maintain equilibrium. The work presented aims to utilize high throughput and immunological data to determine the causal agents of immunoregulation using systems biology tools. Developed methods address this problem at the protein and systems levels using protein-protein interaction (PPI) network methods and ordinary differential equation (ODE) models.

Aim 1, the creation of the first ever disease specific subnetwork, identifies a set of proteins that are enriched for possible antiviral drug targets for influenza A infection. In Aim 2, PPI network controllability analyses are used to identify a set of 24 and 16 proteins acting as regulators of influenza A and SARS-CoV-2 infection, respectively. These proteins are further prioritized as targets in drug development/repurposing based on topology, function, and known targeting compounds. Five previously unidentified compounds are recommended for repurposing to treat COVID-19. Together, Aims 1 and 2 computationally produce efficient and meaningful biological results which align with in vivo findings.

Aim 3 explores influenza A strain-specific dynamics observed in immunological data by using ODE models to elucidate which biological mechanisms are at the root of differential behavior. Two models are constructed and parameterized to explore H1N1 and H5N1 influenza dynamics. Study reveals that only a small number of host functions likely

contribute to the strain-specific response, particularly the production rate of interferon. This finding is informative to future exploration of interferon-based therapeutics.

In total, both approaches are useful in teasing out the drivers of emergent properties of the complex immune response. By determining the consequences of the presence of a single component like interferon on other immune mechanisms, these studies enhance our understanding of disease progression and open the door to knowledgeable treatment design.

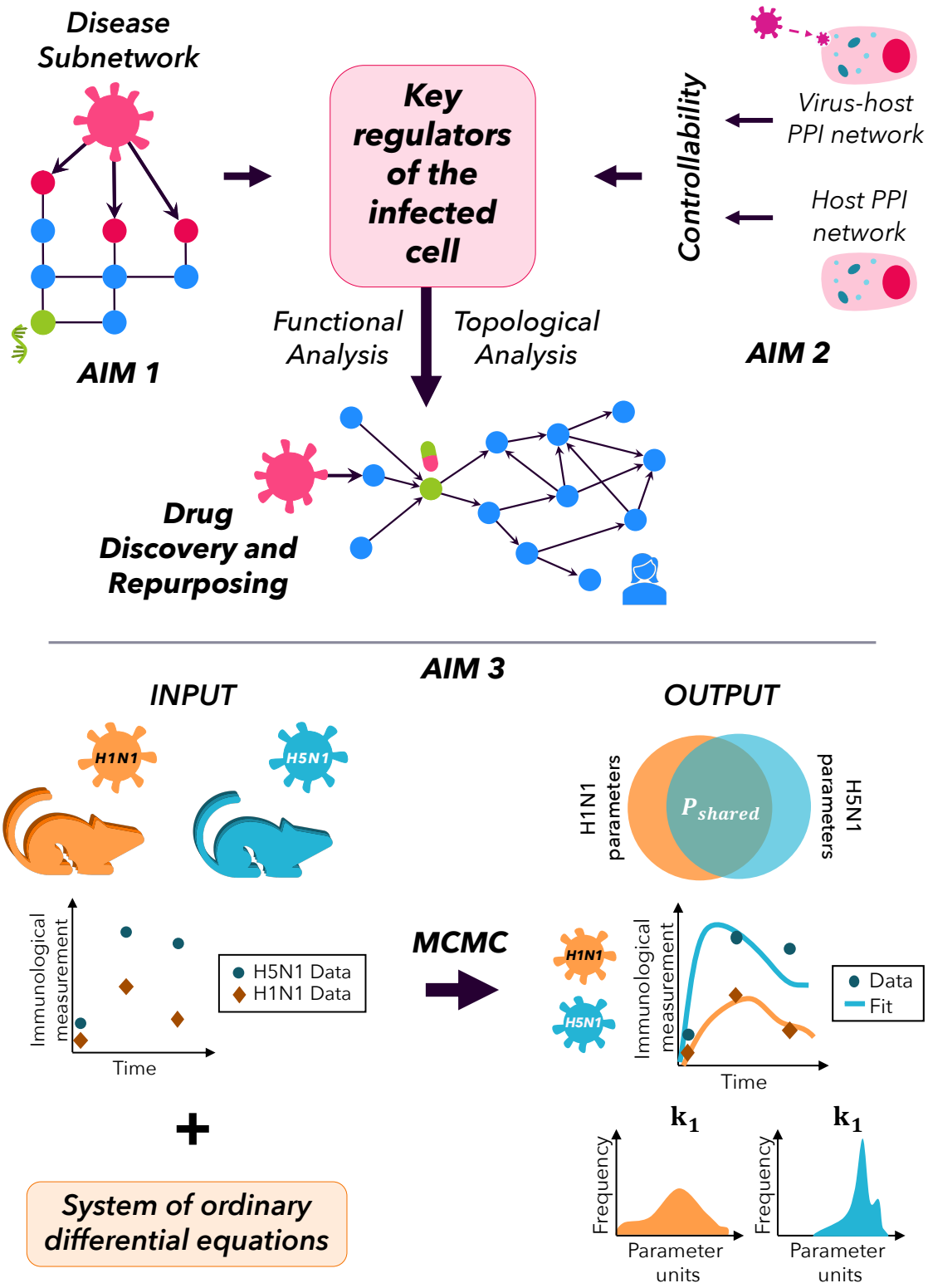


Figure 0.1. Visual Summary of Dissertation Aims. Aims 1 and 2 are the creation of protein-protein interaction network methods to identify disease host factors for drug target discovery. Aim 3 is the exploration of immune response dynamics to multiple strains of influenza virus using mathematical models.

Table of Contents

List of Tables

List of Figures

Preface

I wouldn't be where I am today without the love and support of the people who surround me. My mom, dad, and sister have been essential to my education. I'm lucky to come from a family so unconditionally nurturing and reassuring and thank them for their encouragement. In addition, my higher education experience would have been radically different without the love and support of my partner, Dave. I can't express how thankful I am to have had this group of people as my biggest believers over the past six years.

My time in the Shoemaker Lab at Pitt has shaped me and ability as a researcher. Thank you to Jason for the guidance I've received over the years and the freedom I was awarded to develop both my personal and professional identity as a member of the lab. I'm so grateful for the relationships formed with all of the other members of the Shoemaker Lab (particularly Robert, Muying, Jordan, and Ericka) who have spent up to six years each as my labmates, coding rubber ducks, friends, and co-conspirators.

Thank you to my committee members for coming with me on this journey and ensuring my success in my coming career. Additional thanks to our collaborators whose partnership has enabled me to complete this body of work.

I'm a big believer in the power of communities and couldn't have completed grad school without the ones I've been lucky enough to be a part of. Working to unionize with my fellow graduate workers always reminded me of the importance of my presence in academia and the value of my labor. The personal relationships I forged with members of GSOC were a source of solidarity that sustained me and give me hope for the future.

Thank you to Redhawk Coffee, where most of this research came to fruition, for keeping me caffeinated and allowing me to take residence in your tiny shop.

Thank you to the Howard Hughes Medical Institute for funding not only my work, but my whole self as a disabled researcher through the James H. Gilliam Fellowships for Advanced Study. The teachings and love I've received from the Gilliam fellows have helped me through the worst of days and given me a community that I will cherish for a long time.

Thank you to all my friends, near and far, for continuing to call me your friend when I forget to text back because I'm too distracted debugging my code.

It takes a village to get me out of bed in the morning, so I'd like to thank the 71 personal care aides that have worked for me over the course of my Ph.D. and been a source of true friendship.

On a more serious note, I'd like to bring attention to the fact that disabled children, teens, and adults are implicitly and explicitly taught that we weren't destined for academic pursuits. In elementary school, I watched my mom fight against the idea that it was acceptable to remove me from classes for physical therapy. In middle school, a science teacher asked my mom "what I was ever going to do". We are shown that there are lower expectations for our educations. We are taught that science looks a certain ablebodied way. We are led by example to believe that to succeed in science, you must suffer, and that an inability to withstand it is a flaw in ourselves. Of course, none of this is true, but continues to act as a barrier to equity every day, keeping disabled people from ever entering higher education and driving out the ones who do. I can't give enough thanks to those who have taught me to believe in myself and supported me in advocating for what's right, especially my mom and the disability activism community. I'm proud of my academic accomplishments but feel particularly gratified by my personal growth over time and what I hope are the lasting effects of my work towards true equity in academia.

Last but certainly not least, I'd like to acknowledge that this work was completed on Shawanwaki/Shawnee and Osage land. Academics at the University of Pittsburgh have long benefitted from colonization that displaced the Indigenous people of Pennsylvania, myself included. I call on the community to actively recognize this and intentionally foster relationships with Indigenous peoples which responsibly address land stewardship and move forward in ways which honor their contributions and customs.

1.0 Introduction

1.1 The Host Immune Response to Viral Infection

1.1.1 Viral Respiratory Infection

The prevalence of viral pathogens, particularly those leading to lower respiratory infection and their associated high mortality rates, demand an international effort to control global infection spread. The clinical presentation of respiratory infections vary by virus but share common pathology in wheezing, coughing, fever, and congestion of the upper and lower respiratory tract [1, 2, 3]. Common, yearly epidemics of viruses such as influenza virus make respiratory infection the fourth leading cause of death worldwide [4]. Claiming 2.8 million lives in 2010, respiratory infections are most commonly observed in at-risk populations including the immunocompromised, elderly, and pregnant and in children [5, 6, 7, 8]. Morbidity and mortality in these groups are elevated in comparison to the total population. These populations are also more susceptible to co-infection with bacterial and secondary viral infections such as S. pneumoniae which greatly increase the mortality rate [9].

Up to 75% of novel viruses are considered zoonotic, developing in an animal reservoir and subsequently transmitting to human hosts [10]. Emergent viruses pose a particular threat to public health as the absence of existing immunity can lead to high human-human transmissibility, severe pathology, and high mortality rates. In particular, viral adaptation within animal reservoirs dictates the immune response needed. For example, naturally high levels of inflammation in bats drive intense inflammatory responses in humans infected with viruses that have jumped from bats to humans such as coronaviruses [11]. The onset of emergent virus epidemics necessitates a period of rapid scientific discovery to identity the viral structure, mechanisms of viral replication and spread, and the invoked immune response to better tailor treatment development. Living in continual threat of emergent strains, pandemic viruses have become reality throughout recent history including

pandemic influenza A pH1N1 virus, severe acute respiratory syndrome (SARS), middle eastern respiratory syndrome (MERS), and SARS-CoV-2. The typical timeline for FDA approval from start to finish averages 12 years [12] with a 25% success rate for infectious disease treatments [13]: an infeasible timeline for emergent viral outbreaks. Time and effort, precious commodities in an outbreak, can be saved through the incorporation of knowledge learned from previous study of related viruses and treatments.

Classified within families by genome, size, physiochemical characteristics, and morphology, insight can often be transferred across viruses as many characteristics are conserved including the general steps of the viral life cycle [14]. For enveloped viruses [15], spike proteins on the outside of the viral envelope interact with host lung epithelial cell surface receptors to facilitate viral entry. After endocytosis and the release of viral genetic material, viruses depend on host cell machinery for replication of viral RNA. New viral particles are then packaged and leave the cell to continue the cycle. However, differences in viral protein structure, transcription/translation behavior, host protein interactions, etc. create discrepancies between disease mechanisms and regulation. Understanding the individualized drivers of disease-specific behavior is critical in halting viral transmission and developing therapeutic treatments.

1.1.1.1 Influenza A Virus

Influenza virus is an enveloped, negative-sense single stranded RNA (ssRNA) virus of family Orthomyxoviridae [16]. There are four types (A, B, C, D) with influenza A being the most common and severe in humans during yearly seasonal outbreaks. Common symptoms include fever, chest congestion, body aches, and fatigue. The FDA has approved six total antiviral treatments, although two have been retired due to drug resistance observed in recent strains [17]. Subtype strains are characterized by antigenic surface proteins hemagglutinin (HA) and neuraminidase (NA) that serve as flags for the host immune system to recognize the pathogen [18]. With 18 known HA types and 11 known NA types, there are 198 theoretical strains of influenza A [19]. H1-H3 strains are considered human influenzas, though there is evidence that avian influenzas H5, H7, and H9 are infectious to humans. For example, human

infection with the high pathogenic avian H5N1 influenza has occurred in 16 countries to date in small numbers [20]. H5N1 is noted for its severity in humans, leading to characteristic lower respiratory tract infections with high levels of inflammation and critical damage of lung epithelial tissue [21]. Estimated mortality rates reach 60%, including young populations, a trend atypical of other influenza strains [22]. The threat of H5N1 outbreak and pandemic spread remain on the horizon and the world continues to prepare for the event including ongoing vaccine stockpiling in the United States.

New human strains of influenza A virus emerge through contact with zoonotic hosts including birds, bats, horses, and pigs which serve as reservoirs for mutation. Influenza A is noted for a high level of point mutations within the genome (genetic drift) and genetic recombination of multiple influenza strains (genetic shift), both of which are responsible for the emergence of new viral strains [23]. Vaccines are tailored to circulating viruses each year to provide protection against recent viral evolution [24]. Absence of existing immunity to limit human transmission of animal influenzas increases the likelihood that novel strains reach epidemic and pandemic levels of infection spread.

1.1.1.2 SARS-CoV-2

SARS-CoV-2 is a positive-sense ssRNA virus of family Coronaviridae enveloped within the nucleocapsid (N), membrane (M), and envelope (E) structural proteins [25]. The spike glycoprotein (S) mediates host cell binding and viral entry. Related viruses SARS-CoV (2002) and MERS-CoV (2012) caused pandemics noted for excessive mortality rates (estimated up to 10% and 36%, respectively) [26, 27]. COVID-19, the illness caused by SARS-CoV-2 infection, transmits from human to human through aerosols and droplets shared in close contact [28, 29]. The disease presents both asymptomatically and with mild and severe infection pathology [30, 31]. Mild and severe COVID-19 share many clinical symptoms including fever, fatigue, respiratory congestion, gastrointestinal distress, and loss of taste and smell [32]. Severe cases display extreme lung and cardiovascular damage and neurological impairment. Evidence shows distinct immune response in mild versus severe patients with notable differences of depleted CD8+ T-cells and increased pro-inflammatory

cytokines leading to worsened clinical outcomes and increased mortality rates [31, 33]. Approximately 13% of COVID cases are characterized as "Long COVID" with patients exhibiting extended fatigue, neurological impairments, and more after clearing a severe infection [34].

As of June 2021, the ongoing SARS-CoV-2 COVID-19 pandemic has claimed an estimated 3.7 million lives of the 173 million confirmed cases and continues to infect millions globally [35]. The possibility of reinfection with COVID-19 still remains unknown. In the interest of time, attempts at COVID-19 drug treatment have revolved around repurposing existing drugs [36, 37, 38] and monoclonal antibody treatment [39]. The development of several SARS-CoV-2 vaccines has largely contributed to a downward trend in cases since early 2021 with available vaccines being successful at preventing 65-97% of symptomatic disease [40]. Recent, more transmissible variants have been identified, accelerating the need for advanced treatment options and higher global vaccination rates [41].

1.1.2 The Host Immune Response

The host immune response, categorized into the early innate arm and the late adaptive arm, is a series of mechanisms responsible for the elimination of invading pathogens including viral and bacterial infections. The innate arm is responsible for the initial response and is triggered when foreign material, including viral RNA, is recognized by sensors of characteristic pathogen molecules called pattern recognition receptors. These receptors reside in the cytoplasm and on the surface of epithelial cells, resident macrophages, neutrophils, and dendritic cells [42]. Stimulated RNA sensing pathways trigger the production of type-I interferons by infected host cells, a molecule that alerts surrounding cells of the pathogen's presence and upregulates immunoregulatory pathways including the transcription of "interferon stimulated genes". Pro and anti-inflammatory cytokines secreted by resident macrophages and neutrophils act as chemoattractants to flag the site of infection and dictate leukocyte growth and differentiation [43, 44, 45]. Chemokines released as a product of the activation of pro-inflammatory pathways such as

NFkB are responsible for leukocyte recruitment at the site of infection [46]. New macrophages, neutrophils, and dendritic cells join resident alveolar populations within hours to eliminate cellular debris and viral particles [46, 47]. After several days, natural killer cells identify and induce apoptosis to limit viral replication in infected cells through the release of cytotoxic granzymes [48].

Dendritic cells scour the interstitial space between lung epithelial cells looking for pathogen-associated molecular patterns (PAMPs) [49]. When exposed to a viral pathogen, they mature and migrate to the lymph node, initiating the adaptive immune response [50]. After processing the antigen, it is presented to naïve T cells by major-histocompatibility class (MHC) molecules on the cell surface to induce antigen-specific differentiation [51]. MHC class I molecules activate cytotoxic CD8+ T cells that release cytotoxins to induce apoptosis in infected epithelial cells. MHC class II molecules activate helper CD4+ T cells that release cytokines to aid other lymphocytes. B cells uptake antigen in the lymph node and present it to naïve T cells with MHC class II molecules and helper CD4+ T cells express cytokines to promote eventual B cell differentiation into plasma cells and memory B cells among others [52].

Differentiated, antigen-specific cells typically migrate to the site of infection beginning 3-6 days post-onset of infection [52]. Plasma cells release antibodies that neutralize antigens and memory B cells launch a robust antibody response upon encountering a known antigen. As antigen levels decrease, the release of cytokine and chemokine signals drop off and T cells begin suppressing cytokine release, ceasing immune cell infiltration and ending the immune response to infection [46]. Chronic inflammation caused by dysregulated homeostasis between pro and anti-inflammatory cytokines can lead to extensive tissue damage, acute respiratory distress syndrome, cytokine storm, sepsis, and death [53]. As the timing and magnitude of the invoked immune response is disease specific, even strain specific within viruses [54], understanding the complex dynamics of immunoregulation is necessary to restore homeostasis in the case of dysregulation and develop disease treatments.

1.2 Computational Methods

Computational exploration of biological systems is reliant on supporting data, much of which is generated as a result of recent advancements in experimental methods [55, 56, 57]. In particular, the rise of high throughput processing techniques to simultaneously measure the concentrations of high numbers of genes, proteins, cells, and more with previously intractable repetition has opened the door to a new level of analysis. Resulting omics' data (for example, RNA-seq and microarray transcriptomics or mass spectroscopy-based metabolomics and proteomics) is a descriptive snapshot of complex, multiscale systems that holds the promise of unlocking our biological puzzle upon successful processing of these dense, large-scale datasets [55, 58, 59]. In addition, gene knockdown and knockout studies that grant the ability to observe the effects of downregulated or eliminated expression of specific genes responsible for desirable biological mechanisms have allowed for highly specified *in vivo* exploration [60]. As these historically expensive data collection methods become increasingly affordable and accessible [61], efforts have shifted toward the development of novel methods to analyze this information and achieve greater understanding of biological systems and disease pathology [62].

Novel methods for processing gene expression, protein interaction data, and immune cell counts during influenza A virus infection have exhibited predictive power in characterizing system dynamics [63, 64, 65, 66, 67], predicting disease host factors [68, 69], and optimizing therapeutic targets [70, 71]. These efforts work toward an ultimate goal: to reveal and manipulate the molecular and cellular interactions regulating influenza virus infected cells and terminate infection. While system complexity and experimental or computational limitations [72] have historically stood as barriers to this goal, each step toward discerning the dynamics of infected cells and systems will provide insight into disease pathology and optimize the path to effective clinical treatment.

The lens of scale with which we choose to view viral infection influences the computational direction of our work. Questions of spatial resolution are quickly followed by those of temporal resolution and data availability. For example, immune activity at the lung tissue level is the result of cell signaling pathways activated by continuous regulation

at the genetic and protein scale. Experiments, however, result in discrete measurement timepoints that lend mystery to the events of the in-between. As data collection from animal models requires full lung samples, longitudinal dynamics across time cannot be tracked within a single animal. In addition, evidence that gene transcription does not directly correlate to protein translation [73, 74], inherent time delays built into biological pathways, and the presence of dense, convoluted feedback mechanisms across spatial scales inserts a nonlinear complication into attempts at computational modeling. Multiscale modeling is growing in popularity [75, 76, 77, 78, 79, 80] based in the creation of data processing methods at singular spatial scales across time.

The work contained within develops an understanding of the host response to viral respiratory infection on two spatial scales. First, at the protein-level through the design of protein-protein interaction (PPI) network methods of understanding implicated molecular interactions, and second, at the systems level in the creation and parameterization of mathematical models of the immunoregulation of immune cells and cytokines during viral infection. While both approaches hold predictive power in isolating cell system mechanisms and components with significance to disease progression, the combination provides a comprehensive understanding of viral infection and corresponding host immunoregulation.

1.2.1 Mathematical Modeling

Ordinary differential equations (ODEs) are a field of mathematics used to describe change in state per change in an independent variable, often time [81]. ODEs are commonly employed to determine the species dynamics of biological systems [82, 83, 84]. Equations are constructed from literature understanding of the interactions between each relevant species where kinetic parameter values represent the rates of individual biological processes (for example, the rate of host cell infection by virus, or the rate of natural cytokine decay). These parameters can be interpreted to understand how complex interactions lead to system-level behavior. Common computational solvers for ODEs in MATLAB, python, and Julia use finite differencing methods to estimate derivatives over

small time steps, allowing for fast, powerful estimation of species concentrations over desired timeframes of study [83].

Before reaping the promised benefits of ODE models, kinetic rates must be determined to specify system behavior. Model creation and optimization are an iterative process as parameter fitting may reveal underfitting or overfitting (for example, a linear model failing to capture true trends from non-linear data or a model containing more free parameters than are necessary to explain system dynamics) which require model reformulation. Parameterization can be achieved in two ways. First, rate kinetics available from experimental reports in literature can be directly substituted into the model equations. However, the availability of this data is often limited, and may be derived from experimental conditions that stray too far from those of the training data used for the estimation of the remaining parameters, meaning that this method should be approached with caution.

The second method of parameterization relies on mathematical estimation of the parameter combination that best reproduces experimental training data. Minimized metrics such as sum of squares error can be used to calculate the difference between simulation output and training data across available timepoints to iteratively optimize the best fitting parameter set from models such as regression models [85]. While the least computationally expensive, curve fitting methods fall victim to several pitfalls due to insufficient knowledge of possible parameter space prior to fitting (Figure 1.1). Parameter identifiability, the ability to identify a singular optimal solution opposed to several parameterizations which yield equally minimal error, is required to determine the values of unknown parameters and is difficult to assess without exploring the entirety of parameter space. Additionally, simple optimization methods often fail to escape local minima in their search, ultimately returning non-optimized solutions.

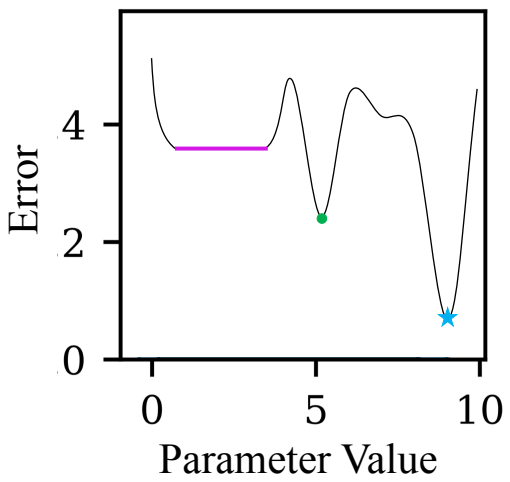


Figure 1.1. Error function output demonstrating unidentifiable solutions (pink), a local minimum (green dot), and the global minimum (blue star).

To address the problems associated with under characterized parameter space, probabilistic methods of parameterization instead return a distribution of possible solutions with associated probabilities of occurrence based on training data. Probabilistic methods are founded on Bayes Theorem,

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{1.1}$$

The posterior probability, $P(\theta|D)$, that a parameter set, θ, can explain dataset D is a function of three probabilities. The prior, $P(\theta)$, details information known about the parameters (often assumed to be a uniform distribution in the event there is no knowledge of the true distribution). The evidence, $P(D)$, indicates if the data was generated by the model, and the likelihood, $P(D|\theta)$, describes how well the data fits parameter set θ. While the posterior cannot be directly calculated from these values, an estimation can be determined through large-scale sampling of multidimensional parameter space.

Markov Chain Monte Carlo (MCMC) is an algorithm that samples probability distributions by stepping stochastic chains through possible parameter space. Advantageously, a Markov Chain's probability at the current step is only a function of the previous step, making past and future parameter sets independent and memoryless. The most commonly implemented MCMC method is the Metropolis-Hastings algorithm [86]. The distance between a parameter set and the training data can be quantified through a customizable objective function which are often based off of the residual sum of squares:

$$E = \sum_i (\bar{y}_i - y_i)^2 \tag{1.2}$$

where $\bar{y}_i$ is simulated output from timepoint i and y_i is experimental data at timepoint i. The likelihood that experimental data could be generated from a given parameter set is a function of error:

$$P(D|\theta) \propto \exp(-E) \tag{1.3}$$

At initialization, the error between training data and simulation output for an initial parameter set, θ, is determined. At each step, each parameter jumps to a new value and the error is re-determined. The algorithm must choose between accepting the new parameter set

and returning to the value of the previous step with a probability of acceptance equal to the ratios of the posterior distributions computed using known likelihood and prior assumptions:

$$\alpha = min(1, \frac{P(\theta_{i+1}|D)}{P(\theta_i|D)}) = min(1, \frac{P(D|\theta_{i+1})P(\theta_{i+1})}{P(D|\theta_i)P(\theta_i)}) \tag{1.4}$$

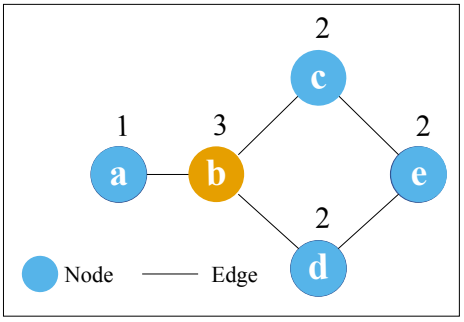
The acceptance probability is compared to a random number, r, from $U[0,1]$ and the new parameter step is accepted for $r < \alpha$ that serves to always accept improved solutions ($\alpha = 1$) and accept inferior solutions with some probability that promotes exploration of parameter space. This sampling process is repeated on a large scale (millions of iterations) to ensure that all multidimensional parameter space is adequately explored before defining the most likely parameter set. To enhance the algorithm's ability to survey possible solutions, parallel tempering MCMC (PTMCMC) explores parameter space by initializing multiple parallel chains and incorporating a chain temperature which influences the size of the jump to the next proposed parameter set and the probability of acceptance [87]. When a high probability solution is identified by a high temperature chain making large jumps, swapping with a low temperature chain works to optimize the solution using small jumps in the area of high probability. Chain swapping is particularly useful for avoiding entrapment in local minima as high temperature chains more efficiently cover all possible parameter space.

While an impossible, infinite sampling would be needed to identify the true posterior and convergence remains impossible to prove, MCMC convergence diagnostics serve as the next best identifier of an optimized solution by assessing whether certain characteristics of convergence hold true [88]. Common graphical methods to check convergence include plotting parameter values over time to ensure the chain moves randomly and not toward a specific end, which would suggest insufficient tuning, and checking resultant parameter distributions to ensure they are not pushed against pre-defined bounds, if applicable. Additional mathematical metrics such as autocorrelation functions can be used to assess the performance of the algorithm. Parameter correlation at the root of convergence issues can be addressed with temperature tuning in PTMCMC [89]. A healthy MCMC study should display a balance between the number of accepted and rejected solutions, a characteristic conveyed by the acceptance ratio. For basic applications of MCMC, an

acceptance ratio of 23% is commonly considered the ideal [90] that balances the needs of sampling versus computational efficiency.

1.2.2 Network Modeling

It would be neat and tidy to attribute observed biological phenomena to the actions of singular molecules within the cellular system, however, the reality is much messier. The interactions between entities from small molecules to proteins, DNA, and cells are governed at the root by a complex set of rules where no two components can be completely separated. With the rise of high throughput screening methods resulting in massive protein-protein interaction (PPI), transcription regulation, and metabolic datasets, the question becomes: how can we model large-scale biological datasets to derive the greatest biological understanding of the systems they represent? Framework from the field of graph theory provides a means to understand the interconnectedness of biological systems [91]. Networks consist of nodes, the components of the network, and edges, the interactions between network components. Edges can be directed, meaning the interaction has a meaningful flow direction, or undirected, and can hold weights representing probabilities. Measures of network topology, or the structure and arrangement of network components, are used to quantify the importance of nodes and edges to the overall function of the system. Common topology measures include distances between nodes, degree (the number



Degree: the number of interactions with node n *(labeled)*

Betweenness: the importance of node n to network information flow

$$B(n) = \sum \frac{s_{ij}(n)}{s_{ij}} \quad \text{for} \quad i \neq j, n \neq i \text{ and } n \neq j$$

where: s_{ij} *is the total number of shortest paths between nodes i and j*

$s_{ij}(n)$ *is the number of* s_{ij} *paths passing through node n*

$$Ex:\ B(b) = \frac{s_{ac}(b)}{s_{ac}} + \frac{s_{ad}(b)}{s_{ad}} + \frac{s_{ae}(b)}{s_{ae}} + \frac{s_{cd}(b)}{s_{cd}} + \frac{s_{ed}(b)}{s_{ed}} + \frac{s_{ce}(b)}{s_{ce}}$$
$$= \frac{1}{1} + \frac{1}{1} + \frac{2}{2} + \frac{1}{2} + 0 + 0 = 3.5$$

Figure 1.2. An example of common topology measures degree and betweenness

of interacting nodes), and betweenness [92] (a measure of the importance of a node to totalnetwork information flow) (Figure 1.2) [93, 94, 95]. Networks have been used to model relevant relationships in a variety of fields including social networks, shipping optimization, and biological systems.

PPI data is an excellent candidate for network representation and study. Curated in annotated databases, both computationally predicted and experimentally verified PPIs are easy and efficient to pick up and utilize [72, 96, 97]. While some computational efforts to direct PPIs have been successful, the majority of available data is undirected and unweighted. Networks constructed from unweighted PPI data create a whole cell model of system behavior by viewing intercellular protein interactions as binary events. In opposition to mathematical modeling methods that require extensive knowledge of biological mechanisms and quantitative kinetic rates, network methods place value on the relationships between proteins in a system much larger than that which could be practically modelled with a system of ODEs. As interdependent states, the exploration of cellular level relationships relies on the effects of network flow on systemic behavior, a view that is highly reflective of the cascade behavior and feedback mechanisms found in biological pathways.

Previous studies have consistently identified disease relevant proteins missed by traditional modeling efforts [98, 99, 100, 101, 102]. While useful, these analyses fall victim to two drawbacks of biological network analysis: their generality and static nature. PPI data is derived from proteome-wide repositories, meaning that most network analyses are not cell type or disease system specific. Additionally, the lack of temporality imposed by both the experimental methods and the collective nature of data curation and organization leads to difficulty advancing dynamic understanding of the system. New methods are needed to address disease specificity and acknowledge the dynamics of highly connected biological systems while maintaining the benefits that come with the availability and qualitative nature of the data.

2.0 Aim 1: Disease Subnetwork Extraction for Host Factor Identification

2.1 Introduction

Yearly outbreaks of influenza A virus have a major impact on public health and the global economy each year [103, 104]. While annual vaccinations provide some protection, vaccination effectiveness is impaired by antigenic drift and availability issues [105, 23]. Recent human infections with highly pathogenic H5N1 and H7N9 subtypes of avian influenza virus have raised concerns about the potential of pandemic spread [106, 107, 108, 109]. Antiviral drugs to target the endocytosis and replication activity of influenza viral proteins are available [110, 111], but they are limited [17] and drug-resistant strains have emerged [112, 113]. Therefore, there is an urgent need to identify drug targets that are robust to virus mutation and drug-mediated selective pressure.

Understanding virus-host interactions in the context of the human protein-protein interaction (PPI) network provides a global perspective into how influenza virus manipulates host proteins and aids in identifying host factors involved in influenza virus replication for drug targeting [102, 114, 115]. The virus-host interaction network, a network containing both host-host interactions and virus-host interactions, is visualized in Figure 2.1 for the human host and influenza A virus' 11 proteins.

A protein's global significance to the PPI network can be assessed by its network centrality, measures of information flow across the network. Common measures used in protein network biology include degree (number of binding partners) and betweenness (the importance of a protein to total network flow) though several others exist [116, 117]. Network centrality has emerged as a valuable tool for studying proteins associated with cancer [118, 119] and drug targeting [119, 120, 98, 121]. PPI network-based approaches have recently been utilized in influenza virus studies to identify and study potential factors involved in virus replication [122, 123, 124, 125, 126] focusing on the host proteins that directly interact with viral proteins. Network studies have demonstrated that virus interacting host proteins tend to have a high network significance based on a variety of

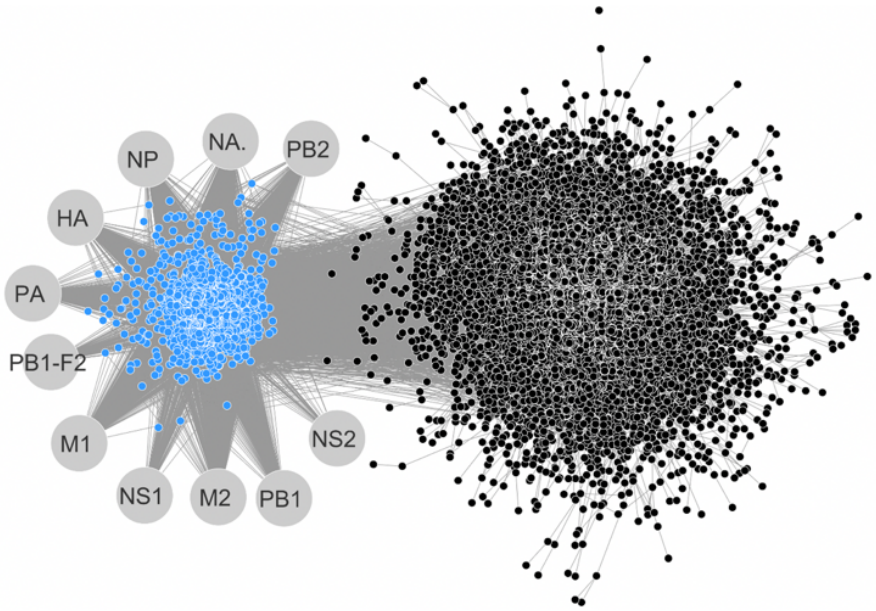


Figure 2.1. The virus interacting network is created from human host-PPI data combined with virus-host protein interaction data for influenza A virus.Here, influenza A virus proteins are depicted in grey, virus interacting host proteins are depicted in blue, and non-virus interacting proteins are depicted in black.

network metrics (including betweenness and degree) for several viruses, including influenza viruses [127] and hepatitis C virus [128]. A comparative analysis of influenza virus protein and host protein interactomes has identified novel host factors that are common across the protein interactomes [129]. Furthermore, meta-analysis studies have developed statistical methods to better identify host factors by leveraging data from several virus replication screens [69]. However, the limits of studying the network topology of virus interacting host proteins to improve host factor identification (i.e., antiviral drug target identification) have likely been reached.

New methods that move beyond whole cell network topology are necessary to assess the potential of proteins downstream of virus interacting proteins to act as disease-specific host factors. Here, we demonstrate a method of identifying a disease-specific subnetwork of the integrated virus-host PPI network and assess its enrichment with potential antiviral drug target candidates.

2.2 Materials and Methods

2.2.1 Protein-Protein Interaction Network Construction and Topology

Protein-protein interaction (PPI) data was downloaded from the Human Integrated Protein-Protein Interaction rEference (HIPPIE) database [130] (version 1.4). Interactions with a confidence score of less than 0.7 were removed to ensure high experimental reproducibility and accuracy. The network was constructed using the igraph package in R. The high confidence interaction data resulted in one large network containing 9,969 proteins and 86 smaller networks which were disconnected from the large network (most with 2 proteins, all containing 7 or fewer). Disconnected networks were removed from the study to eliminate bias in topology calculations. The final human host PPI network contains 9,969 proteins and 57,615 interactions. A previous study identified 1,292 host proteins that co-precipitated with at least one of 11 influenza A virus proteins (viral PB2, PB1, PA, HA, NP, NA, M1, M2, NS1, NS2, and PB1-F2 proteins) [131]. These proteins, referred to as "virus interacting proteins", all mapped to the host PPI network (i.e.: no virus interacting proteins had been removed in the disconnected networks). The topology characteristics of virus interacting and non-virus interacting host proteins within the PPI network were determined using the R igraph library version 1.0.1 [132].

2.2.2 Calculation of Abundance-Degree Correlation

To investigate the effects of correlation between protein abundance and network degree, a linear model was fit to the total network's abundance log_{10} degree data using R 3.2.2's glm function. HEK293 cell abundance data from Geiger et al. [133] was used to avoid the effects of differing cell lines. The correlation slope (0.093) was used to calculate adjusted degree values as follows:

$$adjusted\ degree = (log_{10}(original\ degree) - slope_{lm} * abundance) + intercept_{lm} \quad (2.1)$$

where lm is the linear model. The final values reported are $10^{adjusted\ degree}$.

2.2.3 Protein Network Clustering

A measure of the closeness of virus interacting proteins within the network was calculated using the shortest distances required to connect all virus interacting proteins, creating a distribution of distances. The cumulative distribution details the fraction of host proteins that could be connected to other host proteins that bind the same viral protein in n or fewer steps. A cumulative distribution of distances that result from randomly sampled proteins in the network was used as a control. A set of random proteins of size of the number of proteins that interact with the virus protein of interest (e.g., PB1 has 322 interacting host proteins; therefore, 322 proteins were randomly selected from the network) were pulled per iteration. The distribution of distances connecting the randomly sampled proteins was calculated. This process was repeated 100 times.

To evaluate whether virus interacting proteins are components of a common protein complex, the fraction of all virus interacting protein pairs (735,078 pairs in total) that appear within a protein complex was determined and compared to the fraction of all protein pairs (49,685,496 total pairs) in the PPI network that appear in a protein complex. Mammalian protein complex information was downloaded from CORUM (a comprehensive resource of mammalian protein complex data) [134].

2.2.4 Influenza Virus-Host Subnetwork Construction

A partial siRNA screen to identify key host factors that do not interact directly with viral proteins was performed to address problems of reproducibility resulting from differences in experimental conditions and characteristically high false-negative rates in siRNA screens of influenza virus replication host factors [135]. HEK293 cells were transfected with siRNAs targeting 264 non-virus interacting host factors identified by Karlas et al. [136] (two siRNAs per gene were used; AllStars Negative Control siRNA [Qiagen] was used as a negative control) and then infected with influenza A virus at 24 h post-transfection. The culture supernatants were harvested for virus titration at 48 h post-infection. Virus titers were determined by plaque assay. A protein was defined as a hit if the virus titers decreased by at least two log units upon transfection with an adjusted Pvalue of < 0.01. The viability of siRNA-

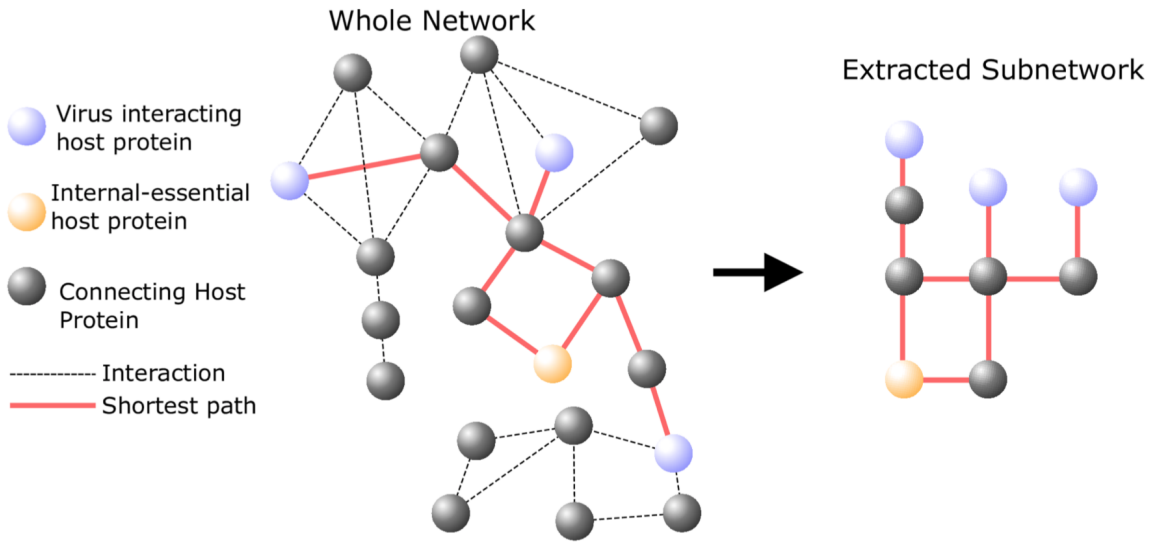


Figure 2.2. A disease specific subnetwork is extracted from the human PPI network. It contains the shortest paths (red) connecting virus interacting proteins (blue) to host proteins that are essential to virus replication (internal-essential proteins, orange). The connecting proteins (black) are candidates for antiviral targets.

transfected cells was assessed using Cell-Titer Glo assay, and downregulation of mRNA levels for the hit proteins was confirmed by quantitative reverse transcription-PCR (qRT-PCR). Of the 264 previously identified host factors tested, 71 significantly downregulated virus replication. Of the 71, 21 of the host factors were identified to directly interact with influenza virus proteins while the remaining 50 host factors downregulated virus growth and do not directly interact with the virus. Proteins of the latter group were labeled as "internal-essential" host factor proteins.

An influenza virus-specific subnetwork (process illustrated in Figure 2.2) was constructed from the shortest paths connecting virus interacting host proteins to internal-essential host factors. The proteins linking internal-essential proteins to virus interacting proteins were labeled "connecting" candidate proteins for evaluation as host factors of virus replication. The resulting subnetwork contains 1,213 virus interacting proteins, 38 internal-essential proteins (12 proteins were not in the PPI network), and 1,643 connecting candidate proteins.

2.2.5 Statistical Analyses

All statistical tests were performed in R 3.2.2 using the functions prop.test, fisher.test, pairwise.t.test or wilcoxon.test (which performs a Mann-Whitney U test) as appropriate. Prop.test and fisher.test both compare outcome proportions between binomial groups with the latter being more precise for small group sizes.

2.2.6 Functional Analysis

A functional enrichment analysis was performed using DAVID 6.8's Functional Annotation tool [137].

2.2.7 Calculation of Host Factor Enrichment

Protein enrichment was assessed by selecting the 78 proteins of the subnetwork with the highest (n = 39) and lowest betweenness (n = 39) and conducting an additional *in vitro* virus replication assay. HEK293 cells were again transfected with siRNAs targeted to the genes of the 78 candidate proteins. The procedure described previously was performed to determine the proportion of connecting proteins tested that are host factors of influenza virus replication. Enrichment is determined by a hit rate defined as the proportion of proteins tested that significantly downregulate virus replication. Differences between hit rates were compared using Pearson's chi-squared test to compare proportions between two binomial groups.

2.3 Results

2.3.1 Virus Interacting Proteins are Central to the PPI Network

Studies have shown that proteins in network positions essential to information flow within the PPI network (high degree or high betweenness) are more likely to be associated with disease [138, 139] or drugs with known, dangerous side effects [140, 141]. Virus proteins

are significantly more likely to interact with host proteins in positions of high regulatory importance. For every protein, the degree (number of interacting proteins) and betweenness [142] (a measure of a protein's importance to efficient network flow) were calculated. On average, the degree of virus interacting host proteins is twice the average degree of all proteins and is significantly higher than the degree of the non-virus interacting proteins of the network (Figure 2.3A; median degree of virus interacting, non-virus interacting, and all proteins: 10, 4, and 5, respectively; Student t-test P value for comparing log-scaled non-virus interacting and virus interacting data: $< 10^{-16}$).

Bias introduced by the relationship between protein abundance and a protein's network degree is documented in literature [143]. An analysis of the correlation between the two variables was performed to ensure that the high degree of virus interacting proteins was not an artifact of protein abundance. Significant correlation exists between abundance and degree in the virus interacting proteins (Pearson correlation coefficient: 0.23; P value: 1.2 x 10^{-13}), subnetwork (Pearson correlation coefficient: 0.10; P value: 6.9 x 10^{-7}), and total network (Pearson correlation coefficient: 0.25; P value: 2.2 x 10^{-16}). (Figure 2.4) After fitting degree and abundance to a linear model to remove this bias, a comparison of the adjusted degrees of the non-virus interacting and virus interacting proteins reveals that the previous conclusions remain: the degree of virus interacting proteins is significantly higher than the degree of non-virus interacting proteins in the network (Figure 2.3B; median degree of virus interacting, non-virus interacting, and all proteins: 5.4, 3.9, and 4.2, respectively; Student t-test P value comparing log-scaled non-virus interacting and virus interacting data: $< 10^{-16}$). Therefore, despite known biases engrained in PPI data, virus interacting proteins interact with a higher number of neighbors than non-virus interacting proteins.

Virus interacting proteins also had a significantly higher betweenness (Figure 2.3C; median betweenness of virus interacting and all proteins: 1,625.1 and 32.8, respectively; Mann-Whitney U test P value for log-scaled data: $< 10^{-16}$). Comparing median betweenness after the removal of proteins with a betweenness of zero, virus interacting proteins still had a significantly higher betweenness though the population medians were closer in value (median betweenness of virus interacting and all proteins: 3,981.1 and 1,584.8, respectively; Mann-Whitney U test P value for the log-scaled data: 8.2 x 10^{-16}).

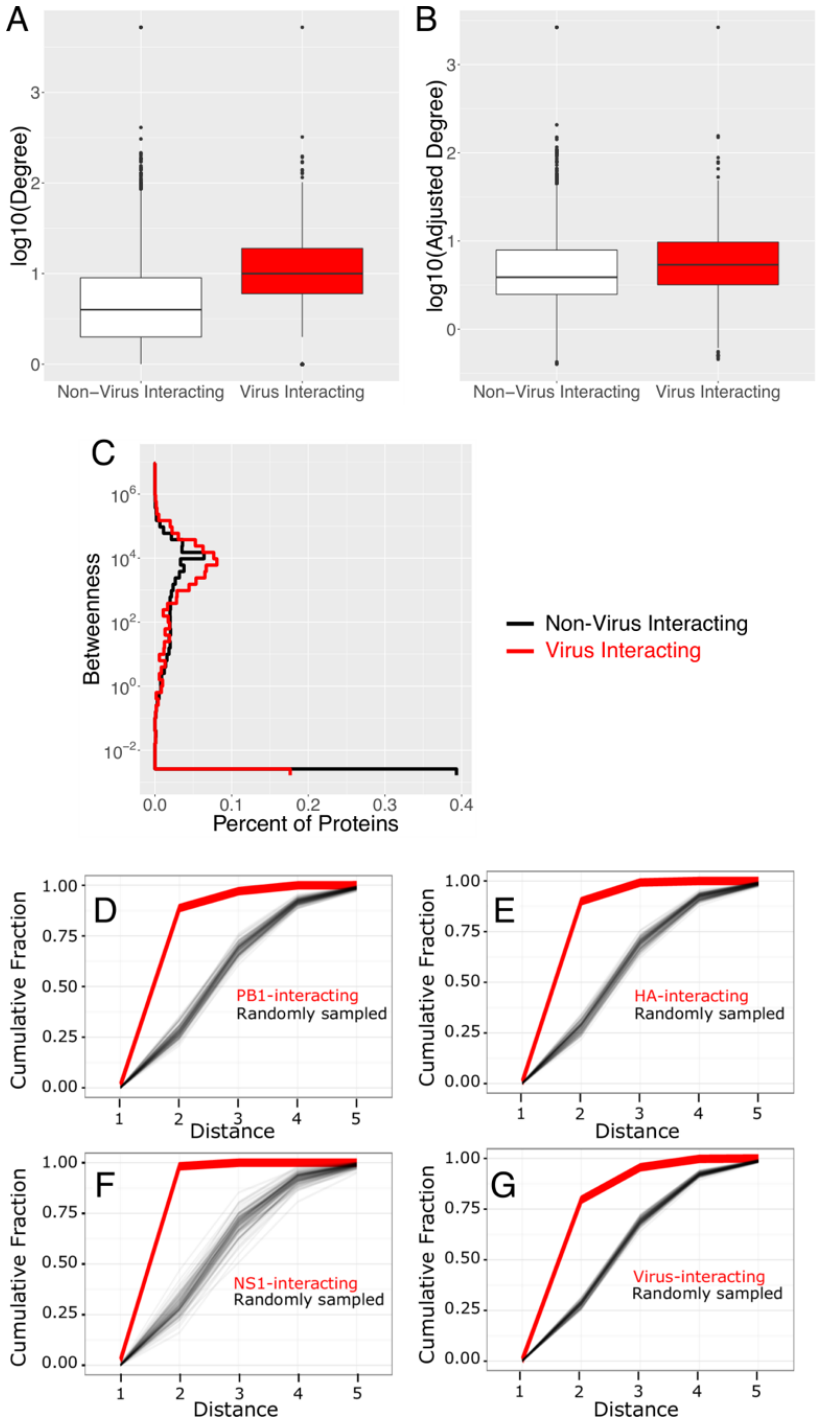


Figure 2.3. The network topological characteristics of virus interacting host proteins. (A to C) Distributions of the degree (A), adjusted degree (B), and betweenness (C) of virus interacting proteins and all proteins in the human PPI network. An ϵ of 0.01 was added to the betweenness to facilitate log scaling. (D to G) The cumulative distributions (thick red lines) of the shortest distances connecting host proteins in the PPI network that interact with viral PB1 (D), HA (E), or NS1 (F) protein or the set of all viral proteins (G). For a control, the cumulative distribution of distances was iteratively determined (N=100) by randomly sampled host proteins in the PPI network (thin black lines). The number of proteins sampled on each iteration was equal to the number of interacting host proteins of each virus protein (or set of viral proteins).

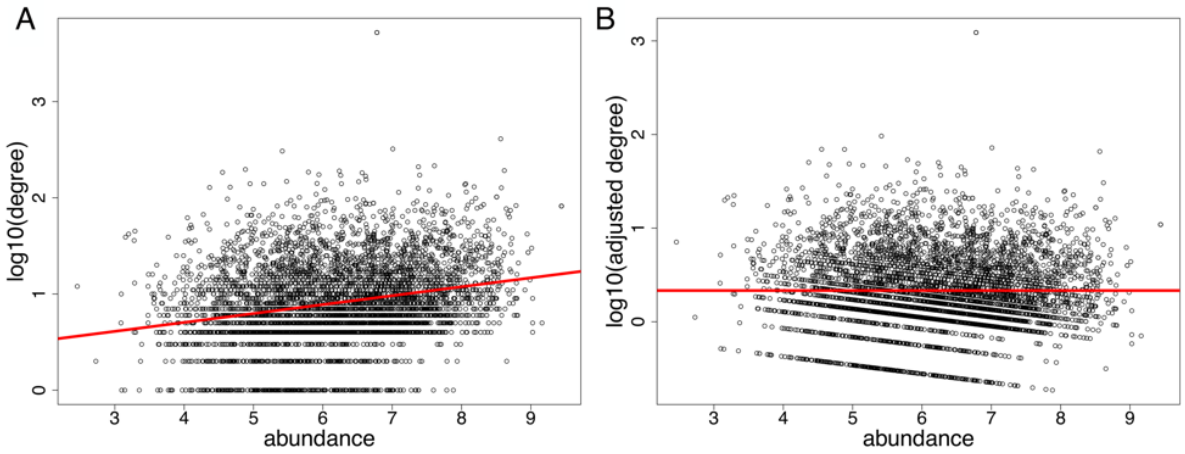


Figure 2.4. Protein abundance versus protein degree for the total network's original degree (A) and adjusted degree (B). Each plot is fit with a linear model in red, demonstrating that the correlation found in the original degree is not present after value adjustment.

The tendency for virus proteins to bind host proteins that had a higher degree and betweenness was consistent when analyzing the interaction partners of each virus protein separately (Pairwise t-test of the log-scaled data: all P values < 0.01 except betweenness of NS2-interacting proteins that was not significantly distinct from the betweenness of the total network). This indicates that influenza virus proteins selectively interact with host proteins that strongly regulate cellular behavior. These results are consistent with published findings for HCV and dengue virus [144, 145] and with a previous study that used a yeast two-hybrid approach to identify influenza virus interacting host proteins for 10 of the 11 virus proteins [127]. Further, these are characteristics that generalize to each virus protein's interacting partner; suggesting that all 11 virus proteins have a role in manipulating cellular machinery.

2.3.2 Virus Interacting Host Proteins are Closely Connected in the Human PPI Network

A previous study suggests that host factors of viral replication are closely clustered within the network but did not assess the topological characteristics of virus interacting host proteins [135]. Functionally related proteins are often observed to be closely clustered in PPI networks [146, 147]. Knowing that influenza virus proteins manipulate multiple host cell functions to

promote replication, these previous studies suggest that the interaction partners of viral proteins should be closely clustered by host function. If this is true, neighboring cluster proteins could serve as possible alternative mechanisms for influenza virus to manipulate necessary host functions.

Clustering analysis finds that virus interacting host proteins are very significantly clustered within the PPI network. The set of proteins that interact with each viral protein are significantly more closely clustered in the network than expected by chance (Figure 2.3D to G, P value < 0.01 comparing the median distance of the virus interacting proteins to the median distance of randomly sampled proteins). Generally, 25% of the randomly sampled proteins are connected by two or fewer interactions, while 88.7% of PB1-interacting proteins, 90.0% of HA-interacting proteins, 98.2% of NS1-interacting proteins, and 79.6% of all host proteins that interact with any influenza virus protein are connected by two or fewer interactions. Collectively, these results support that viral proteins are selectively targeting closely clustered host proteins. Additionally, 1.5% of all virus interacting protein pairs are involved in a complex, whereas only 0.066% of all protein pairs in the PPI are involved in a complex. In sum, influenza virus proteins are closely clustered and 22.4 times more likely to be involved in a protein complex than randomly selected proteins.

2.3.3 Constructing the Influenza A Virus Subnetwork

Network analysis of virus interacting host proteins demonstrates that viral proteins preferentially interact with closely connected host proteins that are in positions central to information flow across the human PPI network, suggesting that it may be possible to exploit network positions to prioritize potential antiviral drug targets. We hypothesized that there exists a disease-specific subnetwork that is likely to be significantly enriched for host factors consisting of pathways that connect virus interacting proteins to key cellular machinery. We further hypothesized that the topology of proteins within this subnetwork may provide an additional advantage in identifying host factors of disease.

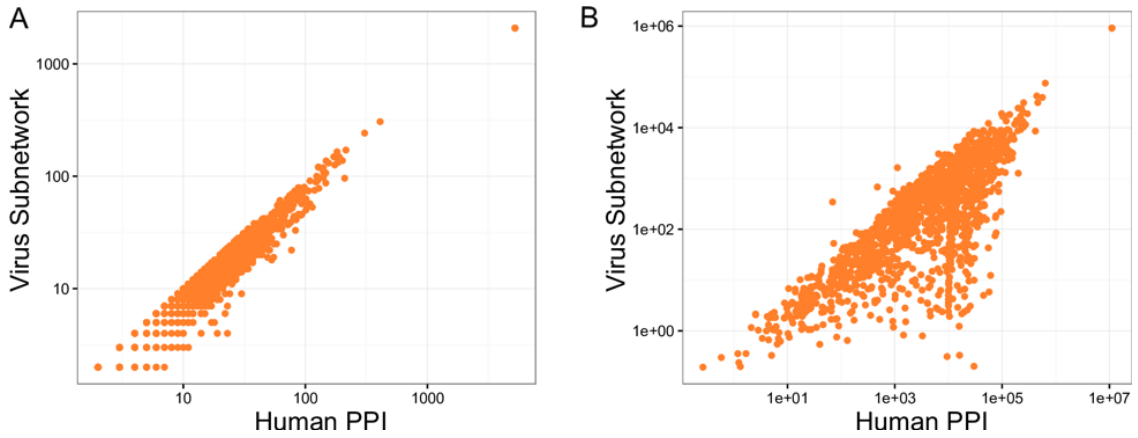


Figure 2.5. Network characteristics of the virus subnetwork. Panels A and B compare the degree and betweenness, respectively, of the connecting proteins in the whole PPI network and the virus subnetwork.

As a result of how the subnetwork is constructed, the mean degree of the virus interacting proteins and the internal-essential proteins were lower than the mean degree of the connecting proteins (ANOVA followed by Tukey post hoc analysis P value: < 0.01). While the degree of connecting proteins does not shift significantly between the total PPI network and the virus subnetwork (Figure 2.5A), some proteins with low betweenness have much lower betweenness in the virus subnetwork compared to the total PPI network (Figure 2.5B). Higher betweenness nodes in the total PPI network do not demonstrate dramatic shifts in the virus subnetwork upon comparison. This shift between the total network and virus subnetwork may reveal proteins that are more or less critical to virus replication that cannot be identified in a standard topological analysis of a PPI network.

2.3.4 Functional Enrichment Analysis of the Influenza Virus-Host Subnetwork

Previous work [131] included a per protein analysis of virus interacting proteins that identified involvement in several stages of the life cycle of influenza virus, particularly in viral replication tasks and export of influenza vRNPs from the nucleus. Here, analysis found that virus interacting host proteins and connecting proteins within the virus subnetwork are functionally distinct (see Tables 2.1 and 2.2 for abbreviated results). Analysis of virus interacting proteins replicated the previous finding that virus interacting

host proteins are primarily associated with housekeeping and viral replication processes [131], whereas analysis of connecting proteins shows association with protein phosphorylation, histone reconfiguration, and immune responses. Specifically, the immune response pathways identified are the stimulatory C-type lectin receptor signaling, T-cell receptor signaling, and Fc epsilon receptor signaling, all of which regulate NF-kB activity. These results suggest that the virus subnetwork contains functional information generally unobserved when considering virus interacting host proteins or internal-essential proteins in isolation.

Table 2.1. Functional enrichment analysis of connecting proteins within the virus subnetwork. Proteins were analyzed using DAVID.

Cluster	**Number of GO Terms**	**Enrichment Score**
Transcription	4	55.4
DNA damage/repair	3	19.2
Protein phosphorylation	19	18.7
Mitosis	5	18.7
Histone reconfiguration	42	14.4
Immune response (C-type lectin receptor signaling pathway, T cell receptor signaling pathway)	3	14.0
Zinc ion binding	4	11.5

Table 2.2. Functional enrichment analysis of virus interacting proteins within the virus subnetwork. Proteins were analyzed using DAVID.

Cluster	**Number of GO Terms**	**Enrichment Score**
Ribonucleoprotein/Viral transcription	13	67.2
Cell-cell adhesion	3	45.0
mRNA splicing	9	41.8
Nucleotide binding	10	30.3
Chaperone/UPR	3	22.1
Viral nucleocapsid	3	19.0
mRNA nuclear export	4	17.5
Nucleotide binding/ATP binding	5	17.3
Translation initiation factors	11	13.2
Proteasome/NF-kB MAPK signaling	23	12.1

2.3.5 Connecting Proteins of the Influenza Virus-Host Subnetwork are More Enriched for Host Factors than Virus Interacting Proteins

To evaluate the significance of the connecting proteins, the observed hit rate from the secondary siRNA screen was compared to the hit rate resulting from several other screens: 1.) the 1,292 virus interacting host proteins in HEK293 cells (hit rate = 299/1,292=0.23) [131], 2.) the 264 host factors in the study by Karlas et al. (71/264 = 0.27) [136], and 3.) a genome-wide screen for influenza virus host factors in A549 cells (287/22,843=0.013) [136] which can be interpreted as the expected hit rate when randomly sampling the PPI network.

The secondary siRNA screen found that connecting proteins were significantly enriched for host factors of influenza virus replication, but demonstrate no statistically significant advantage in selecting proteins by betweenness (Figure 2.6). Of the 78 proteins targeted in the siRNA screen of connecting proteins, a total of 27 significantly reduced viral titers by at least two orders of magnitude, with 15 categorized as high-betweenness proteins and 12

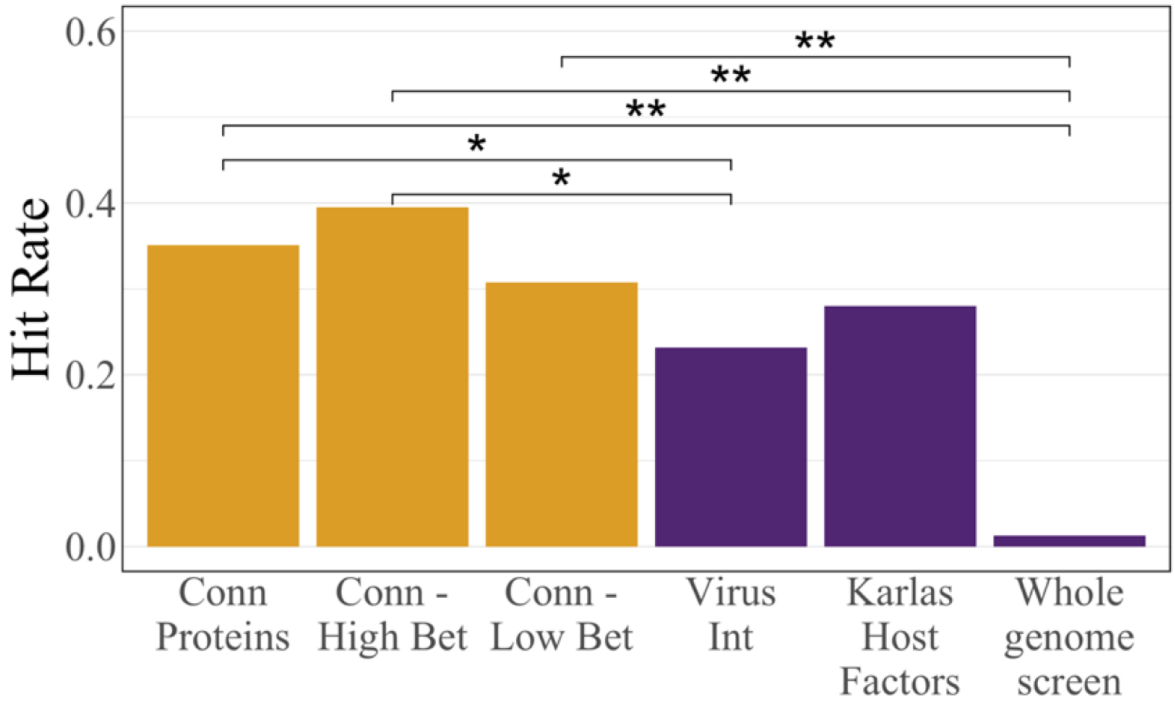


Figure 2.6. Hit rates for all tested connecting proteins and the 39 connecting proteins with highest and lowest betweenness in the virus subnetwork. Hit rates are compared to those observed in a previous screen of virus interacting host proteins [131], from our secondary screen of host factors identified by Karlas et al., and from a genome-wide screen [136]. *p $<$ 0.05 and **p $<$ 0.01.

categorized as low-betweenness proteins. Note that one of the 39 high-betweenness proteins (PLK1) was eliminated from the calculation because both respective siRNAs were cytotoxic. Differences between hit rates were compared using Pearson's chi-squared test. The hit rate of connecting proteins (27/77=0.35) was significantly higher than the hit rate observed in the screen of virus interacting proteins (P value: 0.024) and in the whole-genome screen (P value: $< 2.2x10^{-16}$) but not significantly distinct from the rate observed in rescreening the Karlas host factors (P value: 0.21). When considering the connecting proteins based on their betweenness, the high-betweenness connecting proteins had a hit rate of 0.39 (15/38) which was significantly higher than the hit rates observed in the virus interacting and whole-genome screens (P value: 0.032 and P < 2.2 x 10^{-16}, respectively). The high-betweenness protein hit rate was higher than the rate observed in the screen of host factors by Karlas et al. (33), but not significantly (P value: 0.16). The low-betweenness connecting protein hit rate was lower than that of the high-betweenness connecting proteins (12/39=0.31). The difference in hit rates between high- and low-betweenness proteins was not significant (P value: 0.57). In all, the screening results suggest that proteins connecting virus interacting proteins to host factors of influenza virus replication are highly enriched for host factors themselves-significantly more so than proteins that directly interact with virus proteins. However, the topological information from betweenness does not significantly improve host factor identification.

2.3.6 The Influenza Virus Subnetwork is Enriched for Host Factors Identified in Six Host Factor Screens

To determine whether host factors identified in additional, previous screens are enriched within the virus subnetwork, we compiled a list of host factors of influenza virus replication identified in at least one of six previous screens [136, 148, 149, 150, 151, 152]. A Fisher exact test for enrichment was used to determine whether the connecting proteins or the set of influenza virus interacting proteins are enriched with host factors identified in these studies relative to the abundance of host factors within the total PPI network. Both connecting proteins and the virus interacting proteins are significantly enriched for host factors (P

value: 7.2 x 10^{-5} and 1.1 x 10^{-5}, respectively; odds ratio: 1.4 and 1.5, respectively). There is no significant difference in the enrichment of host factors between connecting proteins and virus interacting proteins (P value: 0.48; odds ratio: 0.92). To ensure the host factors identified in the study by Karlas et al. [136] were not creating bias in the enrichment result, the enrichment analysis was repeated using host factors identified in all studies except Karlas. Again, connecting proteins and virus interacting proteins are significantly enriched for host factors (P value: 1.8 x 10^{-6} and < 3.2 x 10^{-3}, respectively; odds ratio: 1.5 and 1.34, respectively), and no significant difference in the enrichment of host factors between connecting proteins and virus interacting proteins was found (P value: 0.49).

2.4 Summary

Biological network approaches have demonstrated their potential impact on health-related research, including gene/protein characterization, and drug design and side effects [114, 118, 119, 101, 140, 153], yet demonstrations that network information can inform drug target discovery are still limited. The completion of this aim marks the first confirmation that virus and host protein interaction data can be integrated to improve large-scale drug target discovery (specifically antiviral target discovery) and reveal additional insights into virus-host interactions. The positions of virus interacting proteins within the human PPI network suggest that influenza A virus has evolved to interact with proteins that influence network behavior, regardless of the previously unexplored effects of abundance-degree biases in PPI data generation. The virus-specific subnetwork reveals that there is a set of proteins with moderately high betweenness in the total network yet low betweenness in the subnetwork. While these proteins may be of high importance to the total network, this result suggests that they may be less important to the progression of influenza infection than proteins that are of high betweenness to both the total network and virus-specific subnetwork. In this way, the novel subnetwork construction provides further insight when determining important host factors of virus replication.

Virus interacting proteins are closely clustered in the network which may be the result of attempts to manipulate specific biological functions as proteins with shared biological functions tend to cluster in PPI networks [154]. This may signify that influenza virus has parallel available pathways with which to engage with host biological functions. Additionally, protein complex evidence suggests that high degree and high clustering of influenza A virus proteins may be due to their involvement in complexes. From a network viewpoint, it is likely that high incidence of clustering within the PPI network is a result of both the high betweenness and degree of the virus interacting protein group as a whole. Previous studies have found that host factors of virus replication (not necessarily virus interacting host proteins) have been observed to cluster within the PPI network [135]. Further analysis on network clustering host factors of interest is needed to determine whether these two observations are independent of one another.

Functional enrichment analysis of the subnetwork reinforces that while virus interacting proteins are associated with virus replication processes, proteins within the constructed subnetwork are associated with immune response to viral infection. Results for virus interacting proteins largely build on the per protein discussion of virus-host interactions found in previous work [131], identifying involvement in several stages of the viral replication cycle. The functional enrichment analysis of connecting proteins reveals high levels of involvement in the immune response to viral infection, specifically in NF-kB regulating pathways such as stimulatory C-type lectin receptor signaling, T-cell receptor signaling, and Fc epsilon receptor signaling. Influenza virus is known to manipulate host immune response pathways (specifically NF-kB regulating pathways) to promote viral replication [155, 156]. Because previous virus-host PPI network analyses have not studied these connecting proteins as a separate population, their importance to the biology and regulation of the system has been overlooked. The subnetwork construction approach applied in this work isolates additional host biological processes essential to the regulation of virus replication, further demonstrated by siRNA screening results of the connecting proteins. Together, the results suggest that future work in virus-host protein networks can leverage the technique presented here to identify subnetworks with increased biological

relevance to the analyzed phenotypes/conditions and increase predictive power for the purposes of drug discovery.

The conclusion that host-virus interaction data can be leveraged to improve virus replication host factor discovery is unlikely to be affected by off-target concerns associated with siRNA screens. Though off-target effects often challenge siRNA studies, changes to experimental protocols (such as requiring multiple siRNA hits per targeted gene or changing siRNA concentrations) can only moderately improve false-positive rates [157, 158, 159]. The protocol used to generate the data used in this study was not optimal to ensure the characterization of any one targeted gene. As such, the hit rates of gene groups are compared. Protocols between these experiments and those used for comparisons are either identical [131] or very similar [136], suggesting that off-target rates across the tested groups are unlikely to explain the differences in observed hit rates.

The variability and incompleteness of PPI data as well as the limited agreement between influenza virus replication screens are well-known concerns for network-based drug target discovery. False discovery of virus-host interactions and the possibility that virus-host interaction data are skewed toward well-studied networks could also have an effect on the clustering result in virus interacting proteins. However, the enrichment of host proteins important for influenza virus replication within the constructed virus subnetwork demonstrates that even with these possible shortcomings, PPI network analyses have the power to identify important host factors for influenza virus replication. The antiviral protein candidate screen performed in this study further supports the use of PPI data during candidate prioritization with significant hit rates against virus interacting proteins and randomly targeted proteins.

Overall, this PPI network-based study provides insight into the network characteristics of disease relevant proteins beyond those captured directly in virus-host interactions and supports the idea that influenza virus interacts with host proteins in dominant network positions in order to maximally manipulate host cells. Protein predictions from this novel method should be used to guide future experimental design in drug development efforts.

3.0 Aim 2: Network Controllability for Drug Discovery and Repurposing

3.1 Introduction

The key to antiviral drug development lies in identifying key processes in the viral life cycle that can be suppressed to limit infection spread. For example, Oseltamivir and Zanamivir are neuraminidase inhibitors that halt the release of newly manufactured viral material from the host cell [160]. Currently, the FDA recommends the use of four antiviral treatments for the flu, having retired two additional drugs due to drug resistance observed in most recent strains [17]. While antiviral drug development can always benefit from innovation and efficiency, rapid and effective development is of even greater need in the event of emergent viruses and pandemic spread. As COVID-19 spreads worldwide with 179 million cases and 3.87 million deaths occurring between January 2020 and June 2021 [161], there is an urgent need for novel treatment options. There are currently few pharmaceutical treatments for SARS-CoV-2 infection under review [162]. One strategy to accelerate the identification of possible leads is to reposition drugs with known targets and mechanisms that may have been through parts of the FDA approval process [163]. Avoiding this development pipeline known for its low success rate [164] advantageously saves invaluable time and monetary cost. While this is ultimately the fastest way to get treatments to patients in need, the most efficient way to discover drugs with the potential for repurposing is unclear.

In the time since the beginning of 2020, many attempts to predict candidate drugs for repositioning have been made. Given the novel nature of the virus, methods of target prediction have been forced to utilize the limited data that is available or creatively repurpose data from related coronaviruses. *In vitro* screenings of chemical libraries have been used to identify inhibitors of SARS-CoV-2 replication [165, 166] and cellular toxicity [167]. Screenings of experimentally verified SARS-CoV-2 interacting host proteins [168] have elucidated key infection mechanisms which, when compared to drug databases, have predicted a range of possible targets for repurposing. Network analyses using protein

interaction data from up to 13 related human coronaviruses [169, 170] combined with *in vitro* screenings have identified additional sets of cellular pathways to consider for drug repurposing. Topology of protein interactions and drug-gene interactions combined with differential expression and pathway analysis has been used to identify possible mechanisms of action for SARS-CoV-2 infection [171]. With each method integrating varying levels of biological detail, overlap between studies is optimal for ensuring infection-specific relevance and effectiveness.

There is precedent for network studies of many common viruses including hepatitis C [128, 144], severe acute respiratory syndrome (SARS) [172, 173], Human immunodeficiency virus (HIV) [173, 174, 175, 176, 177], and influenza virus [128, 123, 68, 178, 69]. Past work studying the effects of influenza virus in PPI networks has focused on identifying host factors involved in virus replication and improving the prediction of drug targets but often focuses on the analysis of basic topological measurements. While this provides a general overview of the state of the network, it is a static snapshot of the cell and, therefore, fails to capture the dynamic nature of the cell's pathways. To identify how these systems can be manipulated and exploited to manage biological properties, a dynamic understanding of disease progression is required.

Controllability, a concept from classic controls engineering, states that a system is controllable if it can be driven to any final state within state space in finite time given an appropriate external input [179]. This is commonly applied to linear, time invariant dynamic systems:

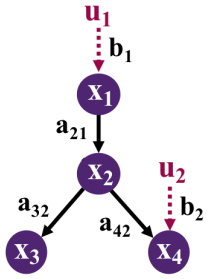
$$\frac{dx(t)}{dt} = Ax(t) + Bu(t) \tag{3.1}$$

where A is an NxN matrix of state coefficients that describes how N molecule states, $x(t)$, interact within the system and B is a matrix of input weights describing how external influences, $u(t)$, impact the system. In general, a system is controllable if the controllability matrix,

$$C = [B, AB, A^2B, ..., A^{N-1}B] \tag{3.2}$$

is full rank, N. This means that the system can be manipulated to reach any desired combination of states within all of state space following the defined input, B. In total, a

a) Classic controllability



$$\dot{x} = Ax + Bu$$

$$\dot{x} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ a_{21} & 0 & 0 & 0 \\ 0 & a_{32} & 0 & 0 \\ 0 & a_{42} & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b_1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & b_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

where:

A = Interaction weight matrix
x = Protein concentation state matrix
B = Input process weight matrix
u = Protein translation process input matrix

Controllability matrix (C):

$$C = [B \quad AB \quad A^2B]$$

$$C = \begin{bmatrix} b_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{21}b_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{21}a_{32}b_1 & 0 \\ 0 & b_2 & 0 & 0 & a_{21}a_{42}b_1 & 0 \end{bmatrix}$$

Rank (C) = 4
Full rank
∴ *System is fully controlled*

Figure 3.1. An example protein-protein interaction network with three proteins and two protein translation process inputs. The state space representation demonstrates that the change in state of a protein's concentration is a function of its current state and an input process. A classic controllability analysis demonstrates that this system is fully controllable and could, therefore, be driven to any possible state change in every protein.

controllability analysis identifies the key components of a system that must be manipulated to drive desired system outcomes [180].

An example PPI network in Figure 3.1 is transformed into its state space matrix representation. With the inclusion of two independent inputs ($u1$ and $u2$), the controllability matrix is full rank. Therefore, the system is fully controllable and it is possible to drive the protein concentrations to any desired state. Applying the idea of controllability to a cell at the onset of infection, a virus aims to control cellular functions (the system of proteins), promote virus replication tasks, and reach a final, infected cell state. Theoretically, controllability can be used as a tool to identify the proteins of the traditional B matrix that the virus must control. While it would be advantageous to interpret the infection from this control perspective, mathematical limits are quickly reached as a result of the system size, preventing the direct application of traditional controllability methods to PPI networks.

Advances in network theory have created alternative methods of network controllability evaluation that survey each node's (protein's) importance in the ability of an external set of inputs to fully control the network. Controllability classification is founded in "driver node" calculations: identifying the network components that must be manipulated for the system to be fully controlled (analogous to determining the non-zero elements of the B matrix in classic controllability). Without manipulation, driver nodes will remain unaffected

by changes to the rest of the system, rendering the total system uncontrollable. Driver nodes are identified using the Hopcroft-Karp algorithm [181] which can be applied to any directed graph in bipartite form. This method calculates the maximum matching of the graph, or, the largest set of network paths where no node is shared by two edges. Because each node can only influence one of its interactors, the identification of these paths dictates the way in which control can propagate through the network. The nodes that are not included in these paths or at the start of these paths are not receiving control from a neighboring node and, therefore, require "driving". A set of driver nodes (size N_D) that is capable of controlling the total network is called a minimum input set (MIS). The MIS is not unique and the number of possible MISs scales exponentially with the size of the network [182]. After a primary MIS is calculated, two methods of controllability node classification can be used.

In robust controllability (by Liu et al. [116]), the MIS is re-calculated (size N'_D) after removing each node from the network. The node is then classified by its effect on the manipulation required to control the network, where an increase in the size of the MIS makes it more difficult to control the network and a decrease in the size of the MIS makes it easier to control the network. This method has previously been applied to many network types such as gene regulatory networks, food webs, citation networks, and PPI networks to better understand what drives the dynamics of each system [177, 116]. While it is useful to observe the structural changes to the network after the removal of singular nodes, this method only considers one possible MIS. A second global controllability method by Jia et al. [182] classifies a node by its role across all possible MISs, placing each node in the broader context of all possible control configurations.

A comparison of the controllability of the human protein-protein interaction network (Host Interaction Network, HIN) and the human network with the addition of influenza A virus or SARS-CoV-2-host protein interactions (Virus Integrated Network, VIN) can be used to identify proteins with unique post-infection roles in driving total cell behavior. Assuming the identified differences are representative of biological changes within the cell (such as changes to gene regulation), the protein predictions have potential as virus-specific drug targets. Here, proteins are identified by topological, controllability, and biological relevance to both influenza A and SARS-CoV-2 infection and recommended for target

prioritization based on previous druggability and relevance to functions such as translation, cellular transport, and the immune response.

3.2 Materials and Methods

3.2.1 Protein-Protein Interaction Network Construction

The directed human protein-protein interaction network was published by Vinayagam et al. [183]. The network was restricted to interactions with a confidence level greater than 0.7 based on the correlation between confidence scores and biological relevance as discussed by Yu and Finley [184]. After construction, the Host Interaction Network (HIN) contains 6,281 proteins and 31,079 interactions. For influenza A virus studies, virus-host interactions from Watanabe et al. [131] were narrowed to 2,592 directed interactions from 11 influenza A virus proteins (HA, M1, M2, NA, NP, NS1, NS2, PA, PB1, PB2, and PB1-F2 proteins) to 752 "virus interacting proteins" preexisting in the HIN. After integration into the HIN, the network contains 6,292 proteins and 33,671 interactions. This network is referred to as the "Virus Integrated Network" (VIN) in the influenza A study.

For SARS-CoV-2, host proteins identified in the SARS-CoV-2-host interactions from Gordon et al. [168] were referenced against the host network. To construct the VIN, 23 of the 27 SARS-CoV-2 proteins tested were added to the network along with 152 interactions with 148 existing host proteins. Four SARS-CoV-2 proteins, spike, nsp11, ORF3b, and ORF7a, had no known interactions with host proteins of the HIN and were omitted from the analysis. In total, the VIN contains 6,304 proteins and 31,231 interactions. All network construction and topological analysis was completed in R 3.6.1 using the igraph package.

3.2.2 Robust Controllability Classification

Methods for robust controllability are sourced from Liu et al.'s work [116]. An example is found in Figure 3.2. For any network with n total nodes, a subset, N_D, of driver nodes is found using a maximum matching algorithm such as Hopcroft-Karp on the bipartite

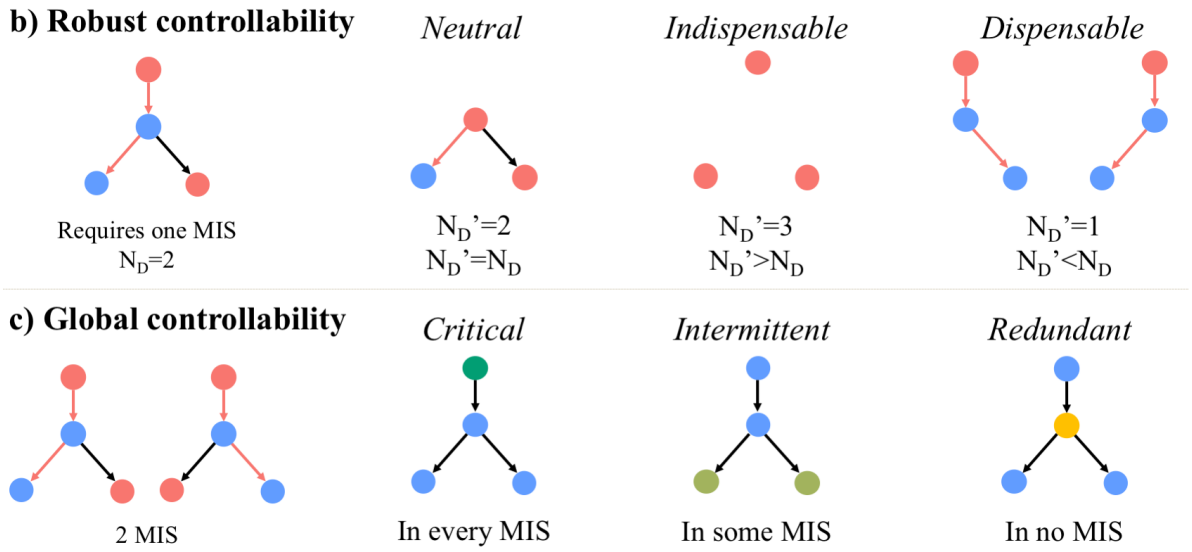


Figure 3.2. Application of robust controllability to determine the robustness of the network after the removal of a protein and global controllability to assess the importance of a protein to all methods of network control. Both methods rely on the calculation of a minimum input set (MIS) of driver nodes which represent the proteins which require an input to be driven to a final state.

representation of the total network [181]. Each node of the network is removed, the driver set is re-calculated (size N_D'), and the removed node is classified by its effect on the changes to the size of the driver set. Increasing the number of driver nodes ($N_D' > N_D$) makes it more difficult to control the network (these nodes are classified as indispensable nodes) and decreasing the number of driver nodes ($N_D' < N_D$) makes it easier to control the network (these nodes are classified as dispensable nodes). A removed node with no effect on the number of driver nodes ($N_D'=N_D$) is classified as a neutral node.

3.2.3 Global Controllability Classification

Calculations for global classification were adopted from Jia et al. [182]. An example is found in Figure 3.2. For any network with n total nodes, a subset, N_D, of driver nodes is found using a maximum matching algorithm such as Hopcroft-Karp on the bipartite representation of the total network [181]. Control adjacent nodes of all N_D are identified iteratively and used to create an input graph as described in Zhang et al. [185]. Global

controllability classifies each node by its membership to all possible minimum input sets (MISs) of the network. Critical nodes are included in all of the network's possible MISs, intermittent nodes are only included in some of the possible MISs, and redundant nodes are not included in any of the possible MISs.

3.2.4 Prediction Validation

A test is needed to ensure that resulting predictions are meaningful and not simply a consequence of network structure. 10,000 random protein sets the size of the number of virus interacting proteins (752 for influenza A virus, 152 for SARS-CoV-2) were pulled from the host proteins of the VIN. The degree and betweenness as well as the controllability classifications of the pseudo-virus interacting proteins were compared to the true values of the virus interacting proteins, driver proteins, and all proteins. Statistical significance was determined with a Mann-Whitney U test.

Controllability-predicted proteins were cross-referenced with Interferome v2.01 [186] to determine which genes were experimentally identified as interferon regulated genes (IRGs), inductors of a fold change in expression greater than two in interferon knockdown studies. For influenza virus, an additional validation was performed by comparing controllability predicted proteins to 6 previous siRNA screens for host factors of influenza A virus replication [150, 135, 136, 152, 149, 131]. Similar analysis could not be performed for SARS-CoV-2 as this study occurred during the COVID-19 pandemic four months after the emergence of the virus and siRNA screenings had yet to be performed. Ingenuity Pathway Analysis (IPA) was also used to analyze the functional networks of the predictions for influenza [187], but could not be completed for SARS-CoV-2 as IPA is proprietary software and could not be accessed in the office during the SARS-CoV-2 global pandemic. For SARS-CoV-2, predicted candidates were prioritized by cross-referencing Drugbank [188] to identify targets with known interactions with drugs that have completed some or all of the FDA pipeline.

3.3 Influenza A

3.3.1 Results

3.3.1.1 Addition of Virus Interactions to Host Network has Wide Reaching Effect

As expected, only the 752 virus interacting proteins display an alteration in degree after the addition of virus interactions to the network (Marked in blue in Figure 3.3a). This shift is significant for the group of virus interacting proteins as compared to all proteins in both the VIN (log scaled median of virus interacting proteins: 1.04; log scaled median of all proteins: 0.70; student t-test of log scaled data P value $< 2.2 \times 10^{-16}$) and the HIN (log scaled median of virus interacting proteins: 0.85; log scaled median of all proteins: 0.70; Student t-test of log scaled data P value: 5.97×10^{-12}). The degree distributions of both networks are scale free (Figure A1).

Because betweenness is sensitive to the information flow through all proteins instead of only neighboring proteins, 2,735 proteins exhibit an increase in betweenness after the addition of influenza virus interactions. Of these proteins, 207 proteins' log betweenness exhibits an increase of 2 or more in the VIN compared to the HIN (Figure 3.3b). This suggests that

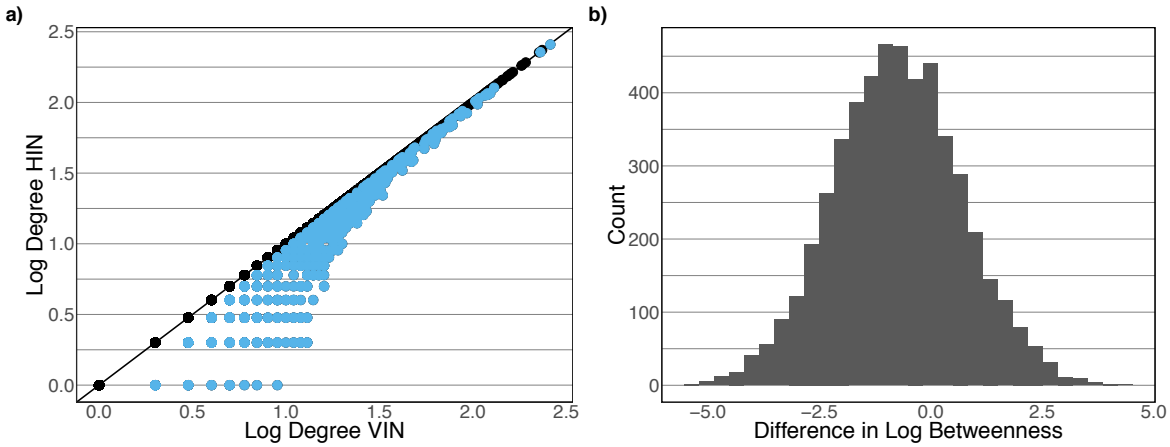


Figure 3.3. a) Degree of the VIN vs degree of the HIN where the virus interacting proteins are marked in blue. The degree distributions of the networks are scale free. b) Difference in betweenness between the VIN and HIN for proteins that exhibit a difference greater than one.

the addition of influenza virus interactions has an effect on network topology that reaches over 3.5 times the number of host proteins that are directly interacting with influenza virus proteins. The betweenness shift in the group of virus interacting host proteins is significant as compared to all proteins in both the VIN (Log scaled median of virus interacting proteins: 3.23; Log scaled median of all proteins: 2.82; Student t-test of log scaled data P value $<$ 2.2 x 10^{-16}) and the HIN (Log scaled median of virus interacting proteins 3.22; Log scaled median of all proteins: 2.82; Student t-test of log scaled data P value: 2.13 x 10^{-15}. This is a result of directed interactions from viral to host proteins, making virus interacting proteins responsible for information flow to the rest of the network.

Driver proteins (nodes) are the foundation of both types of controllability calculations, representing the protein set that must be manipulated for the system to be fully controlled. The proteins are identified through maximum matching algorithms [181]. The HIN and VIN both require N_D = 2,463 driver proteins to achieve full network controllability, suggesting that the magnitude of network control is unchanged by the influence of the influenza virus interactions. However, the identity of driver proteins shifts slightly as the 11 viral proteins replace 11 host proteins within the primary MIS as drivers in the VIN. Table 3.1 lists the identities of the 11 host proteins along with the shortest distance to an influenza virus protein in the network, degree, and betweenness. Of these 11 proteins, only five are directly interacting with influenza virus proteins. One of the remaining proteins is two steps (two interactions and one connecting protein) from any influenza virus protein, and the remaining five proteins are three steps from any influenza virus protein. The number of paths between viral proteins and these proteins are reflective of the number of paths between viral proteins and all host proteins (Fisher test P value: 0.99). This supports the idea that viral interactions have lasting effects on the system's control structure, affecting proteins that are multiple paths away.

Table 3.1. Identities of the proteins that are drivers in the HIN but not the VIN with the shortest number of paths to an Influenza A viral protein. Degree and betweenness of the proteins of the VIN is provided (with the values from the HIN in parenthesis). Only 45% of these proteins are directly interacting with the viral proteins, demonstrating the cascade effect caused by the inclusion of viral interactions

Entrez ID	**Gene Name**	**Shortest Dist. to Virus**	**Deg.**	**Bet.**
10658	CUGBP, Elav-Like Family Member 1 (CELF1)	1	4 (4)	81 (81)
1969	EPH Receptor A2 (EPHA2)	1	14 (13)	93 (0)
6733	SRSF Protein Kinase 2 (SRPK2)	1	6 (2)	6023 (6023)
10318	TNFAIP3 Interacting Protein 1 (TNIP1)	1	7 (7)	115 (115)
2997	Glycogen Synthase 1 (GYS1)	3	4 (4)	384 (384)
10949	Heterogeneous Nuclear Ribonucleoprotein A0 (HNRNPA0)	2	9 (2)	5 (0)
64112	Modulator of Apoptosis 1 (MOAP1)	1	8 (8)	6942 (6931)
10419	Protein Arginine Methyltransferase 5 (PRMT5)	3	26 (17)	6996 (4743)
10262	Splicing Factor 3b Subunit 4 (SF3B4)	3	13 (7)	82 (44)
23321	Tripartite Motif Containing 2 (TRIM2)	3	2 (2)	15 (15)
81603	Tripartite Motif Containing 8 (TRIM8)	3	3 (3)	0 (0)

Lastly, analysis finds that 8.9% of all driver proteins are also virus interacting proteins, meaning the intersection of the two protein groups of interest comprise only 3.5% of the total network. There is a significant increase in the betweenness of driver proteins depending on their status as virus interacting or influenza virus non-interacting proteins (Fisher test P value $< 2.2 \times 10^{-16}$) where there is no significant difference in degree of the same groups (Fisher test P value: 0.7161). This is further evidence that the addition of virus interactions to the network magnifies information flow through the proteins most involved in controlling network behavior.

3.3.1.2 Influenza A Virus Interacts with Proteins That Promote Cellular Control

Robust controllability was calculated for all proteins of the HIN and VIN (as shown in Table 3.2 with and without parentheses, respectively). The addition of influenza virus interactions to the network has no effect on the distribution of classifications of host proteins, and consequently, the virus interacting proteins. Upon entry to the VIN, the 11 influenza virus proteins are classified as neutral, meaning that removing these proteins does not alter the number of driver proteins required to control the VIN ($N_D = N'_D$). This reveals that the removal of singular proteins from the system is not enough to disturb the existing control structure under robust controllability.

While none of the proteins change robust classification between networks, the aforementioned replacement of 11 host driver proteins with viral proteins after the addition

Table 3.2. Robust controllability types of all proteins, driver proteins, and virus interacting proteins in the VIN (HIN in parenthesis).

	All Proteins	**Driver Proteins**	**Virus Interacting Proteins**
Indispensable	1,169 (1,169)	0 (0)	186 (186)
Neutral	2,669 (2,658)	803 (799)	312 (312)
Dispensable	2,454 (2,454)	1,660 (1,664)	254 (254)

of virus interactions creates a small change in robust type distribution for driver proteins. Of the displaced host proteins (deemed "robust proteins", found in Table 3.1), seven are neutral and four are dispensable in the HIN, meaning that their removal from the network does not change the number of driver proteins and reduces the number of driver proteins needed, respectively. All influenza virus proteins are classified as dispensable in the VIN. Of the five robust proteins that are both driver and virus interacting proteins, four are neutral and one is dispensable. The most notable change in degree and betweenness between the HIN and VIN is PRMT5, with an increase of 9 and 2250, respectively. Overall, robust controllability results suggest that the HIN is stable against potential changes in the control structure that could be caused by the addition of influenza virus interactions.

Robust type distributions were plotted against the classification results of virus interacting proteins, driver proteins, and all proteins in the VIN (Figure 3.4a-c). Randomly sampled protein sets closely resemble all proteins of the network, not the true interacting protein set, suggesting that robust controllability behavior of interacting proteins is not a coincidence of network construction (one-sided P value: 0.51, 0.49, and 0.50 for indispensable, neutral, and dispensable, respectively). Virus interacting proteins tend to be indispensable compared to the percentage of all proteins that are indispensable (Figure 3.4a). This suggests that viruses prefer to interact with proteins that are vital to cellular control. Driver proteins are very likely to be dispensable proteins compared to the percent of all proteins that are dispensable (Figure 3.4c). Further, the mean and median log degree and betweenness of the randomly sampled protein sets is significantly lower than the same measurements of the true virus interacting set (P value: $< 2.2 \times 10^{-16}$, $< 2.2 \times 10^{-16}$, Figure 3.5), signifying that virus interacting proteins are in positions of network significance. Overall, the robust controllability results of virus interacting proteins suggest that the virus may be selectively targeting host proteins based on controllability characteristics.

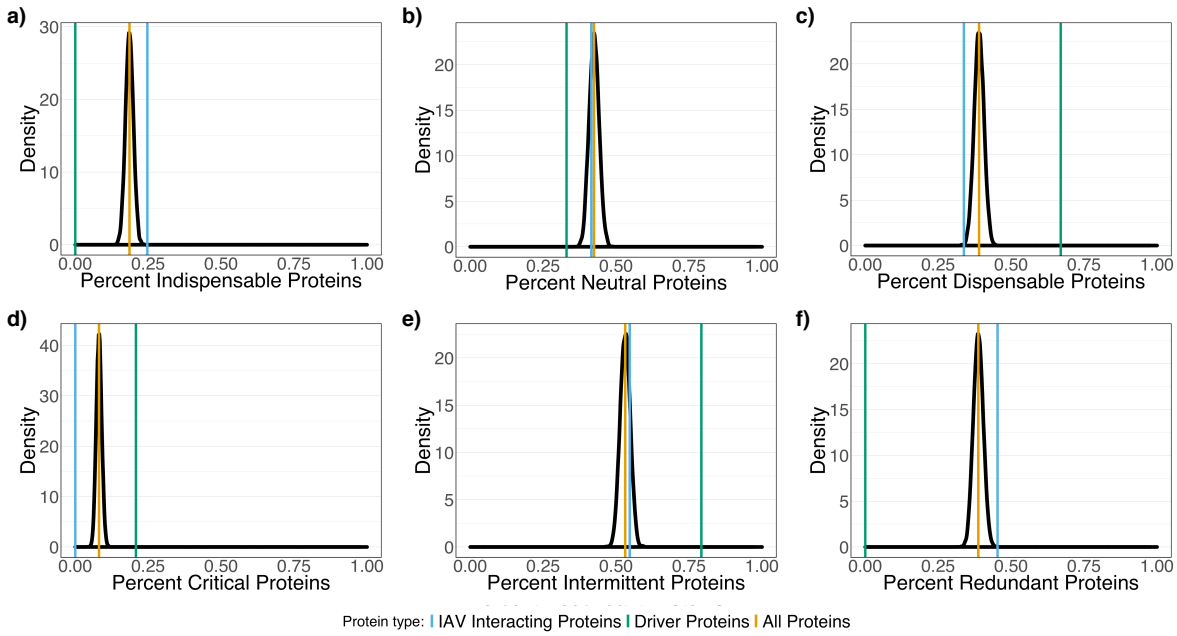


Figure 3.4. a-c) Density plots of distribution of robust controllability type for 10,000 random pulls of 752 proteins (number of virus interacting proteins in network). d-f) Density plots of distribution of global controllability type for 10,000 random pulls of 752 proteins (number of virus interacting proteins in network). Values for virus interacting proteins (blue), driver proteins (green), and all proteins (gold) are pictured for all figures.

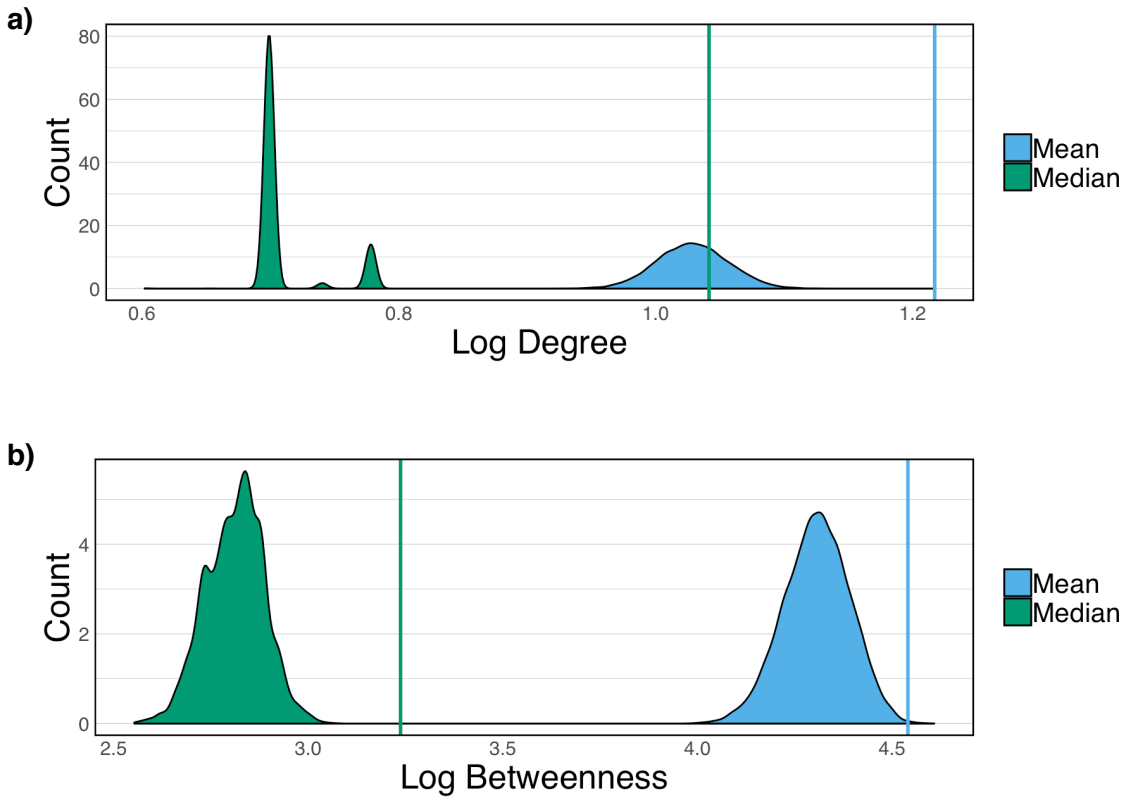


Figure 3.5. Density plots of a) mean (blue) and median (green) log degree of random virus interacting protein sets and b) mean (blue) and median (green) log betweenness of random virus interacting protein. Values for the true influenza virus interaction set shown as vertical lines, evidence that host proteins that directly interact with viral proteins are in positions of network significance.

3.3.1.3 Global Controllability Predicts Key Regulators of Influenza Infection

Global controllability was calculated for all proteins of the HIN and VIN (as shown in Table 3.3 with and without parentheses, respectively). Unlike in robust controllability, there is a small disturbance to global type distributions of host proteins after the addition of virus interactions. 24 host proteins shift from being classified as critical (a member of all MISs) to intermittent (a member of some MISs) proteins. Identities of these proteins (deemed "global proteins") can be found in Table 3.4 along with the shortest distance to an influenza virus protein in the network and protein degree and betweenness. The two most notable changes in degree and betweenness between the HIN and VIN are EPH receptor A2 (EPHA2) with an increase of 1 and 93, respectively, and transferrin receptor (TFRC), with an increase of 3 and 164, respectively. All 24 global proteins are driver and virus interacting proteins which, as mentioned, only comprises 3.5% of the total network. There are only two proteins (EPHA2 and HNRNPA0) that are also members of the robust protein set. 45% of virus interacting proteins are never drivers, suggesting that they are always manipulated by neighboring host proteins within any possible control configuration. Virus interacting proteins are not enriched for driver proteins (Fisher test P value: 0.14).

Table 3.3. Global controllability types of all proteins, driver proteins, and virus interacting proteins in the VIN (HIN in parenthesis).

	All Proteins	**Driver Proteins**	**Virus Interacting Proteins**
Critical	512 (525)	512 (525)	0 (24)
Intermittent	3,342 (3,318)	1,951 (1,938)	411 (387)
Redundant	2,438 (2,438)	0 (0)	341 (341)

Table 3.4. Identities of global Proteins (proteins that shift global classification between the HIN and VIN). All identified proteins are directly interacting with viral proteins. Degree and betweenness of the proteins of the VIN is provided (with the values from the HIN in parenthesis).

Entrez ID	Gene Name	Shortest Dist. to Virus	Deg.	Bet.
56655	DNA Polymerase Epsilon 4, Accessory Subunit (POLE4)	1	2 (1)	1 (0)
30846	EH Domain Containing 2 (EHD2)	1	3 (1)	1 (0)
1969	EPH Receptor A2 (EPHA2)	1	14 (13)	93 (0)
2665	GDP Dissociation Inhibitor 2 (GDI2)	1	3 (1)	2 (0)
51552	RAB14, Member RAS Oncogene Family (RAB14)	1	2 (1)	1 (0)
2091	Fibrillarin (FBL)	1	9 (4)	19 (0)
10949	Heterogeneous Nuclear Ribonucleoprotein A0 (HNRNPA0)	1	9 (2)	5 (0)
3032	Hydroxyacyl-Coa Dehydrogenase/3-Ketoacyl-Coa Thiolase/Enoyl-Coa Hydratase (Trifunctional Protein), Beta Subunit (HADHB)	1	9 (5)	26 (0)
3419	Isocitrate Dehydrogenase 3 (NAD(+)) Alpha (IDH3A)	1	3 (1)	2 (0)
4191	Malate Dehydrogenase 2 (MDH2)	1	3(1)	1 (0)
64949	Mitochondrial Ribosomal Protein S26 (MRPS26)	1	2 (1)	0 (0)
9180	Oncostatin M Receptor (OSMR)	1	6 (5)	18 (0)

Table 3.4. (Continued)

Entrez ID	Gene Name	Shortest Dist. to Virus	Deg.	Bet.
5052	Peroxiredoxin 1 (PRDX1)	1	11 (4)	44 (0)
5213	Phosphofructokinase, Muscle (PFKM)	1	6 (5)	17 (0)
26227	Phosphoglycerate Dehydrogenase (PHGDH)	1	4 (2)	9 (0)
5817	Poliovirus Receptor (PVR)	1	7 (6)	42 (0)
5686	Proteasome Subunit Alpha 5 (PSMA5)	1	6 (5)	11 (0)
5464	Pyrophosphatase (Inorganic) 1 (PPA1)	1	6 (5)	5 (0)
113174	Serum Amyloid A Like 1 (SAAL1)	1	2 (1)	1 (0)
6745	Signal Sequence Receptor Subunit 1 (SSR1)	1	4 (2)	12 (0)
7037	Transferrin Receptor (TFRC)	1	11 (8)	164 (0)
8834	Transmembrane Protein 11 (TMEM11)	1	4 (3)	20 (0)
30000	Transportin 2 (TNPO2)	1	2 (1)	1 (0)
7407	Valyl-Trna Synthetase (VARS)	1	3 (1)	0 (0)

Again, a randomized protein set was created to test if influenza virus may be selectively interacting with host proteins based on controllability characteristics. 10,000 random sets of 752 proteins (the number of virus interacting proteins) were sampled from the host proteins of the VIN. Their global type distributions were plotted against the classification results of virus interacting proteins, driver proteins, and all proteins in the VIN (Figure 3.4d-f). As with the robust classification, the random sets closely resemble the total network (one-sided P value: 0.50, 0.51, and 0.50 for critical, intermittent, and redundant, respectively). While there are no redundant driver proteins by definition, driver proteins are more likely to be intermittent proteins than critical proteins (Figure 3.4d-e), where more than 75% of

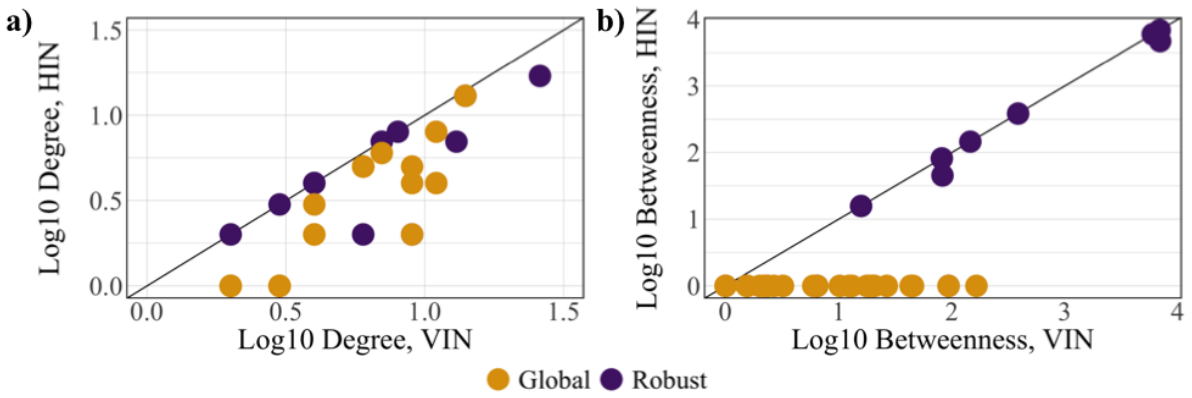


Figure 3.6. A comparison of the a) degree and b) betweenness of the robust (purple) and global (yellow) protein sets identified for influenza A virus. Global proteins are noted for a betweenness of 0 in the HIN while increasing orders of magnitude in the VIN.

all driver proteins are missing from at least one possible MIS. This means the majority of possible driver proteins are controllable by a neighboring protein in at least one MIS. Virus interacting proteins tend to be redundant compared to the total number of proteins that are redundant (Figure 3.4f). This suggests that viruses prefer to interact with proteins that are part of existing control structures to receive input from neighboring proteins.

Overall, global calculations identify a set of proteins for consideration that are more important within the VIN than the HIN. This is demonstrated through a comparison of degree and betweenness for the identified robust and global driver sets in Figure 3.6. Proteins identified in the robust analysis show little deviation in both degree (Figure 3.6a) and betweenness (Figure 3.6b) measures after the addition of virus-host interactions to the network. In contrast, proteins identified in the global analysis show much larger deviations in degree (Figure 3.6a) and betweenness (Figure 3.6b) with all proteins having a betweenness of 0 in the HIN with an up to two log unit increase in the VIN (Table 3.4). Because the identified proteins were not responsible for information flow until the addition of virus-host interactions to the network, this suggests that the global protein set may identify key regulators of host immune response to infection.

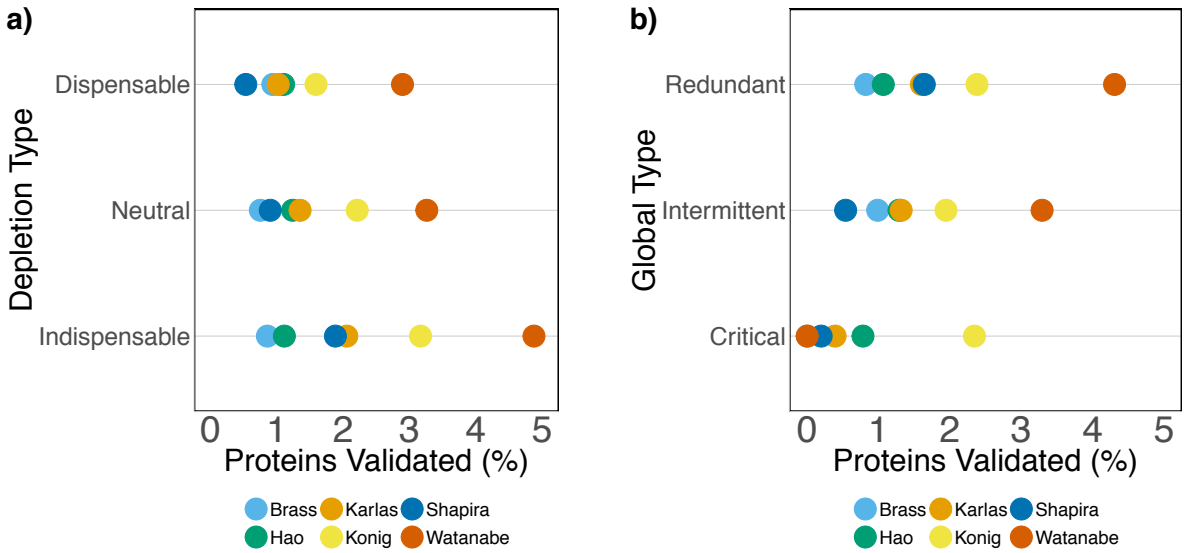


Figure 3.7. Percent of each a) robust classification type and b) global classification type confirmed in 6 siRNA screens (Brass, Karlas, Shapira, Hao, Konig, Watanabe). None of the 6 possible classifications are more than 5% validated in the screenings, suggesting that experimental findings do not favor certain protein controllability types.

3.3.1.4 Partial Genome siRNA Screens Do Not Favor Specific Controllability Classifications

All proteins were checked against 6 siRNA screens for host factors involved in influenza replication (Brass et al. [150], Hao et al. [135], Karlas et al. [136], Konig et al. [152], Shapira et al. [149], and Watanabe et al. [131]), grouped by both robust and global controllability classifications. Less than 5% of all classifications of both types are validated by any of the 6 screens (Figure 3.7), suggesting that no controllability classification is more enriched for host factors than another. This is likely due to the low agreement observed across siRNA studies [135]. However, the driver proteins that change robust and global classification have higher hit rates in siRNA screens, with 2 of 11 changing MIS proteins (SF3B4, SRPK2, 18% validation) and 5 of 24 global-identified proteins (OSMR, PPA1, PSMA5, POLE4, GDI2, 21% validation), though neither are statistically significant results (Fisher P values of 0.685 and 0.252, respectively).

An analysis of both protein sets of interest was performed using Ingenuity Pathway Analysis (IPA) [187]. The network created for the robust protein set identified cellular compromise, cell death, and cell cycle functions. The network created for the global protein

set identified protein synthesis functions, all centered around NF-kB. The global network notably recognizes six proteins (EPHA2, FBL, PFKM, PSMA5, SSR1, and TFRC) for their involvement in the infection of cells (P value: 9.58 x 10^{-4}). Four proteins in the robust network (CELF1, SF384, SRPK2, and HNRNPA0, the last of which appears in both protein sets) were identified for their involvement in mRNA processing (P value: 3.33 x 10^{-6}). Lastly, Interferome v2.01 [186] was used to determine if the 11 robust proteins and 24 global proteins are interferon regulated genes (IRGs). All 11 robust proteins are identified as IRGs and exhibit a 2-fold change in expression when treated with interferon in at least one experimental dataset. 20 of 24 global proteins are identified as IRGs and exhibit a 2-fold change in expression in at least one experimental dataset. 6 global proteins are identified in more than 10 studies. In particular, HNRNPA0 and PPA1 are significantly down regulated in 20 and 63 datasets, respectively. These results point toward the involvement of the predicted protein subsets in immune response events. An overall summary of results can be found in Table 3.5.

Table 3.5. Summary of results for proteins identified in the global controllability analysis.

Quality	**Frequency in Global Protein Set**
Driver protein	100%
virus interacting protein	100%
Identified in robust protein set	8%
Validated in at least one siRNA screen	21%
Cell infection - functional enrichment	25%
mRNA processing - functional enrichment	17%
Interferon regulating gene	83%

3.3.2 Summary

In total, this two-part network controllability analysis for a host protein-protein interaction network (HIN) and an integrated influenza virus-host protein-protein interaction network (VIN) aims to enhance the prediction of antiviral drug targets for

influenza A virus. While robust controllability methods have previously been applied to study PPI networks [177], past analysis focuses only on the classification of virus interacting proteins and does not evaluate changes induced by the addition of virus-host interactions to the network. This study marks the first time a global controllability analysis has been applied to PPI networks. The unique construction of the VIN requires only a minimal disease specific virus-host interaction dataset [131] that represents opportunities for the virus to manipulate host intracellular machinery using protein-protein interactions. Here, analysis of the transition between the healthy and infected network states and further investigation of virus interacting and driver proteins has identified 24 proteins as regulatory markers of the infected state. This protein set is noted for its characteristics in topology, controllability, and functional roles within the infected cell: results that are summarized in Table 3.5. Our workflow observes both the effect of structural changes to the network in the case of potential protein knock outs, as well as each protein's role in all MISs, representing all possible ways of controlling the system. In combination, the network approach and results provide deeper understanding of how changes to cell behavior at the onset of infection are able to occur through the work of a small set of viral proteins. Through understanding the system in this way, we present the possibility to "outsmart" viral attack by dismantling the control structure that allows the viral infection to take hold.

A network representation of the cellular environment demonstrates that the effects of infection (represented by the addition of virus-host interactions) cascade through the system, demonstrated by the alteration of basic topology measures. The betweenness shift between the two networks, particularly in virus interacting proteins, supplies evidence that the topological effect of viral infection is wide reaching (Tables 3.1 and 3.4). Further, a comparison of driver protein betweenness for those that are also virus interacting proteins in comparison to those that are not shows a significant difference. Driver proteins that are virus interacting are not receiving control influence from viral proteins (dictated by the maximum matching requirement that each protein only control a single protein) and require additional external influence to achieve network control. However, the increased betweenness of proteins that are both driver and virus interacting proteins suggests that this group is still of great importance to information flow through the network. This is one

example where differences in network topology measures can emphasize the importance of select proteins that are overlooked by controllability principles.

Controllability analyses confirm that virus interacting proteins are in positions of significance for both types of classification. The increased population of indispensable virus interacting proteins (robust controllability: $N'_D > N_D$, Figure 3.4a) compared to what would be expected by random chance suggests that it would be more difficult for an outside influence (such as viral infection) to control the network after removing the virus interacting proteins opposed to a randomly selected protein. This is logical as virus interacting proteins act as the connection between viral proteins and the host network where control is initiated. The increased population of redundant virus interacting proteins (global controllability: never a driver protein, Figure 3.4f) when compared to the random expectation indicates that more virus interacting proteins are always being manipulated internally than would be expected by chance. This means that they are fully incorporated into the control structure of the VIN. From these two results, one can conclude that virus interacting proteins contribute to both the "gate" (the ease of entering the system) and the "heart" (the proteins responsible for propagating control through the system) of the network control structure during infection. These findings support the idea that viruses are likely to interact with proteins that offer an advantage to total network control.

Similarly, both sets of controllability results demonstrate that driver proteins play interesting roles in the network control structure. The large population of dispensable driver proteins (robust controllability: $N'_D < N_D$, Table 3.2) signifies that the majority of driver proteins are making it more difficult to control the network by requiring more external inputs to control system behavior. In their absence, the number of driver proteins would decrease and it would theoretically be easier for a viral attack to compromise the network control structure. As such, a possible strategy for drug development could be to protect these proteins from repression effects during infection. Over 75% of driver proteins are classified as intermittent (global controllability: sometimes a driver protein, Table 3.3), meaning there is at least one MIS where these driver proteins are not drivers, and receive control influence through internal propagation. This lends itself to the idea of viral escape

routes: under pressure, virus proteins could utilize alternative pathways to maintain system control and reach the goal of hijacking cellular function.

The method of controllability implementation used identifies protein sets of interest through changes to classification between the HIN and VIN. However, robust classification methods do not detect a change between the two networks in this study. As it is a measure of the robustness of the network to structural changes in the absence of each protein, this suggests that the HIN upholds its typical control structure during influenza virus infection. This result could be a consequence of the interaction data used or it may be that the strategy applied here cannot distinguish between the behavior of healthy and diseased states. Knowing the extent of changes to cell behavior within immune response pathways [189, 190, 191], apoptosis signaling [192, 193], and transcriptional processes [194, 195, 196] during infection, the infected cell should be interpreted as a different system. The failure to see this distinction may be a shortcoming of the robust controllability calculation, especially knowing that the 11 robust proteins are not unique due to the method's use of a single MIS.

The 24 proteins identified by global controllability show promise as indicators of regulatory roles specific to the infected state. All global proteins are virus interacting and driver proteins, a high distinction which demonstrates a significant importance to network information flow marked by significantly higher betweenness in the VIN than even driver proteins that are not virus interacting. Additionally, all global proteins have no importance to network flow in the HIN (betweenness = 0) (Table 3.4), suggesting their role in network structure "turns on" after the onset of infection. It is noteworthy that PRDX1 has been implicated in respiratory syncytial virus (RSV) [197], a lower respiratory tract infection that is often associated with influenza virus [198]. Though the number of global proteins identified in existing siRNA screening data is not statistically significant, it should be noted that siRNA screens cover only the partial genome. As such, this type of analysis could be used to direct future experimental studies to save time, money, and effort. IPA analysis reveals that some of the identified proteins hold roles in mRNA processing, an integral part of the influenza virus' ability to spread through processing its own RNA using host machinery [199]. The global protein network is centered around NF-kB, which is

implicated in host immunity with evidence that the virus directly inhibits NF-kB activity [200, 201]. The interferon regulating roles of proteins in a high number of both identified sets (all 11 changing MIS proteins and 20 of 24 global-identified proteins) speak to their high responsibility in controlling infection. PPA1 and HNRNPA0 appear as downregulated in 63 and 20 studies, respectively, when treated with interferon, solidifying their involvement in the host immune response. In total, this evidence suggests that controllability analyses hold power as predictors for important regulators of the host response to influenza infection and, therefore, hold power for drug target prediction.

Existing influenza virus studies using PPI networks require additional data such as differentially expressed gene information [124] or protein context [202] to construct host response networks. Alternative methods such as DeltaNet [203, 204] and ProTINA [205] utilize gene transcription profiles to infer protein drug targets, but rely on the accurate deduction of gene regulatory networks. More recent PPI studies have used network growing functions such as GeneMANIA, STRING, and IPA [125] to predict influenza virus host factors and studied infected cell systems through the integration of screening data with network methods [69, 126]. Approaches incorporating time course data into network analysis have also been explored [206]. While these methods (including basic network metrics such as degree and betweenness of PPI networks) have been successful at identifying disease host factors and in drug target development in the existing body of work, this dual controllability study offers a novel, in-depth analysis of the role of individual proteins in the context of total system function and how possible changes to the system can be interpreted.

3.4 SARS-CoV-2

3.4.1 Addition of Virus Interactions to Host Network Significantly Changes Network Topology

Median log_{10} degree and betweenness of the HIN is 0.699 and 2.945, respectively. Median log_{10} degree and betweenness of the VIN is 0.699 and 2.823, respectively. There is no statistical difference between the degree or betweenness of the HIN and VIN (Wilcoxon rank sum test P values: 0.776 and 0.994, respectively), however, all host proteins have higher betweenness in the VIN compared to the HIN. While viral proteins only interact with 148 proteins, the topological effects are seen across the entire network. Median log_{10} degree and betweenness of host proteins directly interacting with at least one SARS-CoV-2 protein, or "virus interacting proteins", in the HIN are 0.699 and 3.053, respectively. The same values in the VIN are 0.778 and 3.449, respectively. There is a significant difference in mean degree and betweenness distributions of virus interacting proteins compared to the total protein population of the VIN (two sample t-test P value: 3.02 x 10^{-5} and 1.801 x 10^{-6}, respectively). All described degree and betweenness distributions are found in Figure 3.8.

3.4.2 Immune Proteins Become Driver Proteins at the Onset of Infection

Driver proteins are a subset of the network's proteins that must be directly controlled to manipulate total system behavior. This subset, size N_D, is identified through maximum matching algorithms [181] and serves as the first step in both methods of controllability. Calculations identified N_D = 2,463 in the HIN and N_D = 2,466 in the VIN, implying that there is little change to the control structure of the network during infection. All SARS-CoV-2 proteins are driver proteins of the VIN. The 20 host proteins displaced by SARS-CoV-2 proteins as drivers are deemed "displaced proteins". Their identities are listed in Table 3.6. Only five displaced proteins are not virus interacting proteins (HPR, CNNM3, TRIM51, DIP2A, MICA). The removal of these five proteins as drivers of infected cell behavior in the VIN suggests that they have fallen under the control of viral proteins or are part of a host cascade that has been activated in the response. Two proteins are of note: first,

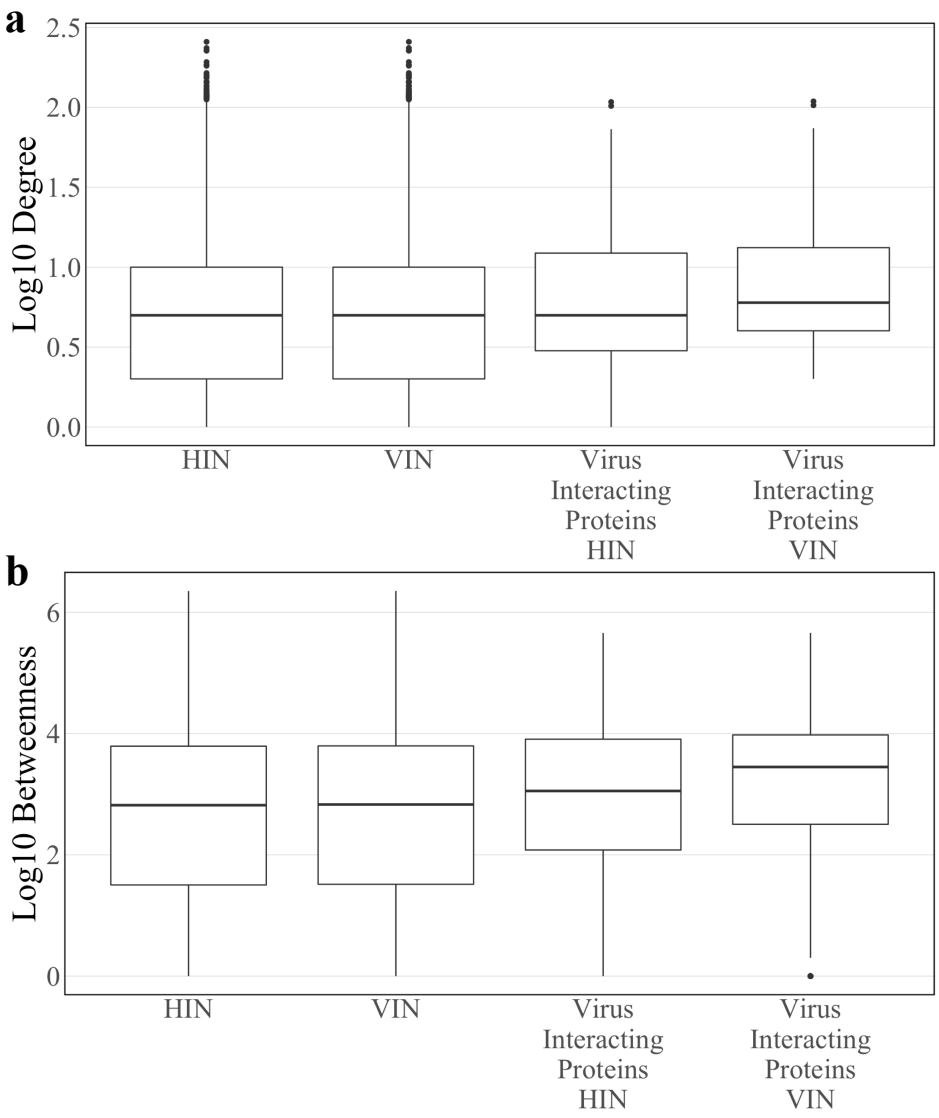


Figure 3.8. The log_{10} (a) degree and (b) betweenness distributions of the Host Interaction Network (HIN) and Virus Interaction Network (VIN) with the corresponding distributions for the subset of SARS-CoV-2 interacting host proteins.

Table 3.6. The identities of the displaced driver proteins: the proteins that are drivers in the HIN, not the VIN. If the proteins are also virus interacting, viral protein interactor is given along with status as an interferon regulated gene (IRG).

Entrez ID	**Gene Name**	**Virus Interaction**	**IRG**
3250	haptoglobin-related protein (HPR)	-	X
23225	nucleoporin 210 (NUP210)	Nsp4	X
26505	cyclin and CBS domain divalent metal cation transport mediator 3 (CNNM3)	-	X
5557	primase (DNA) subunit 1 (PRIM1)	Nsp1	X
23367	La ribonucleoprotein domain family member 1 (LARP1)	N	X
382	ADP ribosylation factor 6 (ARF6)	Nsp15	X
2802	golgin A3 (GOLGA3)	Nsp13	X
949	scavenger receptor class B member 1 (SCARB1)	Nsp7	X
10280	sigma non-opioid intracellular receptor 1 (SIGMAR1)	Nsp6	
84767	tripartite motif-containing 51 (TRIM51)	-	X
3615	inosine monophosphate dehydrogenase 2 (IMPDH2)	Nsp14	X
9470	eukaryotic translation initiation factor 4E family member 2 (EIF4E2)	Nsp2	
55823	VPS11, CORVET/HOPS core subunit (VPS11)	ORF3a, ORF8	X
523	ATPase H+ transporting V1 subunit A (ATP6V1A)	M	X
2876	glutathione peroxidase 1 (GPX1)	Nsp5_C145A	X
23181	disco interacting protein 2 homolog A (DIP2A)	-	X
2150	F2R like trypsin receptor 1 (F2RL1)	ORF9c	X

Table 3.6. (Continued)

Entrez ID	Gene Name	Virus Interaction	IRG
5817	poliovirus receptor (PVR)	ORF8	X
6731	signal recognition particle 72 (SRP72)	Nsp8	X
4276	MHC class I polypeptide-related sequence A (MICA)	-	

TRIM51 is a member of the tripartite interaction motif family of innate immunity regulators [207]. It is previously shown to be highly upregulated in the presence of TLR3 and TLR4 ligands [208] of the viral RNA sensing pathway [209]. Second, MICA is an MHC class I cell surface protein that regulates the activation of both T cells and natural killer cells during a stress response along with other NKG2D ligands such as RAE1 [210, 211]. The displaced protein set was analyzed with Interferome v2.01 [186] to determine their status as interferon regulated genes (IRGs) known to exhibit a fold change in expression greater than two in interferon knockdown studies. All displaced proteins are IRGs with the exception of SIGMAR1, EIF42E, and MICA. The altered role of these immune proteins as drivers of network behavior is representative of the activation of immune response pathways.

A robust controllability analysis was performed on the HIN and VIN to determine the effect of singular protein components on total system behavior. The classification results are shown in Table 3.7. Aside from the addition of viral nodes, there is very little change to the robust controllability of the VIN as compared to the HIN. The majority of all proteins are classified as neutral (VIN: 42.4%, HIN: 42.3%) and dispensable (VIN: 39.0%, HIN: 39.1%), suggesting that most proteins are regulated by neighboring protein pathways (neutral) or make the network easier to control in their absence (dispensable). Conversely, the loss of a small proportion of indispensable proteins (VIN and HIN: 18.6%) would make the network increasingly difficult to regulate. The driver protein population is skewed toward those with dispensable classifications (VIN: 67.2%, HIN: 67.6%) as compared to all proteins. Classifications of virus interacting proteins are similar to those of the total

Table 3.7. The number of proteins in each robust controllability category for all proteins, driver proteins, and SARS-CoV-2 interacting proteins. Values are reported as totals and percent total for the VIN (HIN).

	Indispensable	**Neutral**	**Dispensable**	**Total**
All proteins	1170 (1169)	2675 (2658)	2459 (2454)	6304 (6281)
	18.6% (18.6%)	42.4% (42.3%)	39.0% (39.1%)	100%
Driver proteins	0 (0)	810 (799)	1656 (1664)	2466 (2463)
	0% (0%)	32.8% (32.4%)	67.2% (67.6%)	100%
Virus interacting proteins	32 (30)	57 (57)	59 (61)	148 (148)
	21.6% (20.3%)	38.5% (38.5%)	39.8% (41.2%)	100%

network, eliminating the possibility that viral interactions target proteins that are advantageous to robust controllability. Viral proteins E, nsp5, and nsp10 are classified as dispensable in the VIN. All other viral proteins are classified as neutral.

The nine host proteins that change robust classification after the addition of viral interactions are listed in Table 3.8. Four of the "robust proteins" are also virus interacting proteins (IMPDH2, RAE1, SIGMAR1, NUP210) and three belong to the displaced protein set (IMPDH2, SIGMAR1, NUP210). While only the four virus interacting proteins exhibit an increase in degree from a singular viral interaction, all nine robust proteins demonstrate an increase in betweenness in the infected network, some by orders of magnitude. Of note, the betweenness of IMPDH2 increases from 65 to 4090 and SIGMAR1 reaches 4094 where it has a betweenness of 0 in the HIN. This trend demonstrates the importance of the robust protein set to network information flow and regulation in the infected cell. There is no trend in the changes to classification type for the robust group. Only two robust proteins were identified as IRGs with a fold change greater than two by Interferome (IMPDH2, DYNLT1).

Table 3.8. The identities and topological characteristics of the proteins identified in the robust controllability analysis. Values for degree, betweenness, and classification are given as VIN (HIN). Classification is denoted as indispensable, I; neutral, N; and dispensable, D. Genes that have experimentally shown a fold change in expression greater than 2 during interferon knockdown studies are denoted as interferon regulated genes (IRGs).

Entrez ID	**Gene Name**	**Deg.**	**Bet.**	**Class.**	**IRG**
1174	adaptor-related protein complex 1 sigma 1 subunit (AP1S1)	3 (3)	34.1 (33.0)	D (N)	
3615	inosine monophosphate dehydrogenase 2(IMPDH2)	4 (3)	4090.0 (65.0)	N (D)	X
6993	dynein light chain Tctex-type 1(DYNLT1)	11 (11)	44,093.7 (43,902.7)	D (N)	X
8480	ribonucleic acid export 1(RAE1)	6 (5)	3863.3 (1076.3)	I (N)	
10280	sigma non-opioid intracellular receptor 1(SIGMAR1)	3 (2)	4094.0 (0.0)	I (N)	
10987	COP9 signalosome subunit 5(COPS5)	33 (33)	30,094.4 (29,968.0)	N (I)	
23225	nucleoporin 210(NUP210)	4 (3)	331.1 (325.9)	N (D)	
64326	ring finger and WD repeat domain 2(RFWD2)	3 (3)	207.1 (206.8)	D (N)	
64837	kinesin light chain 2(KLC2)	7 (7)	463.2 (460.3)	D (N)	

To assess whether the robust controllability classifications of the driver and virus interacting proteins are a result of the network's connectivity structure, a randomization analysis was performed. A random set of 148 host proteins representing a "pseudo-virus interacting" protein set was pulled from the network and assessed for robust controllability. The resulting distributions from 10,000 iterations of this process are reported in Figure 3.9a against the true values for all proteins, driver proteins, and virus interacting proteins. Distributions are reflective of the true values of all proteins for all three robust controllability classifications. True values for virus interacting proteins also fall within the distributions for robust classifications implying that there is no regulatory advantage for the particular set of host proteins interacting with SARS-CoV-2 within the robust controllability framework. However, true values for driver proteins fall outside the distributions generated by the pseudo-sets, implying that the groups are distinctly different in regulatory function. A topological analysis including the median and mean values of the same distributions against the true values for virus interacting proteins is shown in Figure 3.10. The mean log_{10} degree and betweenness is significantly higher than the corresponding distribution mean (one-sided t-test P values: 2.2 x 10^{16}, 2.2 x 10^{16}) implying that the virus prefers to interact with proteins that hold significance to network structure.

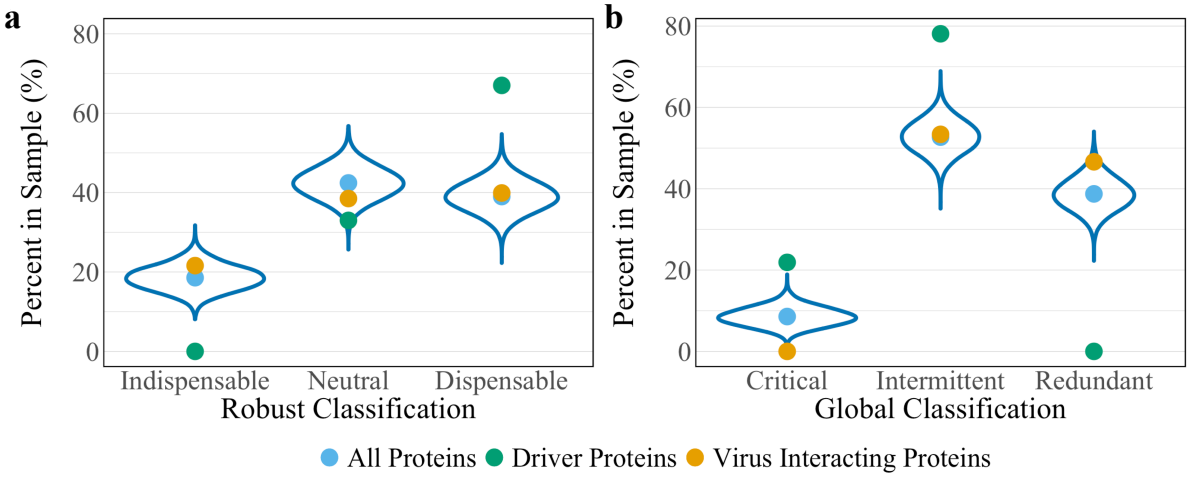


Figure 3.9. Distributions of controllability classification results of 10,000 random "pseudo-virus interacting" protein sets for (a) robust controllability and (b) global controllability. True values for all proteins (blue), driver proteins (green), and virus interacting proteins (yellow) of the VIN are shown as reference for each classification.

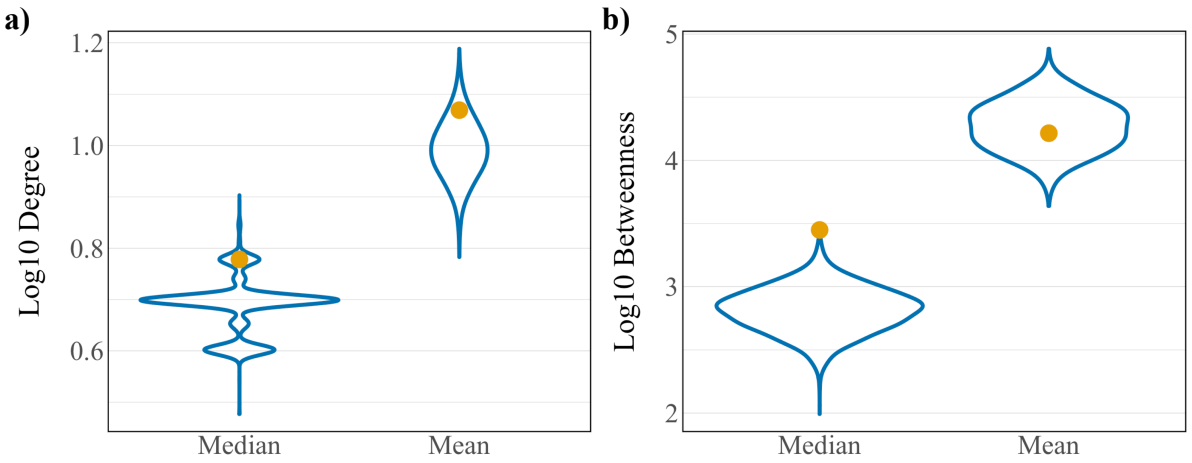


Figure 3.10. Distributions of (a) log_{10} degree and (b) log_{10} betweenness of 10,000 random "pseudo-virus interacting" protein sets. True values for virus interacting proteins are shown in yellow.

3.4.3 Global Controllability Predicts Key Regulators of SARS-CoV-2 Infection

Similarly, a global controllability analysis was performed on the HIN and VIN. Results are shown in Table 3.9. As in the robust controllability analysis, global controllability classifications of the VIN's proteins are almost identical to those of the HIN. Over half (VIN: 52.7%, HIN: 52.8%) of all proteins are classified as intermittent, suggesting that the majority of proteins are able to play a role in cellular regulation. Only a small percentage (VIN: 8.6%, HIN: 8.4%) of all proteins are classified as critical, meaning they are involved in all combinations of network regulators. By definition, driver nodes cannot be redundant, therefore, they are predominately classified as intermittent (VIN: 78.1%, HIN: 80.5%). Unlike the robust analysis, classifications of virus interacting proteins differ slightly from those of the total protein population. The eight critical virus interacting proteins of the HIN become intermittent in the VIN, losing some control over infected network regulation. There is a higher proportion of redundant virus interacting proteins in the VIN (46.6%) compared to both the virus interacting proteins in the HIN and all proteins of both the VIN and HIN (HIN virus interacting proteins: 32.9%, VIN all proteins: 38.7%, HIN all proteins: 38.8%), suggesting that proteins that directly interact with the virus are transitioning into deferential roles after the onset of infection. All 23 viral proteins are classified as critical in the VIN, always holding control of network regulation.

Table 3.9. The number of proteins in each global controllability category for all proteins, driver proteins, and SARS-CoV-2 interacting proteins. Values are reported as totals and percent total for the VIN (HIN).

	Critical	**Intermittent**	**Redundant**	**Total**
All proteins	540 (525)	3322 (3318)	2442 (2438)	6304 (6281)
	8.6% (8.4%)	52.7% (52.8%)	38.7% (38.8%)	100%
Driver proteins	540 (525)	1926 (1983)	0 (0)	2466 (2463)
	21.9% (21.3%)	78.1% (80.5%)	0% (0%)	100%
Virus interacting proteins	0 (8)	79 (75)	69 (65)	148 (148)
	0% (5.4%)	53.4% (50.7%)	46.6% (32.9%)	100%

Eleven host proteins change classification after the addition of viral interactions (Table 3.10). All eleven "global proteins" interact with SARS-CoV-2 proteins with six belonging to the displaced protein set (SCARB1, IMPDH2, PVR, EIF4E2, SIGMAR1, and NUP210). Four global proteins are also identified as robust proteins (IMPDH2, RAE1, SIGMAR1, and NUP210). With the exception of EIF4E2, SIGMAR1, and SAAL1, all members of the set were identified as IRGs with a fold change greater than two by Interferome.

The betweenness of the global protein set increases after the addition of virus-host interactions. In particular, the betweenness of eight of the global proteins (SCARB1, PVR, EIF4E2, CEP135, SIGMAR1, TOR1AIP1, RAB14, and SAAL1) is 0 in the HIN before increasing by several orders of magnitude in the VIN, implying that these proteins are integrated into the network information flow at the onset of infection. Supporting this, seven of the eight shift from critical to intermittent classification (SIGMAR1 becomes redundant) after the integration of viral interactions, indicating new regulation and a loss of control over the network. A topological comparison of the robust and global protein sets within the HIN and VIN (Figure 3.11) demonstrates larger differences between the degree and betweenness of the two networks, making the global controllability analysis a better predictor of the regulators of the infected cell.

Table 3.10. The identities and topological characteristics of the proteins identified in the global controllability analysis. Values for degree, betweenness, and classification are given as VIN (HIN). Classification is denoted as indispensable, I; neutral, N; and dispensable, D. Genes that have experimentally shown a fold change in expression greater than 2 during interferon knockdown studies are denoted as interferon regulated genes (IRGs).

Entrez ID	**Gene Name**	**Deg.**	**Bet.**	**Class.**	**IRG**
949	scavenger receptor class B member 1(SCARB1)	5 (4)	1098.0 (0.0)	I (C)	X
3615	inosine monophosphate dehydrogenase 2(IMPDH2)	4 (3)	4090.0 (65.0)	R (I)	X
5817	poliovirus receptor (PVR)	7 (6)	349.6 (0.0)	I (C)	X
8480	ribonucleic acid export 1(RAE1)	6 (5)	3863.3 (1076.3)	R (I)	X
9470	eukaryotic translation initiation factor 4E family member 2(EIF4E2)	4 (3)	672.1 (0.0)	I (C)	
9662	centrosomal protein 135(CEP135)	2 (1)	46.6 (0.0)	I (C)	X
10280	sigma non-opioid intracellular receptor 1(SIGMAR1)	3 (2)	4094.0 (0.0)	R (C)	
23225	nucleoporin 210(NUP210)	4 (3)	331.1 (325.9)	R (I)	X
26092	torsin 1A interacting protein 1(TOR1AIP1)	2 (1)	41.1 (0.0)	I (C)	X
51552	RAB14, member RAS oncogene family (RAB14)	2 (1)	1.0 (0.0)	I (C)	X
113174	serum amyloid A like 1(SAAL1)	2 (1)	78.2 (0.0)	I (C)	

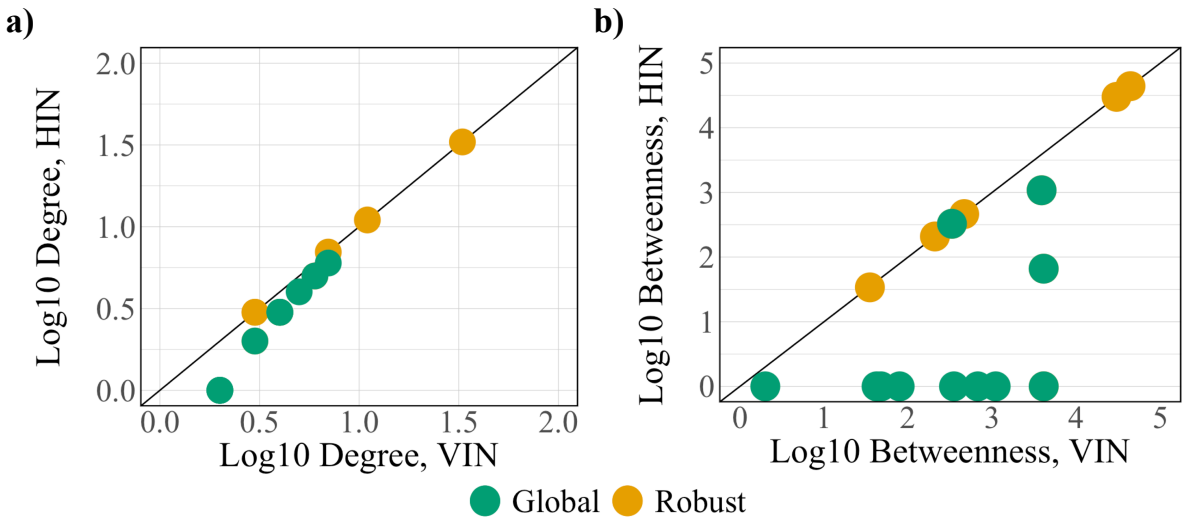


Figure 3.11. A comparison of the (a) degree and (b) betweenness of the robust (yellow) and global (green) protein sets.

A randomization analysis was performed using the same "pseudo-virus interacting" protein sets assessed for robust controllability classifications. The resulting distributions from the 10,000 iterations are found in Figure 3.9b against the true values for all proteins, driver proteins, and virus interacting proteins. Again, random distributions are reflective of the true values from the global controllability of all proteins. While the true value for intermittent virus interacting proteins reflects the random distributions, true values for critical and redundant proteins fall at the tails of the distributions suggesting a regulatory advantage in SARS-CoV-2 interacting with redundant host proteins. True values for driver proteins fall outside the distributions generated by the pseudo-sets, supporting the conclusion that the groups are distinct.

3.4.4 Prioritization of Six Drugs to Repurpose for COVID-19 Treatment

Assuming the identified robust and global proteins are acting as regulators of the infected state, it follows that they have potential as drug targets for SARS-CoV-2 infection treatments. The Drugbank database [188] was used to prioritize the predicted proteins for drug repositioning efforts by assessing which proteins act as targets for existing drugs. Results were compared with the results of drug repurposing and viral inhibition studies

performed by Gordon et al. [168]. Of the 16 combined robust and global proteins, six are drug targets registered in Drugbank (PVR, SCARB1, NUP210, SIGMAR1, IMPDH2, and EIF4E2). NUP210, SIGMAR1, IMPDH2, and EIF4E2 were identified by Gordon et al., though Drugbank identified compounds for each target that were not included in the viral inhibition studies. The compounds associated with two additional targets that were identified by the controllability methods but are unregistered in Drugbank (LARP1 and RAE1) were also previously identified. The targets and associated compounds for all eight genes are found in Appendix Table B1.

A summary of compounds known to target PVR and SCARB1, i.e., the prioritized proteins that have not been recommended in previous repurposing studies are found in Table 3.11. PVR is a known regulator of natural killer cell adhesion to host cells and lytic granule secretion after binding to DNAM-1, a receptor expressed by natural killer cells, T cells, and monocytes [212]. First identified in the context of polio virus, it has also been identified for its role in motility during tumor cell invasion [213]. It is a member of the displaced driver set and the global set and acts as the target for two experimental compounds: myristic acid and sphingosine. A previous study of cytokine storms resulting from influenza virus infection asserts that the use of sphingosine-1-phosphate successfully blunts the overactive inflammatory response, limiting morbidity and mortality [214]. Given the similarly aggressive inflammatory response seen clinically in SARS-CoV-2 infected individuals [215], the prioritization of PVR is noteworthy.

SCARB1 is a cellular membrane protein involved in high-density lipid transport [216] that mediates cell entry of hepatitis C virus as the receptor for the E2 protein [217]. It was identified as both a displaced driver and global protein, and functions as the target for three compounds: phosphatidylserine, tocopherol/vitamin E, and PHA-665752. While not specific to viral infection, SCARB1 acts as a phosphatidylserine receptor on testicular Sertoli cells which induce phagocytosis of spermatogenic cells [218]. SCARB1 also acts as one of the most important transport vehicles for vitamin E in the lung's alveolar cells, the presence of which largely regulates the receptor's expression [219]. Vitamin E is also known to have a positive effect on influenza A viral clearance in the lungs of mice [220].

Table 3.11. Status of drugs known to target controllability predicted proteins. Each target's SARS-CoV-2 protein interactor is given along with its known target function.

Drug (Status)	**Target/Viral Protein**	**Target Function**
Myristic acid (Experimental)	PVR/ORF8	Regulate Natural killer cells, polio virus [212]
Sphingosine (Experimental)		
Phosphatidyl serine (Approved)	SCARB1/Nsp7	Facilitate cell entry, Hepatitis C [216]
Tocopherol/Vitamin E (Approved)		
PHA-665752 (Experimental)	SCARB1/Nsp7	Facilitate cell entry, Hepatitis C [216]
	NUP210/Nsp4	Transport between nucleus and cytoplasm [220]

In addition to SCARB1, PHA-665752 targets NUP210, a nucleopore protein identified in all three protein sets of interest from controllability. Knockout experiments reveal that NUP210 has wide effects on T cell differentiation and response [221]. While not studied in the context of viral infection, PHA-665751 is known to induce apoptosis in both tumor and vascular endothelial cells resulting from non-small cell lung cancer [222].

SIGMAR1 and IMPDH2 were also identified in all controllability predicted groups. Sigma receptors 1 and 2 (SIGMAR1 and SIGMAR2) have been discussed as regulators of cellular stress and the apoptotic response [223]. Several known targeting compounds show evidence of inhibition at the viral replication stage including haloperidol, PB28, and widely discussed hydroxychloroquine. Many Drugbank predicted compounds targeting SIGMAR1 have not been tested for viral inhibition. IMPDH2 has long been a goal for targeting with immunosuppressive treatments, though inhibition with small molecules is notoriously difficult [224]. While many Drugbank-predicted compounds for IMPDH2 have not been tested in this context, mycophenolic acid and ribavirin were assessed in the viral inhibition screen with only the latter displaying active inhibition.

EIF4AE belongs to the 4E family of translation initiation factor proteins which bind to mRNA 5' cap structures to recruit ribosomal recruitment within the cytosol [225]. The 4E family controls the rate of the early steps of the protein translation process. EIF4AE is a member of both the displaced driver set and the global protein set. A different translation regulator, LARP1, was identified as a druggable target for SARS-CoV-2 by Gordon et al., given previous evidence that the downstream effects of a common kinase inhibitor, rapamycin, inhibited MERS-CoV infection by over 60% [226]. Studies show that rapamycin promotes the phosphorylation of EIF4AE, achieving similar inhibition of the mTOR pathway [227]; however, evidence of viral inhibition after rapamycin treatment was inconclusive. Still, several other mRNA translation inhibitor compounds such as ternatin 4 and zotatifin have tested as active inhibitors of SARS-CoV-2, making EIF4AE an interesting prospect for future study.

3.4.5 Summary

Here, a set of drug targets is prioritized for drug repurposing efforts in the global fight against COVID-19. Network controllability methods, with only disease-specific virus-host and host-host protein interaction data, create a large-scale representation of regulatory changes occurring during infection. With no additional biological information, the connectivity of the network is sufficient to predict the most biologically relevant components of the disease system, as evidenced by the high level of overlap between the presented results and the extensive biological analysis performed in Gordon et al.'s study for SARS-CoV-2 [168]. In total, this study demonstrates a simple computational approach to prioritizing drug target predictions with minimal biological context, an advantage in present times where viral understanding and data is even more sparse than usual.

As seen in the previous study of influenza A virus [228], the magnitude of control needed to manipulate the total cell system (number of driver proteins) is comparable in the healthy and infected cellular networks. The small changes in driver proteins between the networks are seen in immunoregulatory proteins that are typically upregulated during viral infection (such as TRIM51 and MICA) and many of the proteins identified in the controllability analyses. This is reflective of the activation of the immune response pathways and their effect on the cell as a whole. With respect to the ratio of resultant classifications in both controllability methods, outcomes are again similar to those achieved with the influenza A virus-host network [228]. This is unsurprising due to the use of the same host network in the analyses. However, the low overlap between the controllability predicted proteins for the two diseases (3/16 proteins, PVR, RAB14, and SAAL1) demonstrates that while the method is easily applied to other viruses, the result is unique.

One limitation of this method is the requirement of high-confidence virus-host protein interaction data where the host proteins exist in the HIN. As experimentally-validated, directed networks are typically smaller than the available undirected networks, the method was unfortunately unable to use over half of the 332 known SARS-CoV-2-host protein interactions (in comparison, the influenza A virus network contains 752 virus-host interactions). Even so, the controllability analysis was able to predict biologically relevant

proteins involved in functions like the cellular stress response, host translation, and cellular transport, proving the robustness of the method.

Of the eight prioritized targets, all but NUP210 exhibit large increases in betweenness after the addition of virus-host protein interactions, placing particular importance on their role in infected cell behavior based on topology. Further, most of the global protein set (including the novel predictions, PVR, and SCARB1) have a betweenness of zero in the HIN, implying that their individual significance to cellular network flow is truly unique to the infected cell. With the majority of the identified proteins being regulated by the interferon response (with a fold change in expression greater than two), this network result translates to immunological significance. The alterations in classification for all global proteins indicate a step down in network control where critical proteins have the most control and redundant the least. Biologically, this could represent viral interruption of normal host function or activation of a new pathway, both being interesting prospects for drug development.

Given the biological relevance of the topologically predicted/controllability target proteins, there is a good reason to pursue these recommendations, either in drug repurposing or in novel drug development. The extended list of untested compounds found in Appendix Table B1 will be considered for further viral inhibition studies, particularly tocopherol/vitamin E which is already approved and has documented positive effects on viral clearance for influenza A [220]. Predictions indicate opportunity to both interfere with the viral replication cycle or to modulate the immune response to infection. Therefore, to most efficiently translate these findings to bedside, knockdown studies or siRNA screens should be used to validate drug predictions for each target. Cell culture studies that track interferon and cytokine activity may further establish a possible mechanism between the proposed targets and immune regulation. By narrowing the pool of drug target candidates with controllability methods, experimental validation will be efficient and timely.

4.0 Aim 3: Modeling Strain-Specific Immunodynamics During Viral Infection

4.1 Introduction

Mathematical (ODE) models of the host immune response in influenza A virus infected lungs have served as a computational method of treatment optimization [63, 229, 66, 230, 231, 232, 67]. ODE modeling is advantageous in tandem with traditional experiments in its ability to determine kinetic rate parameters of influenza infection that are difficult to measure *in vivo*. Many experimental data sources, particularly murine (mouse) models of influenza, are generated from a pool of measures collected from hosts subjected to identical experimental conditions. Because these animals need to be sacrificed to measure cell and cytokine levels, hosts cannot be tracked for the full duration of the infection, making true longitudinal data impossible to obtain. These experiments assume that all animals will mount identical immune responses, though inter-individual variability is a known confounding factor. Mathematical modeling can be used to help fill in gaps in knowledge created by the deficiencies in experimental data.

ODE models can vary substantially in complexity as they depend on the chosen representative components of the immune response and the interactions between them. There are two general types of model: target cell-limited models, in which the healthy epithelial cells acting as a target for the virus are unable to replicate themselves [229, 230] and models in which healthy cells are able to regenerate [63, 66, 231, 232, 67]. Current models feature three fundamental states: healthy epithelial cells, infected epithelial cells, and virus. More states, such as cytokines, immune cells, or antibodies, can be added to the model with additional ODEs and rate parameters. While large models create a more biologically complete picture of the immune response, they require a larger pool of quality data to parameterize the model and can easily fall into stability issues. The question becomes: which interactions comprise a minimum necessary set for characterizing the immune response without losing critical information to simplifying assumptions?

The interferon response is a key regulator of the immune system and is foundational to capturing the innate immune response in a mathematical model. While interferon is essential for viral clearance and has been heavily studied since its discovery in 1957, full understanding of the role of interferon dynamics in clinical pathology remains elusive [233, 234]. Modulating the immune response post infection to control inflammation or pre-infection to provide increased protection for high risk groups has been a major theme in severe influenza infection research [235, 236, 237, 238, 239, 240, 241, 242, 53, 243, 244]. When pre-stimulated with synthetic or natural agonists of the Toll-Like Receptor pathways (specifically TLR3 and TLR4) that activate interferon production prior to infection [245, 246, 247, 248], animals infected with highly pathogenic viruses elicited higher concentrations of interferon in lung epithelial cells, reduced virus titers and significantly improved infection outcomes. Several studies demonstrate dysregulation of the immune response during deadly influenza infections [238, 242, 249, 250] suggesting that immunomodulation prior to infection may be an option for protecting high risk groups.

4.1.1 Model Review

The ability to replicate the effects of interferon-regulating pathways provides a valuable measure of model applicability. In preparation for modeling studies, we reviewed three recently published models (Figure 4.1) of varying complexity surrounding interferon interactions (Appendix C) (Saenz et al., 2010 [229], Pawelek et al., 2012 [63], and Hancioglu et al., 2007 [66]). Five elements of the intrahost immune response are conserved across each model: healthy epithelial cells (H), infected cells (I), virus (V), type I interferon (F), and "resistant cells", that is, epithelial cells with interferon-induced virus resistance (R). While each model has these five features in common, the inflammatory response to viral infection is represented differently, depending largely on model complexity. These differences are particularly apparent in the model-specific incorporation of the production, activity, and depletion of interferon.

In the Pawelek model (Figure 4.1a), interferon has two functions: creating virus-resistant cells when interacting with healthy epithelial cells, and increasing infected cell death when

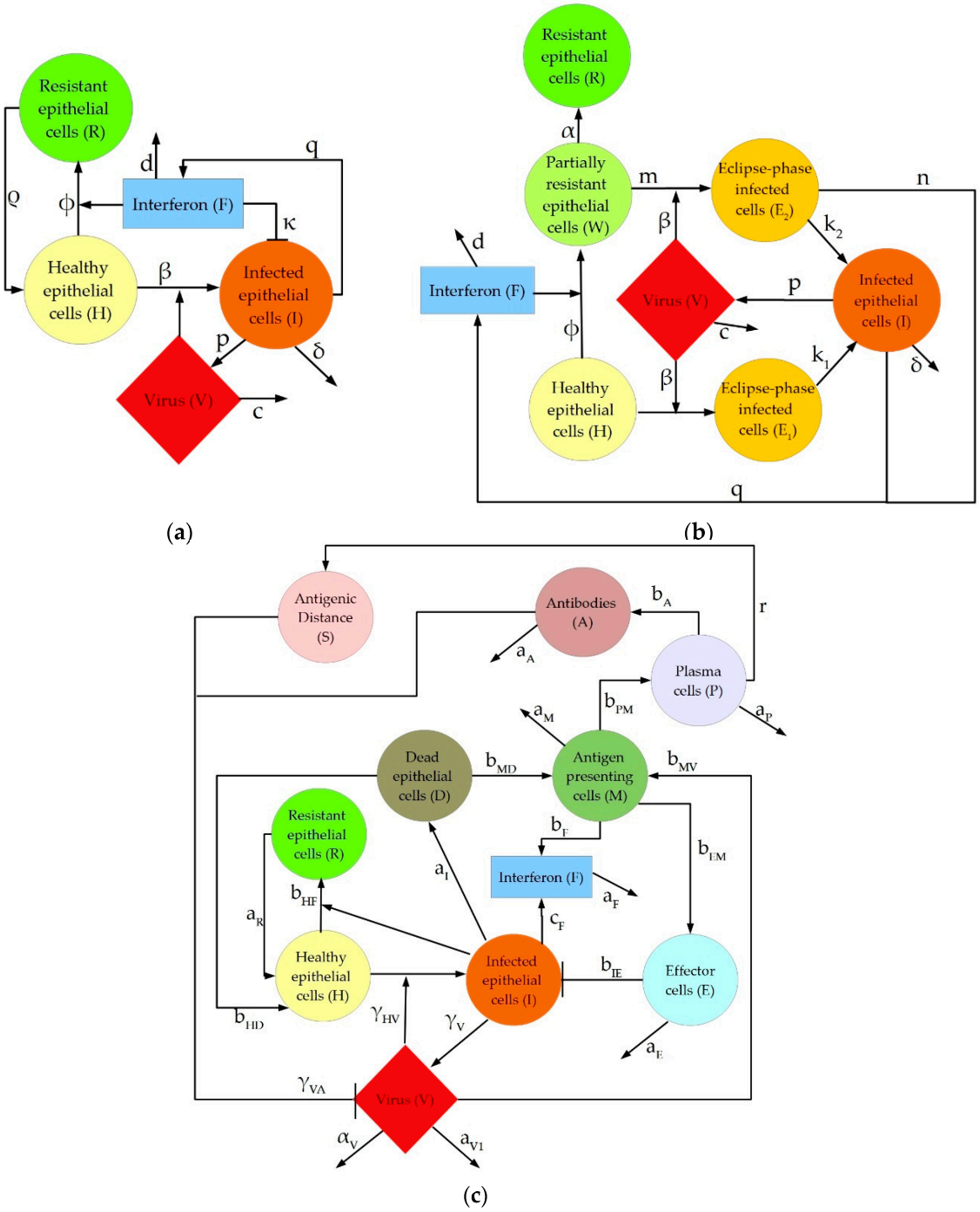


Figure 4.1. Model diagrams for Saenz, Pawelek, and Hancioglu models of the host immune response to influenza A infection. These models all include the interferon response and rely on epithelial cell populations, particularly resistant cells.

interacting with infected epithelial cells. In the Saenz model (Figure 4.1b), interferon leads to the creation of virus-resistant cells but does not impact the infected cells directly. Instead, the infected cells produce more interferon. The Hancioglu model (Figure 4.1c) uses interferon to create resistant cells (as in the other two models) while interferon is produced by infected cells and antigen-presenting cells. In all models, a decrease in interferon levels is caused by a combination of natural decay and absorption into epithelial cells. The Saenz and Pawelek models are trained to experimental data (e.g., cytokine concentrations and immune cell counts) measured in pony lungs infected with H3N8 virus, while the Hancioglu model was fit to certain qualitative behaviors selected from a study of the human response to influenza A virus by Bocharov and Romanyukha [251].

We considered the "systems" perspective of which components of the model most strongly regulate virus replication. Each of the three models shows a sensitivity of the virus to the creation and loss of infected epithelial cells. The virus equation of the Pawelek model is most sensitive to the loss of resistance in epithelial cells, ρ, and the death rate of infected cells, δ (Figure 4.2A). If the infected cells die off too quickly, the virus cannot replicate at a rate high enough to sustain the infection. Similarly, if cells are becoming virus-resistant too quickly, there will not be a sufficient number of cells remaining to become infected and keep the viral titers elevated. In this way, the presence of the virus in the system is predominantly driven by the number of cells currently infected or able to become infected. The Saenz model also emphasizes a low death rate of infected cells, δ, as well as a short eclipse phase for infected cells, k_2 (Figure 4.2B). The duration of the eclipse phase determines the delay in time between the infection of the cell and the subsequent release of virion by the infected cell. The shorter the eclipse phase, the more readily the cells can begin producing virus. As in the Pawelek model, the Saenz model shows that the availability of productively infected cells is vital to the continuation of the infection.

The Hancioglu model also emphasizes the importance of maintaining a large pool of infected cells, but through a different set of parameters than the Pawelek or Saenz models. The infectivity of the virus, γ_{HV}, and the replication rate of the virus, γ_V, are the most sensitive parameters in the model (Figure 4.2C). The Hancioglu model is thus controlling the virus by a high rate of production of infected cells, and not through a diminished rate

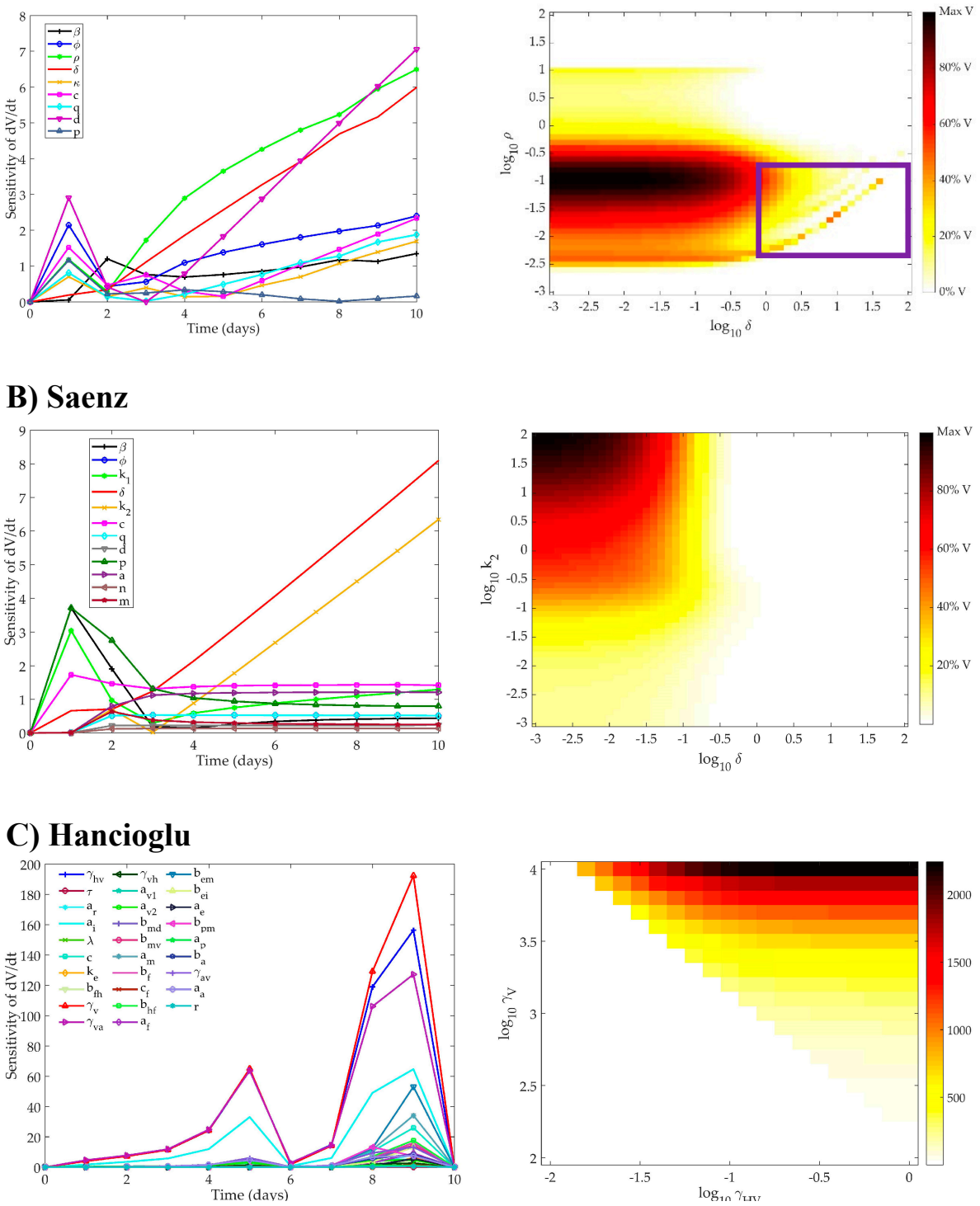


Figure 4.2. Time-dependent sensitivity of virus to each parameter and two-dimensional sensitivity to the two most sensitive parameters in the A) Pawelek, B) Saenz, and C) Hancioglu model. Colors correspond to the maximum amount of virus present over a ten-day simulation. White indicates that the maximum value of V is 0.01, the initial value of the virus. Darker colors indicate higher values of peak virus.

of decay of these cells, as in the other two models. Interestingly, none of the three models shows a strong sensitivity of virus to the concentration of interferon in the system.

This review has shown that simply creating a population of virus-resistant cells is not sufficient to model the impact of interferon on control of virus replication. This is the mechanism by which many current published models, including the three covered here, incorporate the effect of interferon on the immune response. For a truly accurate mathematical model, the model structure should be able to simulate known qualitative behaviors as well as reproduce the quantitative data used to tune the model parameters. Moving forward with this information, ODE models of influenza infection should include a better representation of innate immunity, and possibly more interactions of interferon with other components in the model, to accurately portray the impact of interferon on the system as a whole. Rather than reliance on the creation of virus-resistant epithelial cells to simulate the effect of interferon on the host, interferon should be used to directly diminish the replication rate of the virus, similar to a model proposed by Baccam et al. [230].

4.1.2 The Strain-specific Immune Response

A comparison of influenza A virus strains reveals distinct trends in the severity of clinical outcomes. High pathogenic strains such as H5N1 induce increased levels of inflammation, congestion, and tissue damage extending deep into the alveoli of the lung in comparison to low pathogenic strains such as the seasonal H1N1 that replicate in the larger airways of the upper lungs [252]. High pathogenic strains are generally noted for their virulence, lengthened clearance times, and increased mortality rates, exemplified by H5N1 where an estimated 60% of human cases end in death (the majority of which unexpectedly occur in those under age 65) [22]. Viral titers, immune cell counts, and cytokine counts for high pathogenic strains have higher magnitudes and sustained production compared to low pathogenic strains, indicating disparate immunoregulatory behavior across viral strains [54]. Biologically, influenza virus strains are subtyped by the viral envelope proteins hemagglutinin (HA) and neuraminidase (NA) [253]. The presence of HA leads to a 60% reduction in STAT1 phosphorylation in correlation with IFNAR1 downregulation [254], ultimately downregulating the expression of

interferon stimulated genes and altering the positive feedback relationship between STING and interferon production [255]. Infection with NA-defective influenza virus induces lower levels of type-I interferon [256]. Additionally, studies have identified that a D92E mutation in the influenza A nonstructural NS1 protein of highly pathogenic H5N1 has a repressive effect on interferon production [257]. NS1 has been implicated as a factor in the type-I interferon response time after the discovery that low pathogenic strains grafted with NS1 from high pathogenic strains remain more capable of inhibiting interferon signaling, resulting in failed suppression of viral replication and delay in the overall immune response [258]. Exploration of the relationship between virus replication and interferon production as well as the timing of the innate immune response is needed to identify causes of the *in vivo* correlation between viral strain and severity.

Previous studies have used ordinary differential equation (ODE) modeling methods to understand *in vivo* immunodynamics during influenza infection, though differential strain behavior has never been explicitly explored. Existing models often focus on a range of epithelial cell states as well as interferon production [63, 229, 259, 260, 261], however, states representative of additional cells and cytokines of the innate immune response are limited [66, 67]. ODE modeling efforts are limited in part by the quality of training data which is often in the form of triplicate viral titers, immune cell counts and/or cytokine assays from *in vivo* experiments. The marked absence of data regarding the epithelial cell populations of interest (many models move beyond healthy and infected cells to include refractory cell states that denote stages of interferon production) leads to identifiability problems during parameter estimation and often necessitates that heuristics be imposed on the model to force biologically relevant behavior. All data quality and quantity issues contribute to model error and uncertainty and diminish the overall usefulness of the model for understanding system dynamics. As such, model design should seek to eliminate these sources of error when possible.

To elucidate the biological mechanisms that contribute to strain-specific immunodynamics observed during influenza virus infection, we have performed a shared, multi-strain parameterization of two ODE models. The first aims to assess the inclusion of biological mechanisms representative of both the innate and adaptive immune response.

The second eliminates the adaptive and epithelial states to focus on the interferon feedback of the early innate immune response.

4.2 Multi-strain parameterization of Innate-Adaptive Model

4.2.1 Materials and Methods

4.2.1.1 Model Creation

A mathematical model representative of the immune response to influenza A virus infection was developed (Equations 4.1-4.6, parameter key Appendix D). A brief schematic of the system in Figure 4.3 outlines the interactions included in this six state model of the innate and adaptive immune response. Healthy cells (H) become infected cells (I) at a rate influenced by the virus infectivity parameter (β). The amount of virus (V) present is dictated by three parameters: replication rate (p), depletion by interferon (q), and clearance by antibodies (c). An innate immune response is represented by the presence of macrophages (M) and interferon (IFN), where interferon production is notably influenced by the presence of macrophages ($b_{M,IFN}$). The adaptive immune response is represented by cytotoxic T lymphocytes (CTL). In total, the model contains 15 parameters.

$$\frac{dH}{dt} = -\beta HV + \lambda H(1 - H - I) \tag{4.1}$$

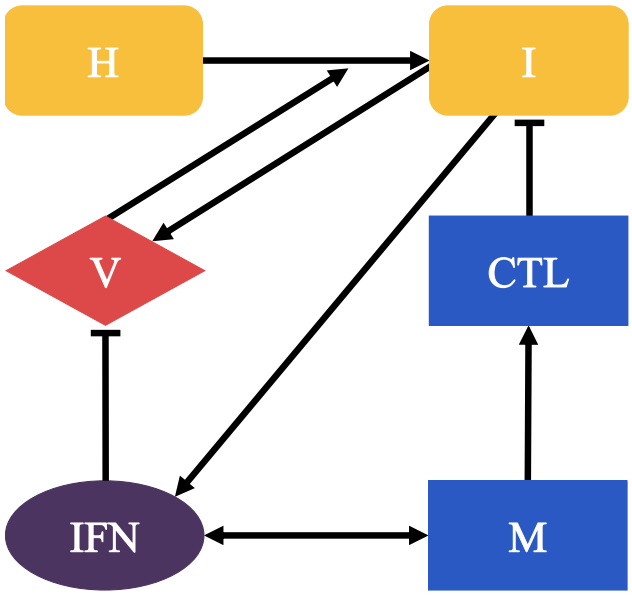
$$\frac{dI}{dt} = \beta HV - \delta I - K_{CTL} CTL * I \tag{4.2}$$

$$\frac{dV}{dt} = pI - cV - qV * IFN \tag{4.3}$$

$$\frac{dM}{dt} = b_{IFN,M} IFN * M - \mu_M (M - M_0) \tag{4.4}$$

$$\frac{dIFN}{dt} = \frac{b_{I,IFN} I^n}{I^n + \theta_{IFN}^n} + \frac{b_{M,IFN} (M - M_0)^n}{(M - M_0)^n + \theta_{IFN}^n} - \mu_{IFN} IFN \tag{4.5}$$

$$\frac{dCTL}{dt} = b_{M,CTL} M - \mu_{CTL} (CTL - CTL_0) \tag{4.6}$$



States:

H: Healthy cells
I: Infected cells
V: Influenza virus
IFN: Interferon
M: Macrophages
CTL: Cytotoxic T Lymphocytes

Figure 4.3. Innate-Adaptive model scheme. Macrophage represent the early, innate response and cytotoxic T lymphocytes represent the late, adaptive response.

4.2.1.2 Multi-strain MCMC Parameterization

Parameter fitting was performed in MATLAB using a shared parameter Markov Chain Monte Carlo algorithm as modified from Price et al [67]. As shown in Figure 4.4, each MCMC study begins with the decision of P_{shared}, the parameters to be kept equal across strains. After solving the ODEs with an initial parameter set, P_0, H1N1 parameters are stepped to new values on the chain and the ODEs are solved. P_{H5N1} is set equal to P_{H1N1} for P_{shared} and the chain is stepped to new values for the remaining parameters of P_{H5N1} before solving. The per strain objective function,

$$Energy = \sum_{x=1}^{X} \sum_{t=0}^{T} \frac{(M_{x,t} - O_{x,t})^2}{\sigma_{x,t}} \tag{4.7}$$

is based on the residual sum of squares: the squared difference between model output, $M_{x,t}$, and observed data, $O_{x,t}$, for each state, x and timepoint, t, over all time points, T, for all species, X. Each time point was divided by the standard error of the corresponding data $\sigma_{x,t}$ to take the variance of triplicate measures into consideration. The model was trained with data for four of the six states (V, CTL, IFN, M) [54]. Virus kinetics used were measured in plaque assay studies. Macrophage and CTL counts were determined using flow cytometry. All data is derived between days zero and seven post-infection from mouse lung infected with 10^5 PFU of H1N1 or H5N1 influenza A virus. Cell data was pre-processed by normalizing raw data per cell type by the reported live cell count and multiplying each value by a constant, 10^7, to return to cell count units. Gene expression data for the probe corresponding to gene *ifnb* was averaged within each time point per strain and used as time course data to inform interferon dynamics. Initial conditions are dependent on data when available (V, CTL, M). Because interferon is initially below the limit of detection, it is initialized at a value of zero. Assuming there is no infection prior to time 0, the healthy cell fraction is initialized to one, and for the same reason, the infected cell fraction is initialized to zero. No heuristics were added to the model. A sensitivity analysis was performed in MATLAB.

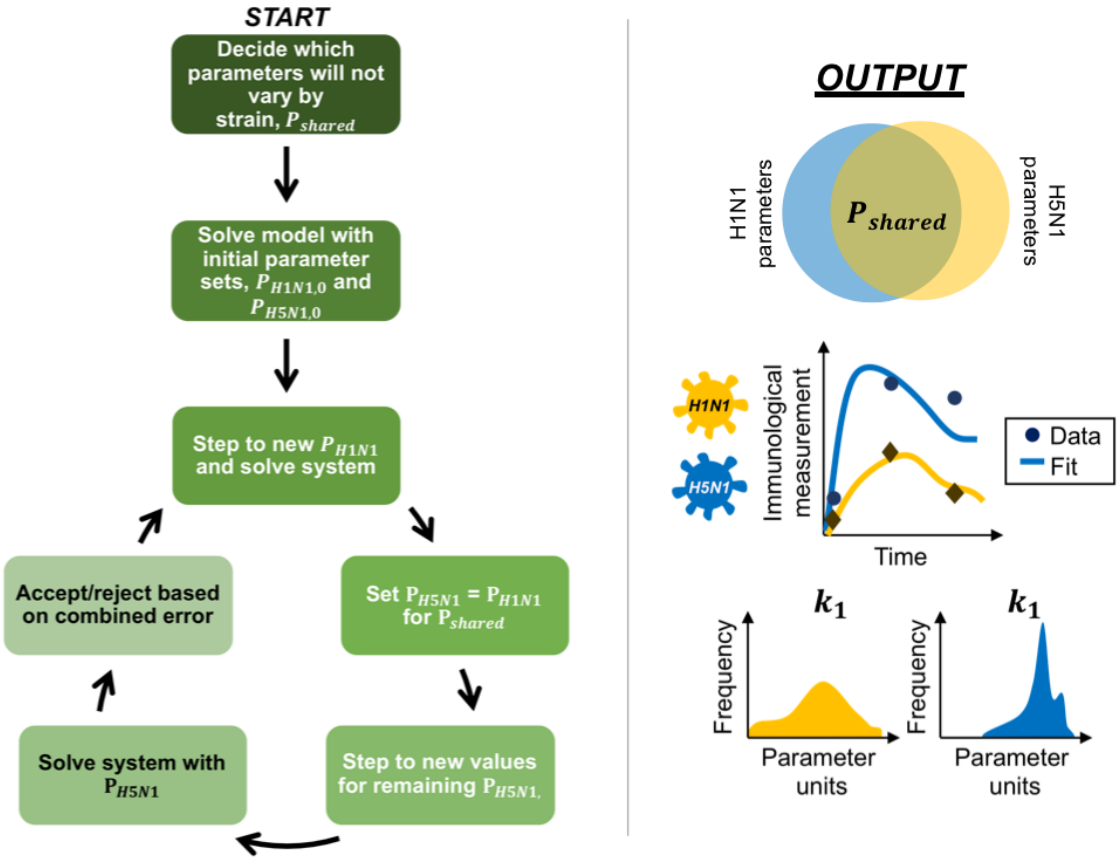


Figure 4.4. A summary of the multi-strain MCMC parameterization method. The method output contains a parameter set, best fit solution, and parameter distributions for each strain of influenza virus.

4.2.2 Results

4.2.2.1 H1N1 Model Cannot Predict H5N1-induced Dynamics with Variance in Virus Replication Parameters

First, the model was fit to H1N1 infection data and evaluated for its ability to simultaneously predict immune response dynamics in H5N1-infected animals when allowing only strain-specific parameterization of the four virus-specific parameters: virus infectivity, β, viral replication rate, p, viral depletion by interferon, q, and viral clearance by antibodies, c. However, no solution was identified in which the H1N1-trained model could predict the H5N1-induced responses (results not shown).

4.2.2.2 H1N1 Model Predicts H5N1-induced Dynamics with Added variance in Macrophage Interferon Production

A second analysis used the H1N1 trained model to predict H5N1-induced dynamics as in section 4.2.2.1 with the addition of the parameter describing the effect of macrophages on interferon production, $b_{M,IFN}$, to the independently trained H5N1 parameter set. This parameter was chosen because *in vivo* data suggests H5N1 infected resident macrophages induce a stronger interferon response [262]. Dynamics of the best parameter fits for H1N1 and H5N1 are found in Figure 4.5a and Figure 4.5b, respectively (1,850,000 training instances; suitable models selected after parameterization burn in). Each fit is shown with the intervals containing 25-75% (dark grey) of all parameter sets and 5-95% (light grey) of all parameter sets. In the H1N1 model fits, all states are well fit to the data through the seven day span of infection with little variance in dynamics caused by the parameter set distribution. The only notable exception of good fit is seen in CTL's where there is no data between days 3 and 7 and the data present exhibits a sharp increase at the later time point.

The same model accurately predicts the immune dynamics induced by the high pathogenic H5N1 strain when only the virus-specific parameters and $b_{M,IFN}$ are allowed to differ from those of the H1N1 parameter set. A large population of infected cells peak around days 1-2 and, unlike H1N1, never return to zero. However, this is concurrent with

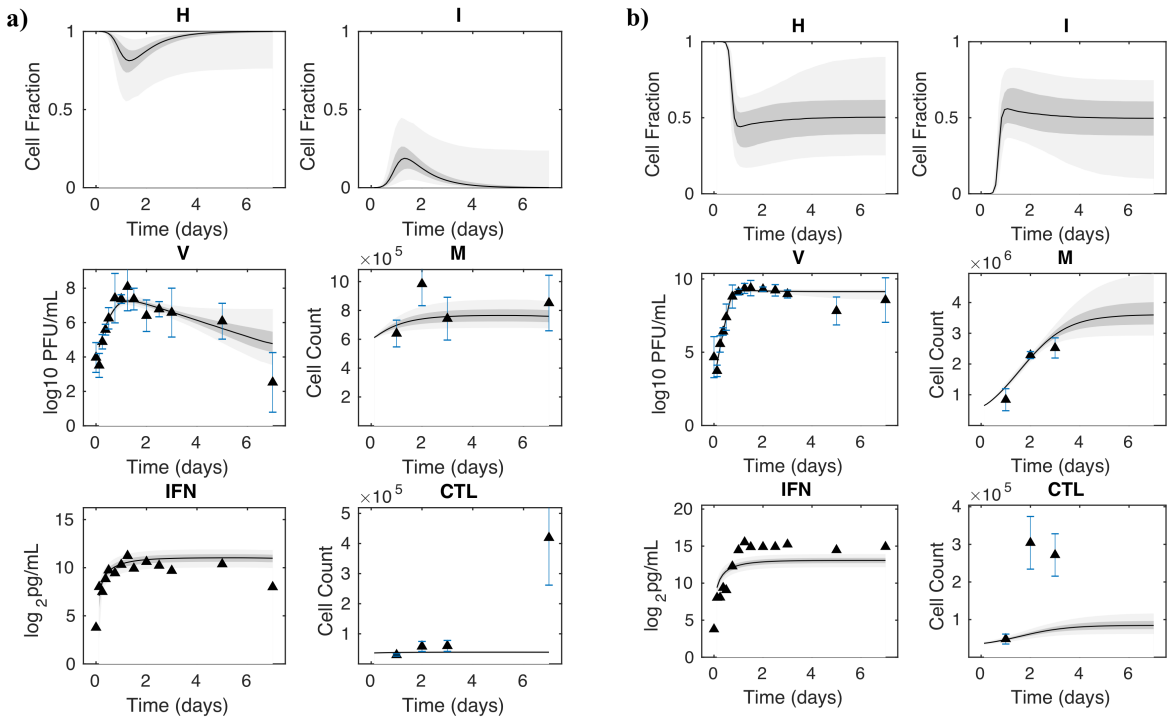


Figure 4.5. Model fits for a) H1N1 and b) H5N1 where virus-specific parameters and $b_{M,IFN}$ were allowed to vary between strains. Fits are plotted against training data. Intervals in light grey and dark grey contain 25-75% and 5-95% of all parameter value sets, respectively.

the high viral load still in place past day 4 which does not occur in H1N1. Additionally, recovery is not expected because H5N1 infected animals die before day 7. Fits are generally in agreement with the data sources, displaying greater variance in parameter sets in comparison to H1N1. Similar to the H1N1 model, the steep increase in CTL counts is not captured in the model dynamics. Overall, the model correctly predicts a more severe response to H5N1 than H1N1 influenza strains.

An analysis of the variance in strain parameter distributions is shown in Figure 4.6 with H1N1 distributions in green and H5N1 in blue. All virus-specific parameters and $b_{M,IFN}$ display a difference in distributions between strains (P value $< 2x10^{-16}$, Kolmogorov-Smirnov test). Distributions for virus infectivity, β, and viral depletion by interferon, q, are slightly elevated in magnitude for H1N1 where distributions for viral replication rate, p, viral clearance by antibodies, c, and the effect of macrophages on interferon production, $b_{M,IFN}$, are elevated by orders of magnitude in H5N1. Narrow distributions suggest that H1N1 is highly sensitive to virus depletion by interferon where H5N1 is sensitive to interferon production in the presence of macrophages.

4.2.2.3 Sensitivity in Virus Infectivity Leads to Variance in Viral Production

A sensitivity analysis was performed to assess the parameters with the greatest effect on output dynamics during influenza infection. Table 4.1 shows the time dependent sensitivity of the virus concentration to each of the 15 model parameters. Seven parameters are most in control of the virus trajectory: virus infectivity, β, the natural death rate of infected epithelial cells, δ, the clearance rate of infected cells by CTLs, K_{CTL}, viral depletion by interferon, q, viral clearance by antibodies, c, activation of macrophages due to interferon, $b_{IFN,M}$, and the decay rate of interferon, $_{IFN}$. Sensitivity peaks before day 2 for most parameters and drops off over the remainder of the infection.

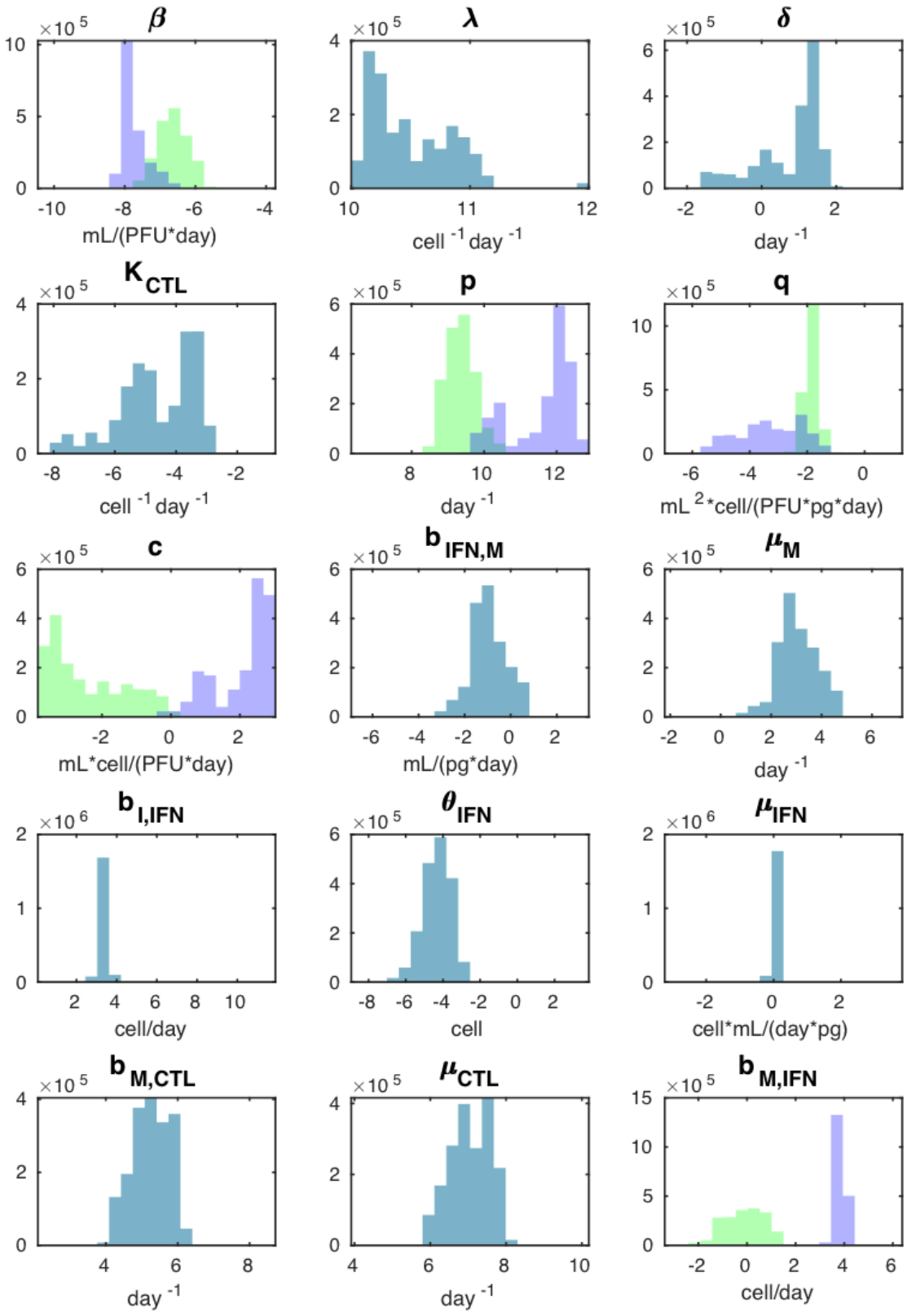


Figure 4.6. Log_{10} parameter distributions for H1N1 (green) and H5N1 (blue) for the model where virus-specific parameters and $b_{M,IFN}$ were allowed to vary between strains.

Table 4.1. Maximum and minimum sensitivity of virus trajectory for all 15 parameters

Parameter	Max. Sensitivity (day)	Min. Sensitivity (day)
β	7.49 x 10^{14} (1.5)	0 (0.1)
λ	1.78 x 10^{-11} (4)	-8.89e-12 (1.4)
δ	0 (0.1)	-5.81 x 10^{6} (1.6)
K_{CTL}	0 (0.1)	-2.16 x 10^{11} (1.6)
p	0.041 (1.5)	0 (0.1)
q	0 (0.1)	-8.12 x 10^{9} (1.6)
c	0 (0.1)	-5.64 x 10^{6} (1.5)
$b_{IFN,M}$	0 (0.1)	-5.43 x 10^{6} (1.9)
mu_M	407.01 (1.9)	0 (0.1)
$b_{I,IFN}$	0 (0.1)	-30013 (1.6)
θ_{IFN}	13.11 (2.4)	-3.51 (4.7)
μ_{IFN}	4.11 x 10^{7} (1.9)	0 (0.1)
$b_{M,CTL}$	0 (0.1)	-132.25 (1.6)
μ_{CTL}	1.99 (1.6)	0 (0.1)
$b_{M,IFN}$	0 (0.1)	-30010 (1.6)

4.2.2.4 Prediction of additional strain-specific kinetics

The existence of multiple solutions is expected when fitting ODE models to limited data. To develop additional hypotheses on virus-specific host immune regulation, a third study assessed the ability of the model to capture influenza dynamics given no parameter restraints between strains. All model parameters were permitted to vary between H1N1 and H5N1 strains during estimation. Otherwise, the analysis used identical conditions (approximately 1,000,000 training instances; suitable models selected after parameterization burn in). This serves to identify if varying additional host-specific parameters could improve model fits or identify additional solutions that fit the virus-specific immune response. Fits for H1N1 and H5N1 dynamics are found in Figure 4.7a and b, respectively. Each fit is shown with the intervals containing 25-75% (dark grey) of all parameter sets and 5-95% (light grey) of all

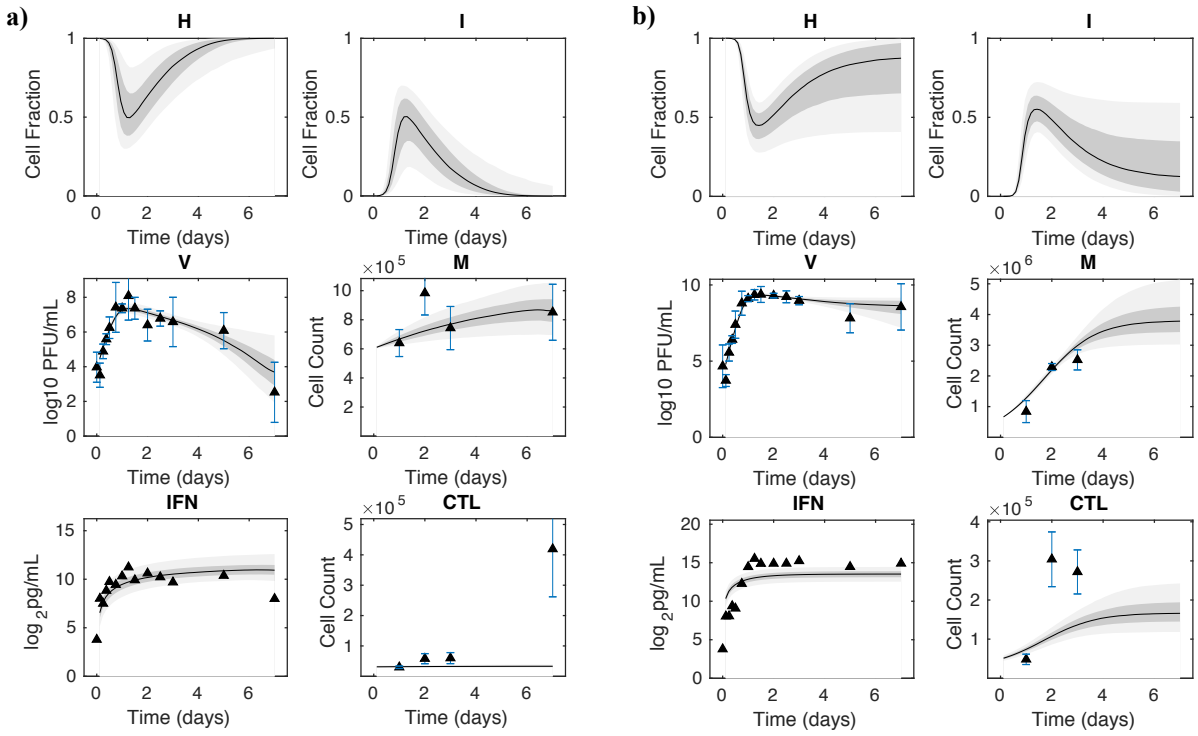


Figure 4.7. Model fits for a) H1N1 and b) H5N1 where all parameters were allowed to vary between strains. Fits are plotted against training data. Intervals in light grey and dark grey contain 25-75% and 5-95% of all parameter value sets, respectively.

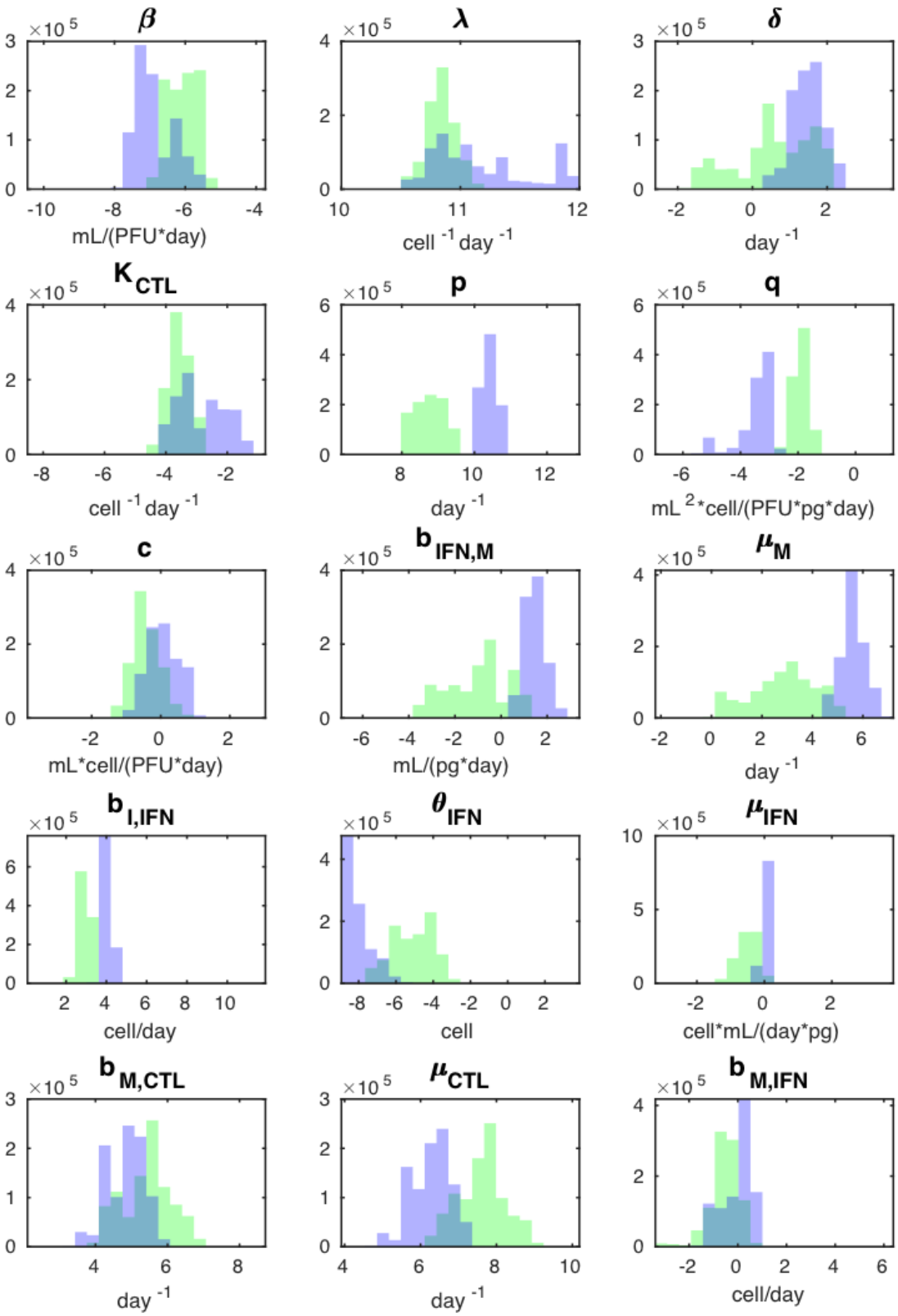


Figure 4.8. *Log*$_{10}$ parameter distributions for H1N1 (green) and H5N1 (blue) for the model where all parameters were allowed to vary between strains.

parameter sets. Overall, fits for both strains are comparable to those generated when only varying the virus-specific parameters (minimum error: 54.0025, section 4.2.2.2: 62.6403). While the H1N1 infected cell count similarly peaks between days 1-2, the magnitude of infected cells is greater (50% of cells infected) as compared to the fraction seen in section 4.2.2.2 (20% of cells infected). Virus, macrophage, and interferon are well fit to the data with tight parameter set distribution, however, the model is still unable to capture the steep increase in late H1N1 infection CTL behaviour. The H5N1 fits for infected cells peak around days 1-2 and while they do not return to zero, they do not display the same high level of infection through day 7 seen in section 4.2.2.2. Fits are in good agreement with data with similar variance in parameter sets as seen in section 4.2.2.2. Notably different, this model is the closest to capturing late stage CTL behaviour, though it still cannot achieve the magnitude of increase depicted in the data. Again, the model correctly predicts a more severe response to H5N1 infection.

Log_{10} parameter distributions for H1N1 (green) and H5N1 (blue) are shown in Figure 4.8. All parameters demonstrate differences between strains (P value $< 2x10^{-16}$, Kolmogorov-Smirnov test). When given total freedom, virus infectivity, β, viral depletion by interferon, q, and viral replication rate, p, retain similar distributions to those produced in the model from section 4.2.2.2. Viral clearance by antibodies, c, and the effect of macrophages on interferon production, $b_{M,IFN}$, show more similarity between strains in comparison to the model from section 4.2.2.2 where the strain distributions show no overlap. As supported in the sensitivity analysis, interferon-related parameters tend to have tight distributions.

4.2.3 Summary

In total, this study has assessed the ability of an ODE model of the influenza virus infection host immune response to accurately capture strain-specific dynamics. A simple model of the immune response was constructed to reproduce most of the dynamics identified in data sources during H1N1 and H5N1 infections when allowing only 5 parameters (those that control virus replication and interferon production by macrophages) to vary between strains. Further, the fits produced are similar in trajectory and timing to those produced

when all 15 parameters are permitted to vary. This comparable performance between models suggests that it is possible to successfully capture strain-specific influenza infection behaviour using the same generalized model with limited alterations to a small parameter set. Further exploration to minimize the size of this parameter set will advance the viability of this process moving forward into more generalized models. For example, while it is logical to allow for variance in viral replication rate, p, between strains physiologically, a sensitivity analysis reveals it has little effect on virus trajectory within the model. In the interest of simplification, the removal of p from the virus-specific parameter set should be explored.

The more severe response predicted in the H5N1 fit is reflective of disease phenotype. The model's viral state predicts a drop in viral load near day 6 in H1N1 infections and a sustained viral production time in H5N1 which agrees with trends identified in viral titer data. It is well recorded in literature that H5N1 infections cause increased damage to lung epithelial cells [263] and the strain is estimated to have a high mortality rate of 60% by the World Health Organization [264].

Two major areas of model development should be further explored. First, positive feedback occurring between macrophage and interferon production within these equations act as a major barrier to model stability. Reformulation of equations should be explored. Second, the model's inability to capture late infection CTL dynamics in all models suggests that the quality of the current CTL data is insufficient as equations are unsuited to capture the dynamics seen at later timepoints. Future work will address these issues and continue to analyse the involvement of host-specific parameters in eliciting a strain-specific response from a single ODE model.

4.3 Multi-strain parameterization of Innate Model

4.3.1 Materials and Methods

4.3.1.1 Model Creation

A deeper look into three commonly referenced models [63, 229, 66] which focus largely on interferon and epithelial lung cell populations reveals a common thread: dependence on epithelial cell states that cannot be measured *in vivo*. Further, despite their mid to large number of states, the models demonstrate mixed ability to respond to changes in interferon levels, suggesting an inadequate representation of the host immune response. Additional attention was given to the quality of available time course data of immune cells and cytokines. Many cells and cytokines share dynamic profiles, a particularly common occurrence amongst cytokines and the monocytes they recruit which transpire within feedback loops [54]. While a model of increased complexity serves as a more complete biological picture, the reality of larger models is parameterizations that often contain intractably sensitive and unidentifiable parameters. Lastly, experiments verify that CCR2+ infiltrating macrophage are at the root of pulmonary immunopathology and mortality [265]. MCP1 is the chemokine responsible for regulating the infiltration of macrophages to the site of infection [266], making it notable for its consequence within the model building process.

Given this collection of evidence, a novel three state ODE model of the innate immune response (Equations 4.8-4.10, parameter key Appendix E) was developed using current immunological knowledge of the early innate immune response to a primary influenza A virus infection (i.e. the animal's first exposure when they have no antibodies for the virus). A schematic of model interactions can be found in Figure 4.9. Virus (V) growth is logistic based on evidence from single-cell viral load florescence experiments [267]. Virus is removed two ways: at a rate $r_{IFN,V}$ corresponding to the activation level of the interferon response representing all downstream methods of viral removal within the immune response and a natural death rate, d_V. Type-I interferon (IFN) is produced at a rate, $p_{V,IFN}$, relative to viral load and decays naturally at rate d_{IFN}. Biologically, macrophages serve as the major producer of MCP1 [268] and display an *in vivo* time delay relative to

the initial interferon response. However, *in vivo* dynamics for MCP1 and macrophage are highly correlated, leading to parameter unidentifiability in early model building trials. Additionally, because available macrophage data is lacking day 5 timepoints, model training could not occur past day 3 for a macrophage state. As such, MCP1 was chosen as a representative state for the macrophage-derived inflammatory response. Several forms of the MCP1 production term were tested including mass action and logistic growth. However, none were able to find a parameterization that reproduced the steep drop in H1N1 MCP1 expression at day 5. Therefore, the production term is estimated as a Hill kinetic dependent with a natural decay rate, d_{MCP1}. Instead of the classic interpretation of the Hill Coefficient, n, as cooperativity in ligand binding [269], it can be interpreted in this context as an activation threshold representing the threshold of interferon needed to induce macrophage production of MCP1, similar to the activation threshold that must be exceeded to induce T cell cytokine production [270, 271].

$$\frac{dV}{dt} = kV(1 - V/K) - r_{IFN,V}V(IFN - IFN_0) - d_V V \tag{4.8}$$

$$\frac{dIFN}{dt} = p_{V,IFN}V - d_{IFN}(IFN - IFN_0) \tag{4.9}$$

$$\frac{dMCP1}{dt} = \frac{k_1(IFN - IFN_0)^n}{k_2 + (IFN - IFN_0)^n} - d_{MCP1}(MCP1 - MCP1_0) \tag{4.10}$$

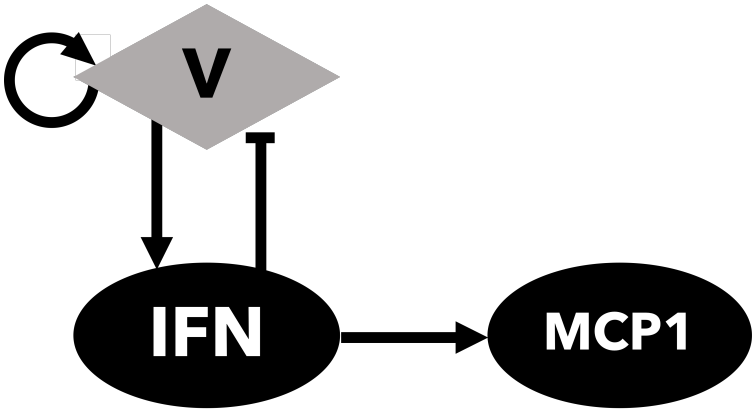


Figure 4.9. Innate Model Schematic.

4.3.1.2 Multi-strain MCMC Parameterization

The model was fit to *in vivo* data from Shoemaker et al. [54] derived from triplicate C57BL/6J mice infected with either the low pathogenic, seasonal A/Kawasaki/UTK-4/09 H1N1 virus (H1N1) or the high pathogenic, pandemic A/Vietnam/1203/04 H5N1 virus (H5N1). The viral state was fit to viral titers (PFU/mg), and while the MCP1 state was fit to protein expression data (ρg/mL), resultant interferon-β protein expression below the assay limits of detectability necessitated that the interferon state be fit to gene expression data. All training data used is the mean of triplicate values. Virus and interferon states are fit to log_{10} and log_2 values respectively and treated as absolute values to improve solvability and convergence. Initial state conditions were set to data values from time zero for model initialization. Lab-developed Parallel Tempering Markov Chain Monte Carlo (PT MCMC) methods were used as described in section 4.2.1.2 to determine parameter distributions for the ten model parameters, regardless of shared parameters, through optimization of the objective function:

$$Energy = \sum_{x=1}^{X} \sum_{t=0}^{T} \frac{(M_{x,t} - O_{x,t})^2}{2O_{x,t}} \tag{4.11}$$

The developed objective function was based on the sum of squares error: one half of the squared difference between model output, $M_{x,t}$, and observed data, $O_{x,t}$, for each state, x and timepoint, t, over all time points, T, for all species, X. Each time point was divided by the corresponding data $O_{x,t}$ to normalize error values. All simulations ran across six chains of temperature 0.99, 0.9, 0.8, 0.4, 0.2, and 0.05 to ensure adequate exploration of parameter space. For shared parameter studies, a modified MCMC routine continued to evaluate new parameters for H1N1 and H5N1 as described while ensuring that chosen shared parameter values were equal across strains before error calculations at each iteration. Mann-Whitney U tests were used to assess the statistical differences between the top 1,000 values of non-shared parameters to determine if the PT MCMC was sampling the same, often non-normal, distribution across both strains.

4.3.1.3 Model Selection

While an objective function conveys the quality of the fit achieved by parameterization, it is incapable of comparing models with varying numbers of estimated parameters because it fails to consider changes in degrees of freedom. Consequently, models with an increasing number of free parameters will result in a lower error and better fit but increase the likelihood of overfitting. The number of free parameters in a shared MCMC parameterized model is

$$P_{free} = P_{shared} + 2 * P_{independent} \tag{4.12}$$

where P_{shared} is the number of parameters shared across cohorts and $P_{independent}$ is the number of parameters that are independently estimated per cohort. As one of the benefits of the shared parameter MCMC is the ability to explore the similarities between cohorts through selected variance of P_{shared} and, therefore, the degrees of freedom of the system, the ultimate task of model selection must employ an equivalent metric to error with accounts for variations in the models' degrees of freedom. Akaike Information Criterion (AIC) is a comparison methodology for multiple competing models that explain the same data set. It is defined [272] as:

$$AIC = -2 * log_{10}(MaximumLikelihood) + 2 * P_{free} \tag{4.13}$$

AIC is based on a maximum likelihood estimate specific to the models being compared. This estimate is directly computed for linear regression fittings [273]. Other estimates must be used, when direct Likelihood functions are not available [274], especially in ODE and PDE systems. Monte Carlo maximum log-likelihood functions have been used previously in linear mixed-effect models to inform AIC [275]. Maximum Likelihood Estimate (MLE) for shared MCMC parameterization is calculated as:

$$MLE = exp(-Energy) \tag{4.14}$$

using the minimum values from the corresponding MCMC error function as defined in Equation 4.3.1.2. Much like error, a lower AIC value signifies a better model while a higher AIC value represents a worse model, however, there is a penalty for the addition of free parameters to limit the effects of overfitting.

4.3.1.4 Global Sensitivity

An extended Fourier Amplitude Sensitivity Testing (eFAST) global sensitivity analysis [276, 277] was performed in Julia Version 1.6.1 with the DiffEqSensitivity v6.44.2 package to determine the output variance of each state as a function of input variance to each parameter. The output of the method is first-order indices, S_p, which represents the variance explained by an input parameter, and total-order indices, S_Tp, which represents the high order interactions between parameters.

4.3.2 Results

4.3.2.1 Quantifying limitations of model goodness of fit

To assess the model's capability to capture the behavior of both strains individually, parameterization was performed allowing all parameters to vary between strains (no parameter sharing, DoF: 20). After 2 million iterations of MCMC fitting, the minimum energies for H1N1 and H5N1 were 2.17 and 1.16, respectively, with a total energy of 3.33 and an AIC value of 46.67. Energy values per iteration are shown in Figure 4.10A and B for H1N1 and H5N1, respectively. These values, referred to as "All Independent", are used as benchmark optimal fits for the remainder of the study. Qualitatively, model outputs are well fit to the data, displaying expected trends in magnitude and timing (Figure 4.11, black line). Standard deviation intervals of the top 1,000 solutions (i.e.: the 1,000 lowest energy parameter sets identified) are narrow, suggesting it is necessary that the biological rates represented by the parameter values operate in a small window of possible values to achieve known system behavior. Comparison of resultant top 1,000 parameter distributions, seen in Figure 4.12, across strain yields significant differences between distribution means (Mann-Whitney test P value $< .001$ for all parameters) indicating that MCMC chains are adequately exploring parameter space and suggesting the possibility of strain-specific kinetics in all represented biological functions.

For comparison, parameterization was performed under identical conditions while enforcing equal parameter values across strains for all parameters (DoF: 10). These "All

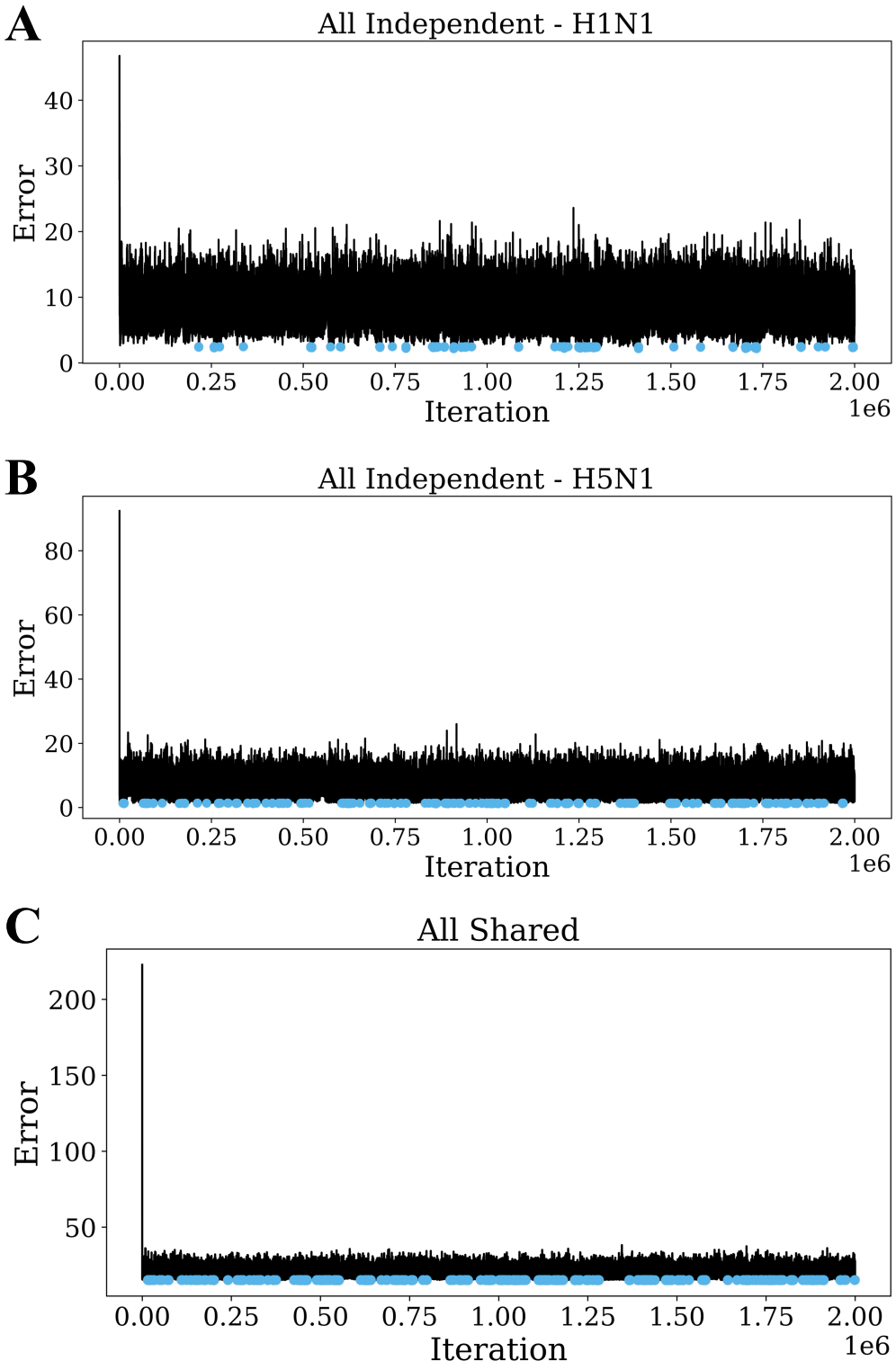


Figure 4.10. Energy values per iteration for A) All Independent H1N1, B) All Independent H5N1, and C) All Shared. The 1,000 lowest energies are marked in blue.

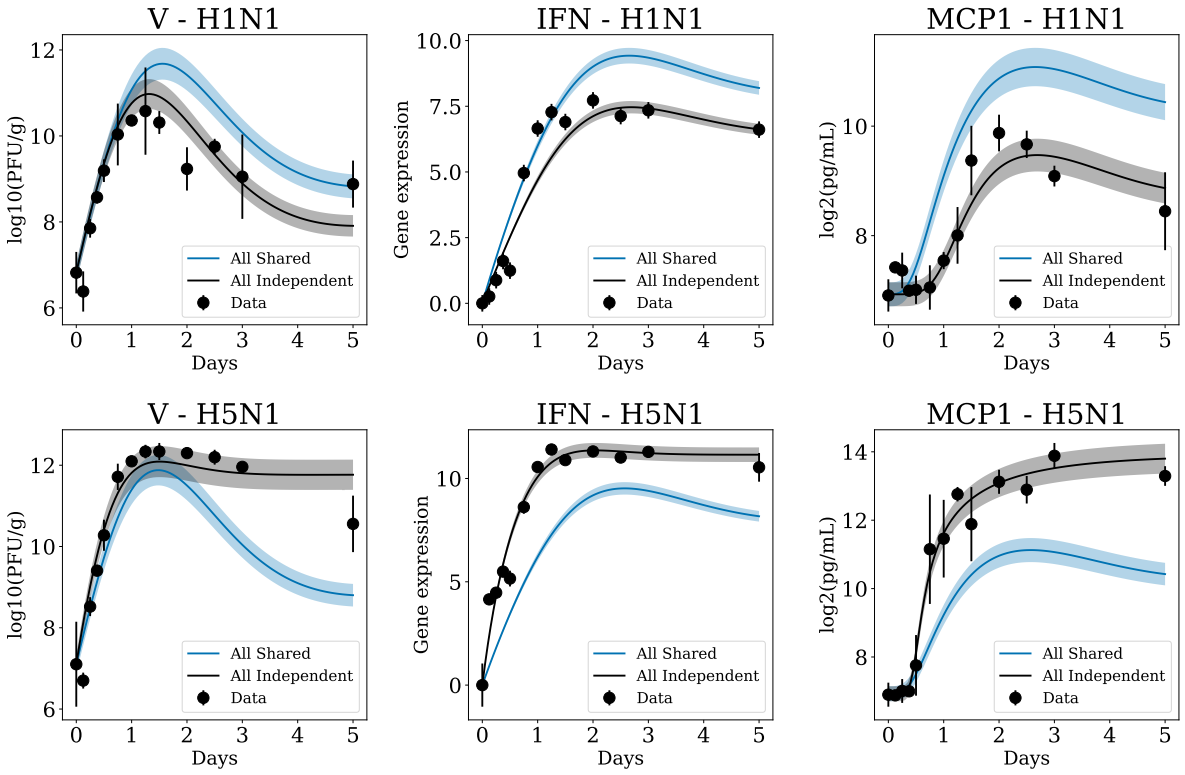


Figure 4.11. Model output for minimum energy parameter set (lines) and corresponding training data (markers) for H1N1 (top row) and H5N1 (bottom row). All Independent results (all parameters allowed to independently estimate across strains) are shown in black and All Shared results (all parameters shared between strains) are shown in blue. Intervals represent standard deviation of the 1,000 lowest energy parameter sets. Data markers are shown with the standard error associated with triplicate data points per timepoint.

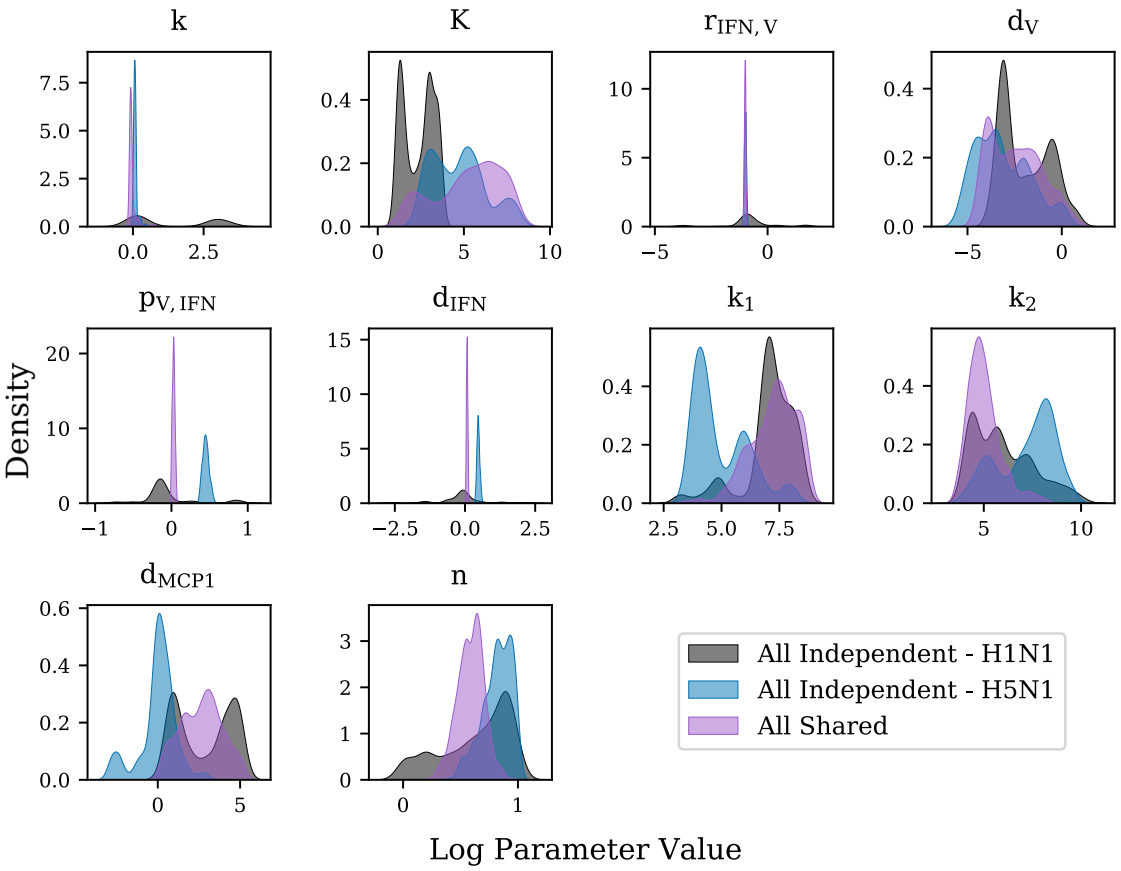


Figure 4.12. Parameter distribution density plots for All Independent H1N1 (black), All Independent H5N1 (blue), and All Shared (purple) resulting from 1,000 lowest energy solutions of each study.

Shared" fits result in a total minimum energy of 15.04 (Figure 4.11, blue line). Energy values per iteration are shown in Figure 4.10C. The associated AIC value, 50.08, suggests that the All Shared parameterization is worse than that of the All Independent. Qualitatively, resultant fits are non-representative of trends and/or magnitudes of experimental data, proving that strain-specific initial state conditions are not enough to reproduce observed dynamics. The top 1,000 parameter sets, purple in Figure 4.12, have significantly different means when compared to all parameter distributions for both H1N1 and H5N1 (Mann-Whitney test $P < .001$). In combination, these results highlight the extrema of the shared parameterization method for the system and serve in contrast to the exploration of shared parameter combinations.

4.3.2.2 Strain Independence in Interferon Production Produced Best Fit with Single Parameter Freedom

Assuming there are parameters where shared values are appropriate across strains, the next step was a relaxation of the All Shared parameterization. For each model parameter, p, 2 million iterations of MCMC parameterization were performed where all parameters were shared across strains with the exception of p (DoF: 11). When compared to the benchmark All Shared and All Independent studies, the contribution of each parameter to observed *in vivo* strain-specific dynamics can be elucidated. Minimum energy solutions for these "Solo Independent" parametrizations of all ten model parameters are found in Figure 4.13 for H1N1 and H5N1 with corresponding energies and AIC values in Table 4.2. Most solutions resemble the undesirable trends seen in the All Shared results of Figure 4.11 including the inability to capture viral titers post day 2 or the stark contrast in day 5 levels of MCP1 between strains. Comparison of resultant top 1,000 parameter distributions across strain yields significant differences between distribution means (Mann-Whitney test $P < .001$ for all independently estimated parameters).

Minimum combined energy values fall between 9 and 13 with the exception of $p_{V,IFN}$, which yields a minimum energy of 6.65. While this is the closest to the combined All Independent energy (3.33) by far, AIC calculations reveal that the resulting value of 35.30

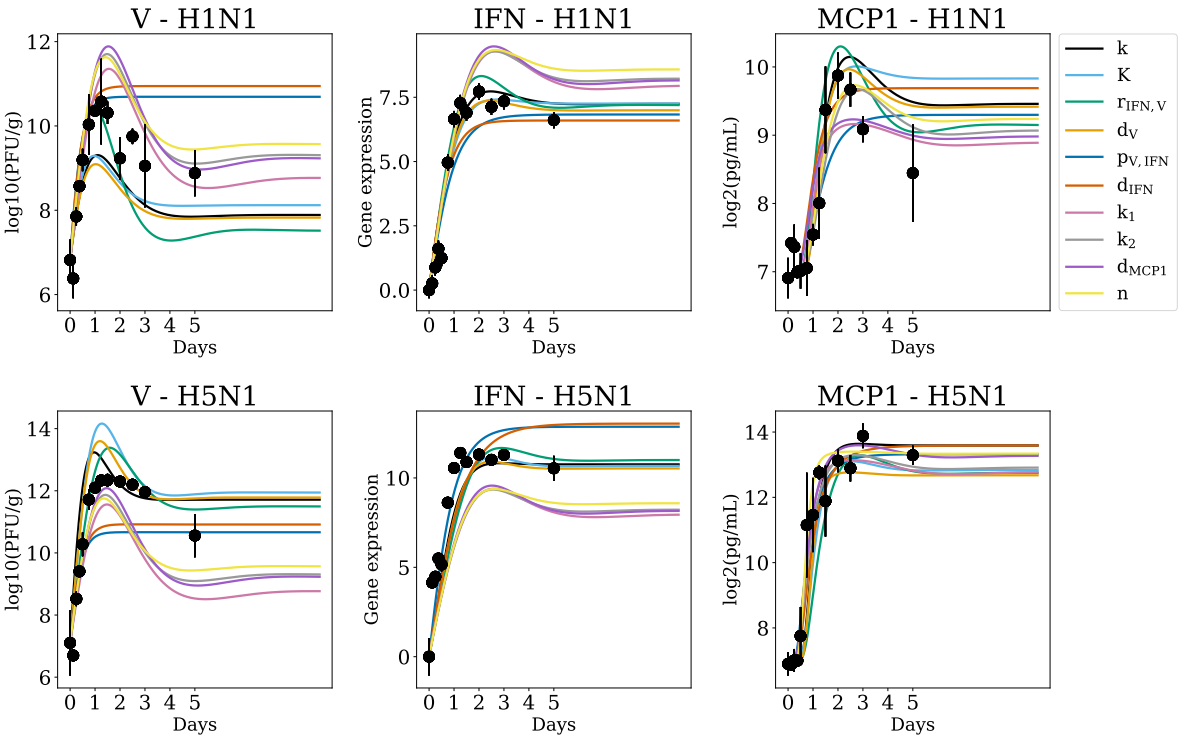


Figure 4.13. Model output for the minimum energy parameter set (line) for solo independent parameterizations and corresponding training data (markers) for H1N1 (top row) and H5N1 (bottom row). Data markers are shown with the standard error associated with triplicate data points per timepoint.

Table 4.2. Minimum error and AIC values for All Independent, All Shared, Solo Independent, and Virus Independent studies. V is representative of four viral parameters: k,K,$r_{IFN,V}$, and d_V.

	Independently Estimated Parameter	**Minimum Error**	**AIC Value**
All Independent	$AllParameters$	3.33	46.67
All Shared	$None$	15.04	50.08
Solo Independent	k	9.37	40.74
	K	9.79	41.59
	$r_{IFN,V}$	10.83	43.66
	d_V	9.65	41.31
	$p_{V,IFN}$	6.65	35.30
	d_{IFN}	10.30	42.61
	k_1	12.36	46.73
	k_2	12.29	46.57
	n	12.28	46.57
	$d_{MCP1,IFN}$	12.37	46.75
Virus Independent	V	9.34	46.68
	$V + p_{V,IFN}$	5.55	41.11
	$V + d_{IFN}$	8.38	46.75
	$V + k_1$	8.86	47.72
	$V + k_2$	8.89	47.79
	$V + n$	8.92	47.85
	$V + d_{MCP1}$	8.89	47.78

for $p_{V,IFN}$ is not only lower than the results of the other nine parameterizations but is lower than that of the Independent study (AIC 46.67). In fact, all Solo Independent studies for parameters of the virus and interferon state equations report AIC values less than that of the Independent study. While the energy calculation values goodness of fit via the distance between simulation and training data, AIC calculation, which analyzes not only parameterization error via likelihood estimate but the impact of the degrees of freedom, suggests that the All Independent parameterization is likely overfitting the model where the same information can be captured using a model of lower degrees of freedom.

4.3.2.3 Independent Estimation of Virus Parameters per Strain Does Not Improve Model Fits

We hypothesized that disparate dynamics between viral strains may be related to virus-based rates such as growth rate, k, or death rate, d_V. To test this, a "Virus Independent" parameterization was performed allowing only viral state parameters k, K, $r_{IFN,V}$ and d_V, denoted as parameter subset V, to independently explore parameter space per strain (DoF: 14). Six additional "Virus-host Independent" parameterizations were performed with the addition of one of the non-viral state parameters to set V (DoF: 15). Qualitatively, the resulting fits are more indicative of the expected dynamic trends, including for day 5 of infection.

The minimum energy solutions for each parameterization are found in Figure 4.14. Corresponding minimum energy and AIC values are found in Table 4.2. Comparison of resultant top 1,000 parameter distributions across strain yields significant differences between distribution means with the exception of $r_{IFN,V}$ in the $V + k_2$ study (Mann-Whitney test $P < .001$ for all independently estimated parameters). Minimum energies associated with Virus-Host Independent parameterizations are lower than that of the Virus Independent equivalent with a marked improvement in the minimum energy, 5.55, associated with independent fitting of V and $p_{V,IFN}$ against both the Virus-Host

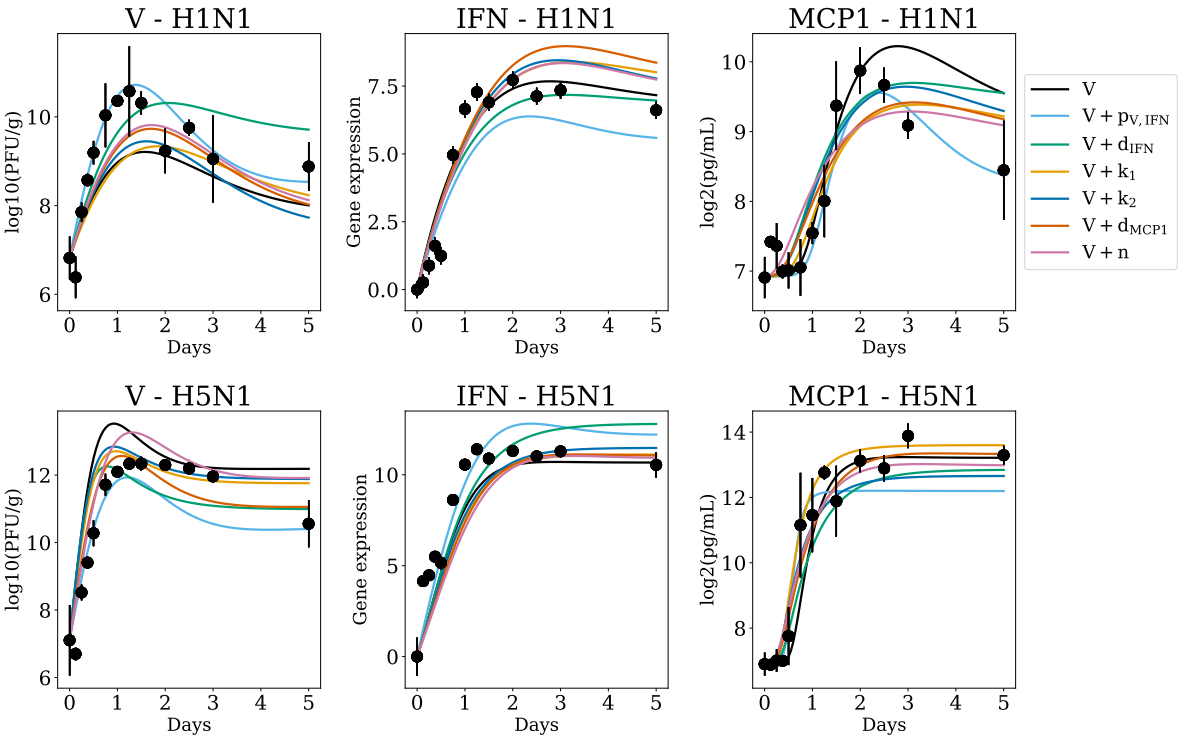


Figure 4.14. Model output for minimum energy parameter set (line) for virus independent parameterizations and corresponding training data (markers) for H1N1 (top row) and H5N1 (bottom row). V is representative of four viral parameters: k, K, $r_{IFN,V}$ and d_V. Data markers are shown with the standard error associated with triplicate data points per timepoint.

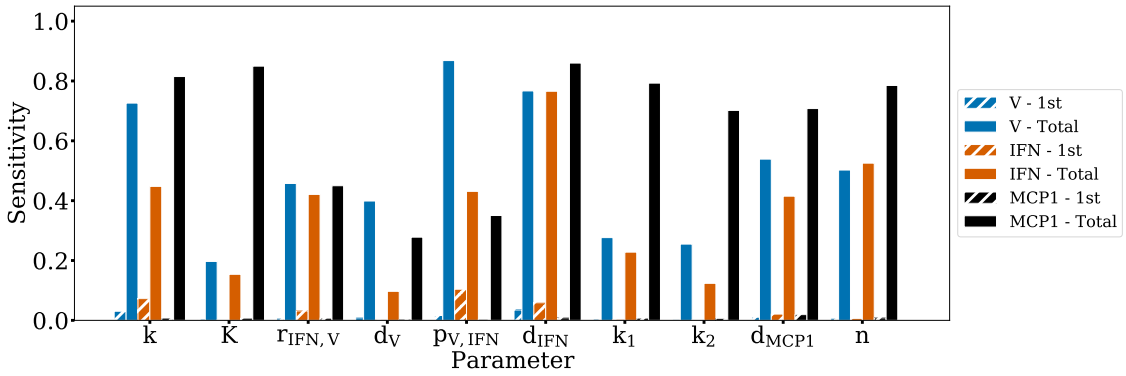


Figure 4.15. First- (dashed) and total-order (solid) indices of the eFAST sensitivity analysis. Indices are reported per model state for each parameter.

Independent and Solo Independent study for $p_{V,IFN}$. Comparing AIC values to baseline values again reveals that Independent estimation of subset V and $p_{V,IFN}$ results in a lower AIC value than the All Independent study.

4.3.2.4 Viral State is Highly Sensitive to Interferon Parameters

The eFAST algorithm was used to quantify overall model sensitivity in the form of fractional variance that can be attributed to model parameters and their interactions. The first-order indices, representing the output variance explained by variance in parameter input p, and total-order indices, representing nonlinear interactions between parameter p and other model parameters, are found in Figure 4.15.

First-order indices remain close to zero for all parameters across states, with the maximum first-order index of 0.1 for the effect of $p_{V,IFN}$ on interferon output, asserting that higher-order, nonlinear interactions between parameters are driving the observed output variance of the model. Total-order sensitivity is much higher for interferon and MCP1 in comparison with many values resting between 0.7 and 0.8, which is unsurprising given the Hill kinetics of the MCP1 state rely only on interferon dynamics. Interferon parameters drive high variance in the viral state, indicative of the positive feedback loop

between viral growth and interferon production. This trend is most clearly seen in the sensitivity of the viral state as a function of variance in $p_{V,IFN}$, the interferon production rate, reiterating the importance of interferon production rate as a driver of overall system outcomes.

4.3.3 Summary

Here, a three state ODE model of the innate immune system investigates the mechanistic roots of differential immunoregulatory behavior observed *in vivo*. This study is made possible by the iterative application of an MCMC parameterization approach to explore the concept of shared parameters across strains as a method of mathematically elucidating the most likely biological drivers of strain-specific behavior in an unbiased manner. The production rate of interferon within the virus-interferon positive feedback relationship, parameter $p_{V,IFN}$, is identified as a major driver of strain-specific dynamics observed in the early innate immune response to highly pathogenic (H5N1) and low pathogenic (H1N1) influenza A virus. Global sensitivity highlights the importance of virus-interferon feedback as the majority of output variance is the result of high order parameter interactions with interferon parameters $p_{V,IFN}$ and d_V being responsible for a large degree of the variance seen in virus production. With broader impact to the field of cohort modeling as a whole, AIC model selection methods reveal that allowing all parameters to explore parameter space simultaneously is detrimental to model performance and asserts that the rates of many biological mechanisms are not significantly strain dependent.

The identification of the virus-interferon feedback, particularly in the context of strain-specific behavior, aligns well with experimental evidence of strain-specific characteristics. HA and NA, the influenza virus envelope proteins which dictate strain characterization, are known regulators of type-I interferon production [254, 255, 256] while H5N1 NS1 protein is noted for its effect on the timing and overall efficacy of the interferon response. One or some combination of these protein mechanisms may be at the root of the identified strain-

sensitivity within the model. Further exploration training the model to additional or mutated strains may prove useful for pinpointing the biological source.

The inequity of trends in model energy and AIC values for the varying levels of shared parameters reveals interesting information about the performance of the constructed model. While energy, a normalized measure of the distance between simulation output and training data, follows a trend of increasing accuracy with increasing degrees of freedom, AIC predicts that the optimal model shares 90% of parameters between strains. As AIC is a measure of both prediction energy and model quality by imposing a penalty for each degree of freedom, results suggest that the All Independent study is overfitting to *in vivo* data. Our hypothesis that disparate kinetics could be explained by viral parameters alone was a logical assumption. However, AIC suggests that variance of a single parameter (Solo Independent parameterizations, DoF: 11) of the virus and interferon states provide a better model than both allowing all parameters to vary (All Independent, DoF: 20) and allowing virus parameters to vary (Virus Independent models, DoF: 14 or 15) with the exception of $V + p_{V,IFN}$. Discovering that the cost of imposing equivalence between three of the four virus parameters between strains outweighs the information lost by allowing freedom in the estimation of all viral parameters at once suggests that viral strains likely share kinetic characteristics. This serves as further proof that the initial hypothesis was incorrect: strain-specific differences likely arise from host-virus interactions and their effects.

Modelers face several limitations and choices in the process of creating mathematical models. First, working with averaged triplicate time course data comes with challenges like missing values and high standard error in the event of the heterogeneity often observed in biological systems. Additionally, data derived from methods such as protein assays often return values below the limit of detection for some or all data points. While alternative measures such as gene expression can be used in place of protein expression as carried out here for training the interferon state, the question of correlation between the two datasets must be acknowledged. Decisions surrounding data transformation are often made to weigh model accuracy and feasibility. For example, data used to train the virus and interferon states are logged values (log_{10} and log_2 respectively) while treated as absolute measures to eliminate problems of mathematical solvability. Lastly, systems where multiple states

are training to data with similar dynamic trends often have parameter identifiability issues that affect the ability of the model to accurately perform. The freedom in these combined decisions remind us that the field of ODE modeling is a constant work in progress, with everchanging best practices and methodologies for optimal model development.

Under these considerations, the deployed combination MCP1-macrophage state representative of the inflammatory response invoked by macrophage activity balances complexity with parameter identifiability concerns. Within the framework of n serving as an activation threshold needed to induce macrophage production of MCP1, resulting values of 5.47 and 9.98 for H1N1 and H5N1, respectively, imply that a higher rate of interferon production is reached in H5N1 infection before the cytokine response is upregulated.

5.0 Conclusions

In aggregate, this dissertation uses classic controls engineering and systems biology modeling tools to elucidate the foundational biological mechanisms of the host immune response driving clinical outcomes in influenza A virus and SARS-CoV-2 infections. Immunoregulation is studied at two levels: a protein level analysis to identify host factors of disease through the creation of two novel PPI network analysis pipelines and a systems level exploration of strain-specific dynamics using MCMC parameterization of ODE models. The resulting analyses hold potential in focusing drug development efforts on the most likely targets and improving patient outcomes. The developed methodologies are widely applicable in future studies of other diseases.

5.1 Disease-specific Subnetwork

The formation and analysis of biological networks from high throughput data such as proteomics yields an advantageous evaluation of the relationships between cellular pathways. Most notably, PPI networks remain unburdened by the need for extensive experimental characterization of protein interaction kinetics that hinder the development of more quantitative modeling methods. Topological analysis, such as degree and betweenness, and the clustering behavior of biological networks remain as foundational systems biology tools for identifying the most important components of the network, however, highly depend on the construction of the network itself. As many efforts utilize whole cell interaction repositories, methods are needed to address the lack of disease-specificity seen in network biology studies.

Aim 1 addresses this gap with the first integration of viral-human protein interactions into the human protein interaction network to create a disease-specific PPI network for influenza A virus. Virus interacting host proteins are found to be closely clustered within the network and are often members of protein complexes, suggesting that influenza virus has

evolved to manipulate specific pathways and mechanisms of the cell to advance the ability to manipulate host machinery. This aligns with previous evidence of clustering in host factors of viral replication [278]

. Further, with additional siRNA data to identify proteins that are vital to viral replication tasks, we create a disease specific subnetwork representative of the pathways activated between virus interacting host proteins and critical host mechanisms used to advance infection. The subnetwork is more highly enriched for host factors of influenza A virus identified in six previous partial genome siRNA screens than the virus interacting protein set and the total network asserting that the subnetwork is a better predictor of disease host factors than the traditionally studied virus interacting proteins. The proteins that are unconfirmed as host factors should be tested in future screens. Several of the highest betweenness proteins of the total network display much lower values in the subnetwork, revealing discord between topological estimation of protein importance in the total cell landscape versus the importance to disease systems of interest. Functionally, the subnetwork is comprised of a distinct group of proteins implicated in immune functions such as NF-kB pathways in contrast to the typically studied virus interacting proteins.

The observation that betweenness does not significantly improve host factor prediction suggests that alternative topology measures should be considered for future studies. There are several reasons why betweenness was selected. Biological pathways are known to have several alternative routes to maintaining cellular operations, a key feature of biological robustness [279, 280, 281]. Biological networks are also theorized to have a bow tie-like structure that suggests a natural bottlenecking within the PPI network near critical cellular machinery [282]. These concepts together suggest targeting bottlenecks (high-betweenness proteins) as a means of mitigating escape via alternative paths. In future work, other network topology measures (e.g., degree, Burt's Constraint, or closeness) could be tested in the subnetwork and subnetwork construction and could be varied to consider different subsets of either the virus interacting proteins or the internal-essential host factors. Even so, the results suggest that construction of the virus-specific subnetwork provides major advantages in host factor discovery and can significantly expand drug candidate repertoires beyond virus interacting proteins. Furthermore, since the connecting

proteins do not directly interact with the virus, they may be more resistant to concerns related to drug-mediated selective pressure.

This analysis proves the ability of a disease-specific subnetwork to narrow the field of network biology to a scope that replicates the specificity of the host response to disease. The proteins identified in this method should be prioritized for drug target development and repurposing.

5.2 Controllability of Virus-Host PPI Networks

While the virus-host network delivers disease specific insights, alone it represents only a static snapshot of a highly dynamic system. Complex behavior within the cell over the course of infection cannot be ignored in determining the causes of cellular dysregulation. A comparison of healthy and infected cell snapshots achieves a more dynamic understanding of the disease system without requiring extensive, time course data used in mathematical modeling. Instead, the focus rests on the differences between the two networks where the observed changes are disease specific. Controllability provides a framework to identify which proteins are responsible for driving total system behavior, a process that contributes to the overall understanding of disease needed to advance drug development.

In Aim 2, controllability methods were applied to both influenza A virus and SARS-CoV-2 infections. For influenza, network-wide increases in topology post infection reveal wide-reaching effects of viral control on cellular behavior during infection, particularly in proteins that are both virus interacting and driver proteins. Controllability results reveal two general conclusions about the roles of these protein groups. First, viruses are more likely to interact with host proteins that offer a greater advantage in guiding total cell behavior. Additionally, driver proteins act as an obstacle in gaining control of the system as they require additional manipulation. Over 75% of driver proteins are able to act as non-driver nodes given certain control configurations, lending to the idea of viral escape routes: alternative pathways of gaining control under suboptimal conditions.

Surprisingly, robust controllability does not identify varying protein regulation between the healthy and infected cell (all classifications are the same for the HIN and VIN). However, global controllability identifies a set of 24 proteins that exhibit notable changes between the HIN and VIN. An increase in topological relevance post-infection implies that the controllability-predicted proteins' importance to the cell is unique to infection. 12 of the 24 proteins are experimentally validated as interferon regulated genes centered around the NF-kB pathway that is known to be inhibited by influenza virus [283, 201]. Though the set of 24 proteins are not significantly enriched for host factors of influenza A replication in past siRNA screening studies, these experiments test only part of the genome by design, meaning the search must be active. In total, these 24 proteins hold topological, functional, and controllability relevance and are recommended for drug targeting efforts.

An analysis of the SARS-CoV-2 virus-host network in comparison to the healthy host network reveals many of the same topological and controllability insights; the majority of controllability classifications remain the same for both viruses. However, the predicted regulators of the cell demonstrate low overlap with those identified for influenza A virus (3/16 predicted regulators), proving that the method is a predictor of disease specific proteins. The changes to controllability classification of the 16 predicted regulators indicate a loss of control that could represent viral interruption of normal host function or activation of a new pathway. As in the analysis of influenza, the majority of predicted proteins are topologically more important to the infected cell system. The usefulness of the developed controllability pipeline was proven in the timely and effective prioritization of drug targets that could be repurposed for COVID-19. Eight controllability-identified proteins are registered as pre-existing drug targets [188, 168], six of which were identified by other groups within the context of SARS-CoV-2 virus and three of which have targeting compounds that are experimentally verified as downregulators of SARS-CoV-2 replication. Relevance to viral infection including the cellular stress response, protein translation, and cellular transport further prove the disease specificity of the predicted proteins. Five compounds of interest targeting PVR, SCARB1, and NUP210 are recommended for viral inhibition studies for the first time. An animal model is needed to assess the potential of

vitamin E and sphingosine-1-phosphate for SARS-CoV-2 viral inhibition, a promising lead as they are known inhibitors of influenza A viral activity [214, 220].

A major advantage of the pipeline developed in this aim is that the generality of this analysis can be used to improve the prediction of drug targets for any pathogen-host interaction given available protein interaction data. The limits of these methods lie in limited availability of large-scale, dependable databases of protein-protein interactions. Foundational maximum matching algorithms for the calculation of driver proteins must be performed with directed networks. While larger directed networks than the network from Vinayagam et al. [183] are available [284], the network used here contains only experimentally derived data opposed to computationally predicted interactions, assuring biological confidence in the results within this study. A robust controllability analysis of the computationally predicted network presented in Uhart et al. [284] finds that 29% of proteins are categorized as indispensable where approximately 20% of proteins in the Vinayagam network are classified as the same, though there is 89% overlap in directed edges between the two networks. This suggests that methods for predicting protein interactions may over represent these key proteins within the analysis, even in combination with experimental results. However, larger networks will move towards a more complete analysis of infected cell behavior and possibly reveal further proteins of interest. Therefore, the future of this field depends on continued establishment of large, confident, directed PPI networks.

5.3 ODE Modeling of Host Immune Response

Experimental quantification of cytokine and immune cell counts along with distinct differences in clinical presentation make it clear that the invoked immune response for influenza A virus may be highly strain specific [54]. H5N1, a severely pathogenic strain, elicits higher weight loss, inflammation, tissue damage, and viral titers when compared to H1N1, a low pathogenic, seasonal strain. However, despite demonstrated differences in presentation, the disease mechanisms underlying these observations are unknown.

Mathematical models of immune response dynamics are uniquely positioned to discern the biological processes responsible for the differences identified in strain specific behavior. However, model parameters have always been trained to data from a singular viral strain. Aim 3 contains a computational assessment of differential immunoregulation to identify the biological drivers of strain-specific behavior and determine if current modeling practices are suitable to predict differences in disease pathology observed *in vivo* between varying strains. A review of three published models of influenza virus infection found that recent models are unable to reproduce expected interferon trends and heavily depend on the infected cell state to downregulate viral replication. These results demonstrate the failure of current modeling standards to sufficiently incorporate the interferon pathway, knowledge that guides future model development efforts.

Two models of varying complexity were trained simultaneously to data from low pathogenic H1N1 strain and the highly pathogenic H5N1 strain of influenza A virus. The first model is representative of both the innate and adaptive immune response. Parameterization of the model found that strain-specific estimation of virus-related parameters was not sufficient to produce differential immunodynamics seen across strains. The addition of parameter $b_{M,IFN}$, the effect of the presence of macrophages on interferon production, resulted in simulations comparable to those observed *in vivo*, suggesting the surrounding mechanisms can be seen as a contributor to strain-specific behavior. This may be the result of evidence that H5N1 infected resident macrophages prompt an increased interferon response [262]. Further, a sensitivity analysis reveals that some virus-related parameters such as viral replication rate have little effect on strain-specific virus trajectory during infection, suggesting that only certain viral characteristics dictate strain behavior.

Driven by the inability of the first model to reproduce late infection CTL counts and the error introduced to the system by the lack of training data for epithelial cells, a pared down model was used to re-explore the strain specificity of the innate immune response. MCMC parameterization finds no well fit solution when forcing all parameters to be equal between H1N1 and H5N1 fits asserting that there are strain-specific mechanisms represented within the three-state model. Testing strain independence in each parameter alone, and in combination with the four virus parameters revealed the importance of the interferon

response to the differential immunodynamics. AIC was used to compare the quality of each parameterization to account for the varying number of shared parameters across strains. Strain-dependent estimation of parameter $p_{V,IFN}$, the production rate of interferon, yields the optimal fit of the parameterizations explored, again asserting that host dynamics may be at the root of observed differences. Comparing the rankings of error and AIC reveals that while independent estimation of all parameters across strains results in the lowest error values, the cost incurred by the additional degrees of freedom is high resulting in the lowest AIC ranking. Biologically, this suggests that mechanistic differences in the host immune response are minimal in number. A global sensitivity analysis supports this idea showing minimal sensitivity to changes in parameters themselves. Instead, state variance can be attributed to the interactions between parameters, particularly parameters of the interferon state that contribute largely to the viral state. In total, results support the attribution of strain-specific immunodynamics to mechanisms of interferon production.

The three-state model could be expanded to include additional states as long as identifiability remains intact. Alternatively, the addition of parameters that are representative of immune functions affecting the virus, interferon and MCP1 states could be integrated into the existing equations such as the induction of interferon in infected macrophages that has been noted as a marker of severity in H5N1 infection specifically [285]. The shared parameter MCMC framework is highly generalizable to other cohorts of data including age or sex specific studies, making it a highly valuable tool for investigating disparate kinetics between groups of interest and, ultimately, the drivers of observed clinical behavior. Additionally, conclusions from cohort specific studies may prove useful for informing and simplifying future modeling work with additional cohorts. For example, if modeling efforts turn toward an emergent influenza A virus strain in the future, one may hypothesize that the parameters shared between other influenza strains may hold true. Investigation may support this hypothesis, saving time and effort, or prove the hypothesis false, indicating a mechanistic difference in the emerging strain. In total, the results enclosed serve as a starting point model of the early innate immune response and a prototype for powerful comparative studies.

The results of these three aims offer a greater understanding of the immunodynamics of viral infection and contribute novel tools to the field of systems biology. The power of computational methods lie in the ability to invent and build on existing methodologies of processing big data to efficiently extract meaningful biological conclusions. Through collaborative work with immunologists and experimental biologists as well as computer scientists and bioinformaticians, the future promises advances in the treatment of disease to the great benefit of global public health.

5.4 Publications Resulting from this Dissertation

- Ackerman, E. E. et al. Network-Guided Discovery of Influenza Virus Replication Host Factors. MBio (2018). doi:10.1128/mBio.02002-18
- Mochan, E., Ackerman, E. & Shoemaker, J. A Systems and Treatment Perspective of Models of Influenza Virus-Induced Host Responses. Processes (2018). doi:10.3390/pr6090138
- Ackerman, E. E., Alcorn, J. F., Hase, T. & Shoemaker, J. E. A dual controllability analysis of influenza virus-host protein-protein interaction networks for antiviral drug target discovery. BMC Bioinformatics 20, 297 (2019).
- Ackerman, E. E., Mochan, E. & Shoemaker, J. E. Strain-Specific Immune Response to Influenza Virus Infection. in IFAC-PapersOnLine 52, 101106 (Elsevier B.V., 2019).
- Ackerman, E. E. & Shoemaker, J. E. Network controllability-based prioritization of candidates for SARS-CoV-2 drug repositioning. Viruses 12, 1087 (2020).
- Ackerman, E. E., Weaver, J. A., & Shoemaker, J. E. Disparate Interferon Production Rate Drives Strain-Specific Immunodynamics of Influenza A Virus. (In preparation).

Appendix A Scale Free Network

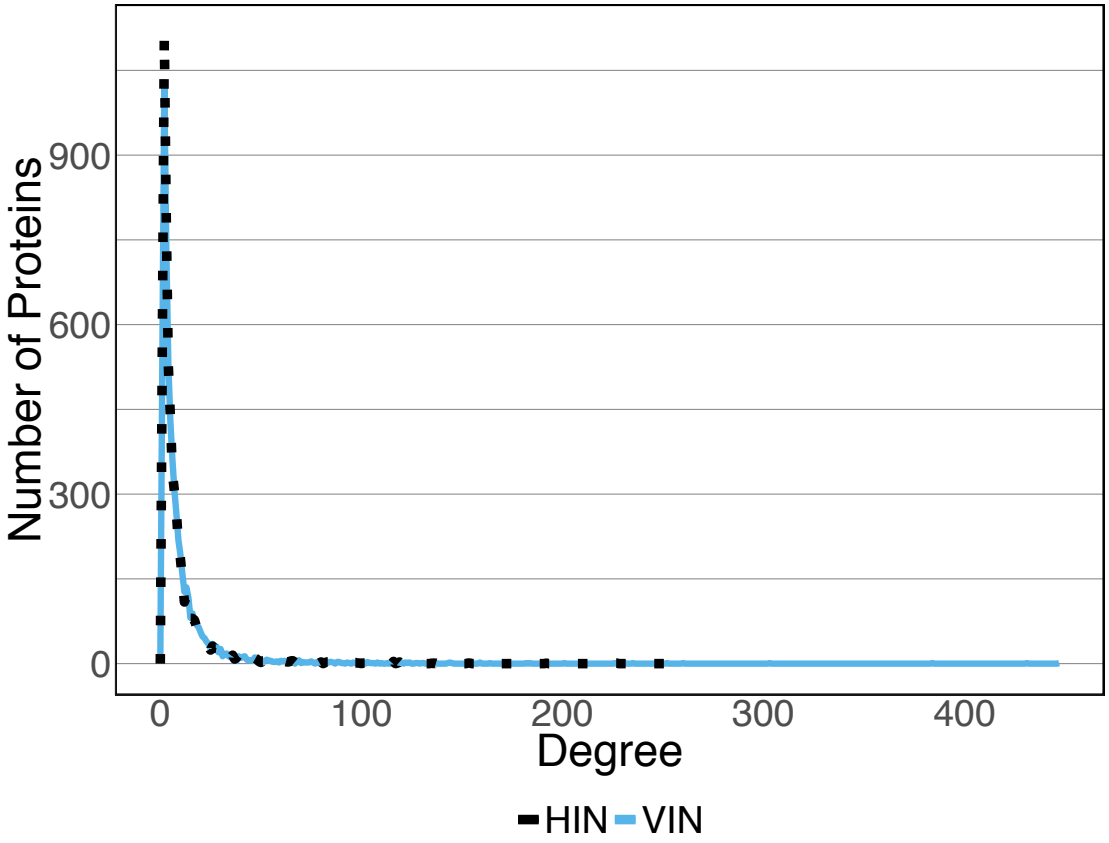


Figure A1. Degree distribution of network proteins

Appendix B Drugs Targeting Prioritized Targets for COVID-19

Table B1. The targets and associated compounds for all eight genes prioritized for COVID-19 treatment.

Compound Name	***Target Gene***	***Drug Status***	***Identified by Drugbank /research associations***	***Identified in Gordon et al (2020) (**active viral inhibition, *untested)***
Ternatin 4 (DA3)	EIF4AE	Pre-clinical		X**
Zotatifin (eFT226)	EIF4AE	Clinical Trial		X**
4E2RCat	EIF4AE	Pre-clinical		X
4E1Rcat	EIF4AE	Pre-clinical		X
7-methyl-GpppA	EIF4AE	Experimental	X	
7-methyl-7,8-dihydroguanosine-5'-diphosphate	EIF4AE	Experimental	X	
7-methyl-guanosine-5'-triphosphate	EIF4AE	Experimental	X	
LY2275796	EIF4AE	Investigational	X	
S-[(1-Hydroxy-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl] methanesulfonothioate	EIF4AE	Experimental	X	

Table B1. (Continued)

Compound Name	***Target Gene***	***Drug Status***	***Identified by Drugbank /research associations***	***Identified in Gordon et al (2020) (**active viral inhibition, *untested)***
Rapamycin (Sirolimus)	EIF4AE, LARP1	Approved	X	X
Mycophenolic acid	IMPDH2	Approved	X	X**
Ribavirin	IMPDH2	Approved	X	X
Merimepodib	IMPDH2	Clinical Trial		X
Mycophenolate mofetil	IMPDH2	Approved	X	
NADH	IMPDH2	Approved	X	
Selenazole-4-carboxyamide-adenine dinucleotide	IMPDH2	Experimental	X	
6-Chloropurine Riboside, 5'-Monophosphate	IMPDH2	Experimental	X	
Inosinic Acid	IMPDH2	Experimental	X	
VX-148	IMPDH2	Investigational	X	
Azathioprine	IMPDH2	Approved	X	
RapaLink-1	LARP1	Pre-clinical		X
Sapanisertib (INK128/MlN128)	LARP1	Clinical Trial		X
Selinexor	NUP210, RAE1	Approved		X*

Table B1. (Continued)

Compound Name	***Target Gene***	***Drug Status***	***Identified by Drugbank /research associations***	***Identified in Gordon et al (2020) (**active viral inhibition, *untested)***
Myristic acid	PVR	Experimental	X	
Sphingosine	PVR	Experimental	X	
Phosphatidyl serine	SCARB1	Approved	X	
Tocopherol	SCARB1	Approved	X	
alpha-Tocopherol acetate	SCARB1	Approved	X	
Tocofersolan	SCARB1	Approved	X	
Vitamin E	SCARB1	Approved	X	
PHA-665752	SCARB1, NUP210	Experimental	X	
E-52862	SIGMAR1	Clinical Trial		X
RS-PPCC	SIGMAR1	Pre-clinical		X
PD-144418	SIGMAR1	Pre-clinical		X**
Dextromethorphan	SIGMAR1	Approved	X	X
Haloperidol	SIGMAR1	Approved	X	X**
PB28	SIGMAR1	Pre-clinical		X**
Siramesine	SIGMAR1	Clinical Trial		X**
Cloperastine	SIGMAR1	Approved		X**
BD1008	SIGMAR1	Pre-clinical		X
Carbapentane	SIGMAR1	Approved		X
Ifenprodil	SIGMAR1	Approved		X

Table B1. (Continued)

Compound Name	***Target Gene***	***Drug Status***	***Identified by Drugbank /research associations***	***Identified in Gordon et al (2020) (**active viral inhibition, *untested)***
Progesterone	SIGMAR1	Approved		X
Clemastine	SIGMAR1	Approved		X**
Hydroxychloroquine	SIGMAR1	Approved		X**
Olanzapine	SIGMAR1	Approved		X
Pimozide	SIGMAR1	Approved		X
Rimcazole	SIGMAR1	Pre-clinical		X
Phencyclidine	SIGMAR1	Illicit	X	
Noscapine	SIGMAR1	Approved	X	
Pentazocine	SIGMAR1	Approved	X	
Dimethyltryptamine	SIGMAR1	Experimental, illicit	X	
Remoxipride	SIGMAR1	Approved	X	
Amitriptyline	SIGMAR1	Approved	X	
Prasterone	SIGMAR1	Approved	X	
Captodiame	SIGMAR1	Experimental	X	
Cocaine	SIGMAR1	Approved, illicit	X	
Pentoxyverine	SIGMAR1	Approved	X	
Hydrocodone	SIGMAR1	Approved, illicit, investigational	X	

Appendix C Review Model Equations

PAWELEK

$$\begin{aligned}
\mathrm{H}' &= -\beta\mathrm{HV} - \phi\mathrm{HF} + \rho\mathrm{F} \\
\mathrm{I}' &= \beta\mathrm{HV} - \delta\mathrm{I} - \mathrm{KIF} \\
\mathrm{R}' &= \phi\mathrm{HF} - \rho\mathrm{F} \\
\mathrm{V}' &= \mathrm{pI} - \mathrm{cV} \\
\mathrm{F}' &= \mathrm{qI} - \mathrm{dF}
\end{aligned} \tag{C.1}$$

SAENZ

$$\begin{aligned}
\mathrm{H}' &= -\beta\mathrm{HV} - \phi\mathrm{HF} \\
\mathrm{E}'1 &= \beta\mathrm{HV} - \mathrm{k}_1\mathrm{E}_1, \\
\mathrm{W}' &= \phi\mathrm{FH} - \mathrm{m}\beta\mathrm{VW} - \alpha\mathrm{W} \\
\mathrm{E}'2 &= \mathrm{m}\beta\mathrm{VW} - \mathrm{k}'\mathrm{E}_2\mathrm{R}' = \alpha\mathrm{W} \\
\mathrm{I}' &= \mathrm{k}_1\mathrm{E}_1 + \mathrm{k}'\mathrm{E}_2 - \delta\mathrm{I}, \\
\mathrm{V}' &= \mathrm{pI} - \mathrm{cV}\mathrm{F}' = \mathrm{nqE2} - \mathrm{dF} + \mathrm{qI}
\end{aligned} \tag{C.2}$$

HANCIOGLU

$$\begin{aligned}
H' &= b_{HD}D(H+R) + a_R R - \gamma_{HV}VH - b_{HF}FH \\
I' &= \gamma_{HV}VH - a_I I - b_{IE}EI \\
R' &= b_{HF}FH - a_R R \\
V' &= \gamma_V I - \gamma_{VA}SAV - \gamma_{VH}HV - \alpha_V V - \tfrac{a_{v1}V}{a_{v2}V+1} \\
M' &= (1-M)\left(b_{MD}D + b_{MV}V\right) - a_M M \\
F' &= b_F M + c_F I - b_{FH}HF - a_F F \\
E' &= b_{EM}ME - b_{EI}IE + a_E(1-E) \\
P' &= b_{PM}MP + a_P(1-P) \\
A' &= b_A P - \gamma_{AV}SAV - a_A A \\
S' &= rP(1-S)
\end{aligned} \tag{C.3}$$

Appendix D Innate-Adaptive Model Parameters

Parameter	Meaning
β	Infectivity of the virus
λ	Replication rate of healthy epithelial cells
δ	Natural death rate of infected epithelial cells
K_{CTL}	Clearance rate of infected cells by CTLs
p	Viral replication rate
q	Depletion of free virions by interferon
c	Clearance of virus by antibodies
$b_{IFN,M}$	Activation of macrophages due to IFN
mu_M	Decay rate of macrophages
$b_{I,IFN}$	Increase in IFN due to infected cells
θ_{IFN}	Saturation of IFN increase
$b_{M,IFN}$	Increase in IFN due to macrophages
μ_{IFN}	Decay rate of IFN
$b_{M,CTL}$	Increase in CTLs due to macrophages
μ_{CTL}	Decay rate of activated CTLs
M_0	Initial condition for macrophages
CTL_0	Initial condition for CTLs
n	Hill kinetic parameter; constant

Appendix E Innate Model Parameters

Parameter	Meaning
k	Viral growth rate
K	Maximum Viral Capacity
$r_{IFN,V}$	Viral inhibition by interferon response
d_V	Viral decay rate
$p_{V,IFN}$	Interferon production rate resulting from viral presence
d_{IFN}	Interferon decay rate
k_1	Maximum rate of MCP1 production
k_2	Dissociation constant in MCP1 production
n	Activation threshold of interferon for macrophage production of MCP1
$d_{MCP1,IFN}$	MCP1 decay rate

Bibliography

[1] S. B. Petkova, R. Yuan, S. W. Tsaih, W. Schott, D. C. Roopenian, and B. Paigen, "Genetic influence on immune phenotype revealed strain-specific variations in peripheral blood lineages," *Physiological Genomics*, 2008.

[2] M. T. Ferris, D. L. Aylor, D. Bottomly, A. C. Whitmore, L. D. Aicher, T. A. Bell, B. Bradel-Tretheway, J. T. Bryan, R. J. Buus, L. E. Gralinski, B. L. Haagmans, L. McMillan, D. R. Miller, E. Rosenzweig, W. Valdar, J. Wang, G. A. Churchill, D. W. Threadgill, S. K. McWeeney, M. G. Katze, F. Pardo-Manuel de Villena, R. S. Baric, and M. T. Heise, "Modeling Host Genetic Regulation of Influenza Pathogenesis in the Collaborative Cross," *PLoS Pathogens*, 2013.

[3] E. E. Ackerman, E. Kawakami, M. Katoh, T. Watanabe, S. Watanabe, Y. Tomita, T. J. Lopes, Y. Matsuoka, H. Kitano, J. E. Shoemaker, and Y. Kawaoka, "Network-Guided Discovery of Influenza Virus Replication Host Factors," *mBio*, 2018.

[4] CDC, "Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths in the United States 20182019 influenza season," 2019.

[5] G. Soleimani and M. Akbarpour, "Clinical presentation of novel influenza a (H1N1) in hospitalized children," *Iranian Journal of Pediatrics*, 2011.

[6] T. Uyeki, "Human Infection with Highly Pathogenic Avian Influenza A (H5N1) Virus: Review of Clinical Issues," *Clinical Infectious Diseases*, 2009.

[7] A. Martínez, N. Soldevila, A. Romero-Tamarit, N. Torner, P. Godoy, C. Rius, M. Jané, and . Domínguez, "Risk factors associated with severe outcomes in adult hospitalized patients according to influenza type and subtype," *PLoS ONE*, 2019.

[8] R. Allard, P. Leclerc, C. Tremblay, and T. N. Tannenbaum, "Diabetes and the severity of pandemic influenza A (H1N1) infection," *Diabetes Care*, 2010.

[9] L. Groop and F. Pociot, "Genetics of diabetes - Are we missing the genes or the disease?," *Molecular and Cellular Endocrinology*, 2014.

[10] World Health Organization, "The Control of Neglected Zoonotic Diseases A route to poverty alleviation," tech. rep., World Health Organization, 2006.

[11] A. Banerjee, M. L. Baker, K. Kulcsar, V. Misra, R. Plowright, and K. Mossman, "Novel Insights Into Immune Systems of Bats," 1 2020.

[12] G. A. Van Norman, "Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs," *JACC: Basic to Translational Science*, vol. 1, pp. 170–179, 4 2016.

[13] C. H. Wong, K. W. Siah, and A. W. Lo, "Estimation of clinical trial success rates and related parameters," *Biostatistics*, vol. 20, pp. 273–286, 4 2019.

[14] H. R. Gelderblom, *Structure and Classification of Viruses*. University of Texas Medical Branch at Galveston, 1996.

[15] A. J. Cann, "Replication of Viruses," in *Encyclopedia of Virology*, pp. 406–412, Elsevier Ltd, 1 2008.

[16] B. Kumar, K. Asha, M. Khanna, L. Ronsard, C. A. Meseko, and M. Sanicas, "The emerging influenza virus threat: status and new prospects for its therapy and control," 2018.

[17] "Influenza (Flu) Antiviral Drugs and Related Information — FDA."

[18] I. Kosik and J. W. Yewdell, "Influenza hemagglutinin and neuraminidase: Yinyang proteins coevolving to thwart immunity," 4 2019.

[19] CDC, "Influenza Type A Viruses — Avian Influenza (Flu)."

[20] CDC, "Highly Pathogenic Asian Avian Influenza A(H5N1) Virus — Avian Influenza (Flu)."

[21] J. Guarner and R. Falcón-Escobedo, "Comparison of the pathology caused by H1N1, H5N1, and H3N2 influenza viruses," 11 2009.

[22] CDC, "Highly Pathogenic Asian Avian Influenza A(H5N1) in People — Avian Influenza (Flu)."

[23] J. Treanor, "Influenza Vaccine Outmaneuvering Antigenic Shift and Drift," *New England Journal of Medicine*, vol. 350, no. 3, pp. 218–220, 2004.

[24] A. T. Harding and N. S. Heaton, "Efforts to improve the seasonal influenza vaccine," 6 2018.

[25] L. Buonaguro, M. Tagliamonte, M. L. Tornesello, and F. M. Buonaguro, "SARS-CoV-2 RNA polymerase as target for antiviral therapy," 2020.

[26] P. A. Rota, M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Peñaranda, B. Bankamp, K. Maher, M. h. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J. L. DeRisi, Q. Chen, D. Wang, D. D. Erdman, T. C. Peret, C. Burns, T. G. Ksiazek, P. E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Günther, A. D. Osterhaus, C. Drosten, M. A. Pallansch, L. J. Anderson, and W. J. Bellini, "Characterization of a novel coronavirus associated with severe acute respiratory syndrome," *Science*, vol. 300, pp. 1394–1399, 5 2003.

[27] L. Lu, W. Zhong, Z. Bian, Z. Li, K. Zhang, B. Liang, Y. Zhong, M. Hu, L. Lin, J. Liu, X. Lin, Y. Huang, J. Jiang, X. Yang, X. Zhang, and Z. Huang, "A comparison of mortality-related risk factors of COVID-19, SARS, and MERS: A systematic review and meta-analysis," 10 2020.

[28] P. Dabisch, M. Schuit, A. Herzog, K. Beck, S. Wood, M. Krause, D. Miller, W. Weaver, D. Freeburger, I. Hooper, B. Green, G. Williams, B. Holland, J. Bohannon, V. Wahl, J. Yolitz, M. Hevey, and S. Ratnesar-Shumate, "The influence of temperature, humidity, and simulated sunlight on the infectivity of SARS-CoV-2 in aerosols," *Aerosol Science and Technology*, vol. 55, no. 2, pp. 142–153, 2021.

[29] A. A. Rabaan, S. H. Al-Ahmed, M. K. Al-Malkey, R. A. Alsubki, S. Ezzikouri, F. Hassan Al-Hababi, R. Sah, A. Al Mutair, S. Alhumaid, J. A. Al, F. Alshahrani, D. Bahadur Shrestha, M. Isaqali Karobari, and S. Arabia, "Airborne transmission of SARS-CoV-2 is the dominant route of transmission: droplets and aerosols King Fahad Medical City," Tech. Rep. 1.

[30] Q. X. Long, X. J. Tang, Q. L. Shi, Q. Li, H. J. Deng, J. Yuan, J. L. Hu, W. Xu, Y. Zhang, F. J. Lv, K. Su, F. Zhang, J. Gong, B. Wu, X. M. Liu, J. J. Li, J. F. Qiu, J. Chen, and A. L. Huang, "Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections," *Nature Medicine*, vol. 26, pp. 1200–1204, 8 2020.

[31] F. Zhang, R. Gan, Z. Zhen, X. Hu, X. Li, F. Zhou, Y. Liu, C. Chen, S. Xie, B. Zhang, X. Wu, and Z. Huang, "Adaptive immune responses to SARS-CoV-2 infection in severe versus mild individuals," *Signal Transduction and Targeted Therapy*, vol. 5, pp. 1–11, 12 2020.

[32] T. Struyf, J. J. Deeks, J. Dinnes, Y. Takwoingi, C. Davenport, M. M. Leeflang, R. Spijker, L. Hooft, D. Emperador, S. Dittrich, J. Domen, S. R. Horn, and A. Van den Bruel, "Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease," 7 2020.

[33] J. Liu, S. Li, J. Liu, B. Liang, X. Wang, H. Wang, W. Li, Q. Tong, J. Yi, L. Zhao, L. Xiong, C. Guo, J. Tian, J. Luo, J. Yao, R. Pang, H. Shen, C. Peng, T. Liu, Q. Zhang, J. Wu, L. Xu, S. Lu, B. Wang, Z. Weng, C. Han, H. Zhu, R. Zhou, H. Zhou, X. Chen, P. Ye, B. Zhu, L. Wang, W. Zhou, S. He, Y. He, S. Jie, P. Wei, J. Zhang, Y. Lu, W. Wang, L. Zhang, L. Li, F. Zhou, J. Wang, U. Dittmer, M. Lu, Y. Hu, D. Yang, and X. Zheng, "Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients," *EBioMedicine*, vol. 55, p. 102763, 5 2020.

[34] C. H. Sudre, B. Murray, T. Varsavsky, M. S. Graham, R. S. Penfold, R. C. Bowyer, J. C. Pujol, K. Klaser, M. Antonelli, L. S. Canas, E. Molteni, M. Modat, M. Jorge Cardoso, A. May, S. Ganesh, R. Davies, L. H. Nguyen, D. A. Drew, C. M. Astley, A. D. Joshi, J. Merino, N. Tsereteli, T. Fall, M. F. Gomez, E. L. Duncan, C. Menni, F. M. Williams, P. W. Franks, A. T. Chan, J. Wolf, S. Ourselin, T. Spector, and C. J. Steves, "Attributes and predictors of long COVID," *Nature Medicine*, vol. 27, pp. 626–631, 4 2021.

[35] World Health Organization, "Coronavirus disease (COVID-19) Situation Report 118," 2020.

[36] M. A. Bakowski, N. Beutler, K. C. Wolff, M. G. Kirkpatrick, E. Chen, T.-T. H. Nguyen, L. Riva, N. Shaabani, M. Parren, J. Ricketts, A. K. Gupta, K. Pan, P. Kuo, M. Fuller, E. Garcia, J. R. Teijaro, L. Yang, D. Sahoo, V. Chi, E. Huang, N. Vargas, A. J. Roberts, S. Das, P. Ghosh, A. K. Woods, S. B. Joseph, M. V. Hull, P. G. Schultz, D. R. Burton, A. K. Chatterjee, C. W. McNamara, and T. F. Rogers, "Drug repurposing screens identify chemical entities for the development of COVID-19 interventions," *Nature Communications*, vol. 12, p. 3309, 12 2021.

[37] T. U. Singh, S. Parida, M. C. Lingaraju, M. Kesavan, D. Kumar, and R. K. Singh, "Drug repurposing approach to fight COVID-19," 12 2020.

[38] G. Ciliberto, R. Mancini, and M. G. Paggi, "Drug repurposing against COVID-19: Focus on anticancer agents," 5 2020.

[39] FDA, "Know Your Treatment Options for COVID-19 — FDA."

[40] Institute for Health Metrics and Evaluation, "COVID-19 vaccine efficacy summary — Institute for Health Metrics and Evaluation."

[41] S. S. Abdool Karim and T. de Oliveira, "New SARS-CoV-2 Variants Clinical, Public Health, and Vaccine Implications," *New England Journal of Medicine*, vol. 384, pp. 1866–1868, 5 2021.

[42] L. Guillot, R. Le Goffic, S. Bloch, N. Escriou, S. Akira, M. Chignard, and M. Si-Tahar, "Involvement of Toll-like receptor 3 in the immune response of lung epithelial cells to double-stranded RNA and influenza A virus," *Journal of Biological Chemistry*, 2005.

[43] P. J. Delves and I. M. Roitt, "Advances in immunology: The immune system (Second of two parts)," 2000.

[44] U. A. Maus, M. Audrey Koay, T. Delbeck, M. Mack, M. Ermert, L. Ermert, T. S. Blackwell, J. W. Christman, D. Schlöndorff, W. Seeger, and J. Lohmeyer, "Role of resident alveolar macrophages in leukocyte traffic into the alveolar air space of intact mice," *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 2002.

[45] F. Tian, Y. Han, J. Song, J. Lei, X. Yan, N. Xie, J. Wang, J. Zhao, X. Liang, D. Zhong, Y. Zhou, X. Wang, and X. Li, "Pulmonary resident neutrophils regulate the production of GM-CSF and alveolar macrophages," *FEBS Journal*, 2016.

[46] L. Chen, H. Deng, H. Cui, J. Fang, Z. Zuo, J. Deng, Y. Li, X. Wang, and L. Zhao, "Inflammatory responses and inflammation-associated diseases in organs," 2018.

[47] D. Hirayama, T. Iida, and H. Nakase, "The phagocytic function of macrophage-enforcing innate immunity and tissue homeostasis," 2018.

[48] F. J. Culley, "Natural killer cells in infection and inflammation of the lung," 2009.

[49] T. V. Condon, R. T. Sawyer, M. J. Fenton, and D. W. H. Riches, "Lung dendritic cells at the innate-adaptive immune interface," *Journal of Leukocyte Biology*, 2011.

[50] M. P. Matheu, J. R. Teijaro, K. B. Walsh, M. L. Greenberg, D. Marsolais, I. Parker, H. Rosen, M. B. Oldstone, and M. D. Cahalan, "Three Phases of CD8 T Cell Response in the Lung Following H1N1 Influenza Infection and Sphingosine 1 Phosphate Agonist Therapy," *PLoS ONE*, 2013.

[51] S. M. Kaech, E. J. Wherry, and R. Ahmed, "Effector and memory T-cell differentiation: Implications for vaccine development," 2002.

[52] D. C. Parker, "T Cell-Dependent B Cell Activation," *Annual Review of Immunology*, 1993.

[53] J. R. Tisoncik, M. J. Korth, C. P. Simmons, J. Farrar, T. R. Martin, and M. G. Katze, "Into the Eye of the Cytokine Storm," *Microbiology and Molecular Biology Reviews*, 2012.

[54] J. E. Shoemaker, S. Fukuyama, A. J. Eisfeld, D. Zhao, E. Kawakami, S. Sakabe, T. Maemura, T. Gorai, H. Katsura, Y. Muramoto, S. Watanabe, T. Watanabe, K. Fuji, Y. Matsuoka, H. Kitano, and Y. Kawaoka, "An Ultrasensitive Mechanism Regulates Influenza Virus-Induced Inflammation," *PLoS Pathogens*, 2015.

[55] A. Yang, M. Troup, and J. W. Ho, "Scalability and Validation of Big Data Bioinformatics Software," 1 2017.

[56] C. S. Greene, J. Tan, M. Ung, J. H. Moore, and C. Cheng, "Big data bioinformatics," 12 2014.

[57] R. M. Ward, R. Schmieder, G. Highnam, and D. Mittelman, "Big data challenges and opportunities in high-throughput sequencing," *Systems Biomedicine*, vol. 1, pp. 29–34, 1 2013.

[58] E. C. Nice, "From proteomics to personalized medicine: The road ahead," 4 2016.

[59] P. Angerer, L. Simon, S. Tritschler, F. A. Wolf, D. Fischer, and F. J. Theis, "Single cells make big data: New challenges and opportunities in transcriptomics," 8 2017.

[60] D. Prawitt, L. Brixel, C. Spangenberg, L. Eshkind, R. Heck, F. Oesch, B. Zabel, and E. Bockamp, "RNAi knock-down mice: An emerging technology for post-genomic functional genetics," 2004.

[61] A. O'Driscoll, J. Daugelaite, and R. D. Sleator, "'Big data', Hadoop and cloud computing in genomics," 2013.

[62] J. Fan, F. Han, and H. Liu, "Challenges of Big Data analysis," 2014.

[63] K. A. Pawelek, G. T. Huynh, M. Quinlivan, A. Cullinane, L. Rong, and A. S. Perelson, "Modeling within-host dynamics of influenza virus infection including immune responses," *PLoS Computational Biology*, vol. 8, no. 6, 2012.

[64] H. Y. Lee, D. J. Topham, S. Y. Park, J. Hollenbaugh, J. Treanor, T. R. Mosmann, X. Jin, B. M. Ward, H. Miao, J. Holden-Wiltse, and others, "Simulation and prediction of the adaptive immune response to influenza A virus infection," *Journal of virology*, vol. 83, no. 14, pp. 7151–7165, 2009.

[65] H. Miao, J. A. Hollenbaugh, M. S. Zand, J. Holden-Wiltse, T. R. Mosmann, A. S. Perelson, H. Wu, and D. J. Topham, "Quantifying the Early Immune Response and Adaptive Immune Response Kinetics in Mice Infected with Influenza A Virus," *Journal of Virology*, vol. 84, no. 13, pp. 6687–6698, 2010.

[66] B. Hancioglu, D. Swigon, and G. Clermont, "A dynamical model of human immune response to influenza A virus infection," *Journal of Theoretical Biology*, vol. 246, no. 1, pp. 70–86, 2007.

[67] I. Price, E. D. Mochan-Keef, D. Swigon, G. B. Ermentrout, S. Lukens, F. R. Toapanta, T. M. Ross, and G. Clermont, "The inflammatory response to influenza A virus (H1N1): An experimental and mathematical study," *Journal of Theoretical Biology*, 2015.

[68] J. E. Shoemaker, S. Fukuyama, A. J. Eisfeld, Y. Muramoto, S. Watanabe, T. Watanabe, Y. Matsuoka, H. Kitano, and Y. Kawaoka, "Integrated network analysis reveals a novel role for the cell cycle in 2009 pandemic influenza virus-induced inflammation in macaque lungs.," *BMC systems biology*, vol. 6, p. 117, 1 2012.

[69] S. Tripathi, M. O. Pohl, Y. Zhou, A. Rodriguez-Frandsen, G. Wang, D. A. Stein, H. M. Moulton, P. Dejesus, J. Che, L. C. Mulder, E. Yángüez, D. Andenmatten, L. Pache, B. Manicassamy, R. A. Albrecht, M. G. Gonzalez, Q. Nguyen, A. Brass, S. Elledge, M. White, S. Shapira, N. Hacohen, A. Karlas, T. F. Meyer, M. Shales, A. Gatorano, J. R. Johnson, G. Jang, T. Johnson, E. Verschueren, D. Sanders, N. Krogan, M. Shaw, R. König, S. Stertz, A. García-Sastre, and S. K. Chanda, "Meta- and Orthogonal Integration of Influenza "oMICs" Data Defines a Role for UBR4 in Virus Budding," *Cell Host and Microbe*, vol. 18, no. 6, pp. 723–735, 2015.

[70] D. K. Arrell and A. Terzic, "Network systems biology for drug discovery," 2010.

[71] A. Pujol, R. Mosca, J. Farrés, and P. Aloy, "Unveiling the role of network and systems biology in drug discovery," 2010.

[72] E. Klipp and W. Liebermeister, "Mathematical modeling of intracellular signaling pathways," 2006.

[73] S. P. Gygi, Y. Rochon, B. R. Franza, and R. Aebersold, "Correlation between Protein and mRNA Abundance in Yeast," *Molecular and Cellular Biology*, vol. 19, pp. 1720–1730, 3 1999.

[74] D. Greenbaum, C. Colangelo, K. Williams, and M. Gerstein, "Comparing protein abundance and mRNA expression levels on a genomic scale," 8 2003.

[75] D. E. Kirschner, S. T. Chang, T. W. Riggs, N. Perry, and J. J. Linderman, "Toward a multiscale model of antigen presentation in immunity," *Immunological Reviews*, vol. 216, pp. 93–118, 2007.

[76] A. Bouchnita, G. Bocharov, A. Meyerhans, and V. Volpert, "Towards a multiscale model of acute HIV infection," *Computation*, vol. 5, p. 6, 3 2017.

[77] R. W. Gregg, F. Shabnam, and J. E. Shoemaker, "Agent-based modeling reveals benefits of heterogeneous and stochastic cell populations during cGAS-mediated IFNβ production," *Bioinformatics*, vol. 37, pp. 1428–1434, 6 2021.

[78] T. J. Sego, J. O. Aponte-Serrano, J. F. Gianlupi, S. R. Heaps, K. Breithaupt, L. Brusch, J. Crawshaw, J. M. Osborne, E. M. Quardokus, R. K. Plemper, and J. A. Glazier, "A modular framework for multiscale, multicellular, spatiotemporal modeling of acute primary viral infection and immune response in epithelial tissues and its application to drug therapy timing and effectiveness," *PLoS Computational Biology*, vol. 16, p. e1008451, 12 2020.

[79] S. A. Prokopiou, L. Barbarroux, S. Bernard, J. Mafille, Y. Leverrier, C. Arpin, J. Marvel, O. Gandrillon, and F. Crauste, "Multiscale modeling of the early CD8 T-cell immune response in lymph nodes: An integrative study," *Computation*, vol. 2, pp. 159–181, 9 2014.

[80] F. S. Heldt, T. Frensing, A. Pflugmacher, R. Gröpler, B. Peschel, and U. Reichl, "Multiscale Modeling of Influenza A Virus Infection Supports the Development of Direct-Acting Antivirals," *PLoS Computational Biology*, vol. 9, p. 1003372, 11 2013.

[81] C. Chicone, *Ordinary differential equations with applications*. 2006.

[82] Z. Ji, K. Yan, W. Li, H. Hu, and X. Zhu, "Mathematical and Computational Modeling in Complex Biological Systems," 2017.

[83] P. Städter, Y. Schälte, L. Schmiester, J. Hasenauer, and P. L. Stapor, "Benchmarking of numerical integration methods for ODE models of biological systems," *Scientific Reports*, vol. 11, p. 2696, 12 2021.

[84] D. E. Gratie, B. Iancu, and I. Petre, "ODE analysis of biological systems," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7938 LNCS, pp. 29–62, Springer Verlag, 2013.

[85] J. S. Hodges, *Richly Parameterized Linear Models*. Chapman and Hall/CRC, 4 2016.

[86] S. Chib and . E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," Tech. Rep. 4, 1995.

[87] S. Gupta, L. Hainsworth, J. Hogg, R. Lee, and J. Faeder, "Evaluation of Parallel Tempering to Accelerate Bayesian Parameter Estimation in Systems Biology," in *Proceedings - 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018*, vol. 2018, pp. 690–697, Institute of Electrical and Electronics Engineers Inc., 6 2018.

[88] S. Sinharay, "Assessing Convergence of the Markov Chain Monte Carlo Algorithms: A Review," tech. rep., 2003.

[89] W. D. Vousden, . W. M. Farr, and I. Mandel, "Dynamic temperature selection for parallel tempering in Markov chain Monte Carlo simulations," *MNRAS*, vol. 455, pp. 1919–1937, 2016.

[90] G. O. Roberts, A. Gelman, and W. R. Gilks, "WEAK CONVERGENCE AND OPTIMAL SCALING OF RANDOM WALK METROPOLIS ALGORITHMS 1," Tech. Rep. 1, 1997.

[91] M. Koutrouli, E. Karatzas, D. Paez-Espino, and G. A. Pavlopoulos, "A Guide to Conquer the Biological Network Era Using Graph Theory," 1 2020.

[92] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[93] S. P. Borgatti, "Centrality and network flow," *Social Networks*, 2005.

[94] M. G. Everett and S. P. Borgatti, "The centrality of groups and classes," *Journal of Mathematical Sociology*, 1999.

[95] A. del Sol, H. Fujihashi, and P. O'Meara, "Topology of small-world networks of protein-protein complex structures," *Bioinformatics*, vol. 21, pp. 1311–1315, 4 2005.

[96] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Muüller, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors," *Nature Biotechnology*, vol. 20, no. 4, pp. 370–375, 2002.

[97] B. B. Aldridge, J. M. Burke, D. A. Lauffenburger, and P. K. Sorger, "Physicochemical modelling of cell signalling pathways," 2006.

[98] M. Zhu, L. Gao, X. Li, Z. Liu, C. Xu, Y. Yan, E. Walker, W. Jiang, B. Su, X. Chen, and H. Lin, "The analysis of the drug-targets based on the topological properties in the human protein-protein interaction network," *Journal of Drug Targeting*, vol. 17, no. 7, pp. 524–532, 2009.

[99] A. Vinayagam, J. Zirin, C. Roesel, Y. Hu, B. Yilmazel, A. A. Samsonova, R. A. Neum??ller, S. E. Mohr, and N. Perrimon, "Integrating protein-protein interaction networks with phenotypes reveals signs of interactions," *Nature Methods*, vol. 11, no. 1, pp. 94–99, 2014.

[100] X. He and J. Zhang, "Why do hubs tend to be essential in protein networks?," *PLoS Genetics*, 2006.

[101] T. J. Lopes, J. E. Shoemaker, Y. Matsuoka, Y. Kawaoka, and H. Kitano, "Identifying problematic drugs based on the characteristics of their targets," *Frontiers in Pharmacology*, 2015.

[102] A.-L. Barabasi, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease.," *Nature reviews. Genetics*, vol. 12, no. 1, pp. 56–68, 2011.

[103] A. J. van Hoek, A. Underwood, M. Jit, E. Miller, and W. J. Edmunds, "The impact of pandemic influenza H1N1 on health-related quality of life: A prospective population-based study," *PLoS ONE*, vol. 6, no. 3, pp. 1–6, 2011.

[104] M. L. Perdue and R. A. Bright, "United States of America Department of Health and Human Services support for advancing influenza vaccine manufacturing in the developing world," *Vaccine*, vol. 29, no. SUPPL. 1, pp. A48–A50, 2011.

[105] M. M. Davis, P. M. Wortley, S. M. Ndiaye, M. G. Woods, and S. J. Clark, "National availability of influenza vaccine among medical subspecialty practices," *American Journal of Preventive Medicine*, vol. 26, pp. 307–310, 4 2004.

[106] G. Neumann, H. Chen, G. F. Gao, Y. Shu, and Y. Kawaoka, "H5N1 influenza viruses: outbreaks and biological properties.," *Cell research*, vol. 20, pp. 51–61, 1 2010.

[107] R. Gao, B. Cao, Y. Hu, Z. Feng, D. Wang, W. Hu, J. Chen, Z. Jie, H. Qiu, K. Xu, X. Xu, H. Lu, W. Zhu, Z. Gao, N. Xiang, Y. Shen, Z. He, Y. Gu, Z. Zhang, Y. Yang, X. Zhao, L. Zhou, X. X. Li, S. Zou, Y. Y. Zhang, L. Yang, J. Guo, J. Dong, Q. Li, L. Dong, Y. Zhu, T. Bai, S. Wang, P. Hao, W. Yang, J. Han, H. Yu, D. Li, G. F. Gao, G. Wu, Y. Wang, Z. Yuan, and Y. Shu, "Human infection with a novel avian-origin influenza A (H7N9) virus.," *The New England journal of medicine*, vol. 368, no. 20, pp. 1888–97, 2013.

[108] Q. Li, L. Zhou, M. Zhou, Z. Chen, F. Li, H. Wu, N. Xiang, E. Chen, F. Tang, D. Wang, L. Meng, Z. Hong, W. Tu, Y. Cao, L. Li, F. Ding, B. Liu, M. Wang, R. Xie, R. Gao, X. Li, T. Bai, S. Zou, J. He, J. Hu, Y. Xu, C. Chai, S. Wang, Y. Gao, L. Jin, Y. Zhang, H. Luo, H. Yu, J. He, Q. Li, X. Wang, L. Gao, X. Pang, G. Liu, Y. Yan, H. Yuan, Y. Shu, W. Yang, Y. Wang, F. Wu, T. M. Uyeki, and Z. Feng, "Epidemiology of human infections with avian influenza A(H7N9) virus in China.," *The New England journal of medicine*, vol. 370, no. 6, pp. 520–32, 2014.

[109] H. L. Yen and R. G. Webster, "Pandemic influenza as a current threat," 2009.

[110] W. L. Davies, R. R. Grunert, R. F. Haff, J. W. McGahen, E. M. Neumayer, M. Paulshock, J. C. Watts, T. R. Wood, E. C. Hermann, and C. E. Hoffmann, "Antiviral Activity of 1-Adamantanamine (Amantadine)," *Science*, vol. 144, pp. 862–863, 5 1964.

[111] F. G. Hayden, "Perspectives on antiviral use during pandemic influenza.," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 356, no. 1416, pp. 1877–1884, 2001.

[112] D. Tamura, K. Mitamura, M. Yamazaki, M. Fujino, M. Nirasawa, K. Kimura, M. Kiso, H. Shimizu, C. Kawakami, S. Hiroi, K. Takahashi, M. Hata, H. Minagawa, Y. Kimura,

S. Kaneda, S. Sugita, T. Horimoto, N. Sugaya, and Y. Kawaoka, "Oseltamivir-resistant influenza a viruses circulating in Japan.," *Journal of clinical microbiology*, vol. 47, pp. 1424–7, 5 2009.

[113] R. a. Bright, D. K. Shay, B. Shu, N. J. Cox, and A. I. Klimov, "Adamantane resistance among influenza A viruses isolated early during the 2005-2006 influenza season in the United States.," *JAMA : the journal of the American Medical Association*, vol. 295, no. 8, pp. 891–894, 2006.

[114] M. Kotlyar, K. Fortney, and I. Jurisica, "Network-based characterization of drug-regulated genes, drug targets, and toxicity.," *Methods (San Diego, Calif.)*, vol. 57, pp. 499–507, 8 2012.

[115] M. J. Korth, N. Tchitchek, A. G. Benecke, and M. G. Katze, "Systems approaches to influenza-virus host interactions and the pathogenesis of highly virulent and pandemic viruses," 2013.

[116] Y. Y. Liu, J. J. Slotine, and A. L. Barabási, "Controllability of complex networks," *Nature*, vol. 473, pp. 167–173, 5 2011.

[117] R. S. Burt, "Structural Holes and Good Ideas," *The American Journal of Sociology;*, vol. 110, no. 2, pp. 349–399, 2004.

[118] P. F. Jonsson and P. A. Bates, "Global topological features of cancer proteins in the human interactome," *Bioinformatics*, vol. 22, no. 18, pp. 2291–2297, 2006.

[119] T. Hase, H. Tanaka, Y. Suzuki, S. Nakagawa, and H. Kitano, "Structure of protein interaction networks and their implications on drug design," *PLoS Computational Biology*, vol. 5, no. 10, 2009.

[120] M. a. Yildirim, K.-I. Goh, M. E. Cusick, A.-L. Barabási, and M. Vidal, "Drug-target network.," *Nature biotechnology*, vol. 25, pp. 1119–26, 10 2007.

[121] T. J. S. Lopes, J. E. Shoemaker, Y. Matsuoka, Y. Kawaoka, and H. Kitano, "Identifying problematic drugs based on the characteristics of their targets," *Frontiers in Pharmacology*, vol. 6, 2015.

[122] S. Jin, Y. Li, R. Pan, and X. Zou, "Characterizing and controlling the inflammatory network during influenza A virus infection.," *Scientific reports*, vol. 4, p. 3799, 2014.

[123] M. H. Schaefer, T. J. S. Lopes, N. Mah, J. E. Shoemaker, Y. Matsuoka, J. F. Fontaine, C. Louis-Jeune, A. J. Eisfeld, G. Neumann, C. Perez-Iratxeta, Y. Kawaoka, H. Kitano, and M. A. Andrade-Navarro, "Adding Protein Context to the Human Protein-Protein Interaction Network to Reveal Meaningful Interactions," *PLoS Computational Biology*, 2013.

[124] J. E. Shoemaker, S. Fukuyama, A. J. Eisfeld, Y. Muramoto, S. Watanabe, T. Watanabe, Y. Matsuoka, H. Kitano, and Y. Kawaoka, "Integrated network analysis reveals a novel role for the cell cycle in 2009 pandemic influenza virus-induced inflammation in macaque lungs," *BMC Systems Biology*, vol. 6, 2012.

[125] B. Taye, C. Vaz, V. Tanavde, V. A. Kuznetsov, F. Eisenhaber, R. J. Sugrue, and S. Maurer-Stroh, "Benchmarking selected computational gene network growing tools in context of virus-host interactions," *Scientific Reports*, vol. 7, no. 1, p. 5805, 2017.

[126] N. S. Heaton, N. Moshkina, R. Fenouil, T. J. Gardner, S. Aguirre, P. S. Shah, N. Zhao, L. Manganaro, J. F. Hultquist, J. Noel, D. H. Sachs, J. Hamilton, P. E. Leon, A. Chawdury, S. Tripathi, C. Melegari, L. Campisi, R. Hai, G. Metreveli, A. V. Gamarnik, A. García-Sastre, B. Greenbaum, V. Simon, A. Fernandez-Sesma, N. J. Krogan, L. C. Mulder, H. van Bakel, D. Tortorella, J. Taunton, P. Palese, and I. Marazzi, "Targeting Viral Proteostasis Limits Influenza Virus, HIV, and Dengue Virus Infection," *Immunity*, vol. 44, no. 1, pp. 46–58, 2016.

[127] L. Meyniel-Schicklin, B. de Chassey, P. André, and V. Lotteau, "Viruses and Interactomes in Translation," *Molecular & Cellular Proteomics*, vol. 11, no. 7, p. M111.014738, 2012.

[128] M.-A. Germain, L. Chatel-Chaix, B. Gagné, . Bonneil, P. Thibault, F. Pradezynski, B. de Chassey, L. Meyniel-Schicklin, V. Lotteau, M. Baril, and D. Lamarre, "Elucidating Novel Hepatitis C VirusHost Interactions Using Combined Mass Spectrometry and Functional Genomics Approaches," *Molecular & Cellular Proteomics*, 2014.

[129] L. Wang, B. Fu, W. Li, G. Patil, L. Liu, M. E. Dorf, and S. Li, "Comparative influenza protein interactomes identify the role of plakophilin 2 in virus restriction," *Nature Communications*, vol. 8, 2017.

[130] M. H. Schaefer, J. F. Fontaine, A. Vinayagam, P. Porras, E. E. Wanker, and M. A. Andrade-Navarro, "Hippie: Integrating protein interaction networks with experiment based quality scores," *PLoS ONE*, vol. 7, no. 2, pp. 1–8, 2012.

[131] T. Watanabe, E. Kawakami, J. E. Shoemaker, T. J. S. Lopes, Y. Matsuoka, Y. Tomita, H. Kozuka-Hata, T. Gorai, T. Kuwahara, E. Takeda, A. Nagata, R. Takano, M. Kiso, M. Yamashita, Y. Sakai-Tagawa, H. Katsura, N. Nonaka, H. Fujii, K. Fujii, Y. Sugita, T. Noda, H. Goto, S. Fukuyama, S. Watanabe, G. Neumann, M. Oyama, H. Kitano, and Y. Kawaoka, "Influenza virus-host interactome screen as a platform for antiviral drug development.," *Cell host & microbe*, vol. 16, no. 6, pp. 795–805, 2014.

[132] Gabor Csardi and Tamas Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Sy, 2006.

[133] T. Geiger, A. Wehner, C. Schaab, J. Cox, and M. Mann, "Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins," *Molecular & Cellular Proteomics*, vol. 11, no. 3, p. M111.014050, 2012.

[134] A. Ruepp, B. Brauner, I. Dunger-Kaltenbach, G. Frishman, C. Montrone, M. Stransky, B. Waegele, T. Schmidt, O. N. Doudieu, V. Stümpflen, and H. W. Mewes, "CORUM: the comprehensive resource of mammalian protein complexes.," *Nucleic acids research*, vol. 36, pp. 646–50, 1 2008.

[135] L. Hao, Q. He, Z. Wang, M. Craven, M. a. Newton, and P. Ahlquist, "Limited agreement of independent RNAi screens for virus-required host genes owes more to false-negative than false-positive factors.," *PLoS computational biology*, vol. 9, p. e1003235, 1 2013.

[136] A. Karlas, N. Machuy, Y. Shin, K.-P. Pleissner, A. Artarini, D. Heuer, D. Becker, H. Khalil, L. A. Ogilvie, S. Hess, A. P. Mäurer, E. Müller, T. Wolff, T. Rudel, and T. F. Meyer, "Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication.," *Nature*, vol. 463, pp. 818–22, 2 2010.

[137] D. W. Huang, B. T. Sherman, Q. Tan, J. Kir, D. Liu, D. Bryant, Y. Guo, R. Stephens, M. W. Baseler, H. C. Lane, and R. A. Lempicki, "DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists," *Nucleic Acids Research*, vol. 35, 7 2007.

[138] X. He and J. Zhang, "Why do hubs tend to be essential in protein networks?," *PLoS Genetics*, vol. 2, no. 6, pp. 0826–0834, 2006.

[139] J. Zhao, T.-H. Yang, Y. Huang, and P. Holme, "Ranking candidate disease genes from gene expression and protein interaction: a Katz-centrality based approach.," *PloS one*, vol. 6, no. 9, p. e24306, 2011.

[140] X. Wang, B. Thijssen, and H. Yu, "Target essentiality and centrality characterize drug side effects.," *PLoS computational biology*, vol. 9, p. e1003119, 1 2013.

[141] T. Hase, H. Tanaka, Y. Suzuki, S. Nakagawa, and H. Kitano, "Structure of protein interaction networks and their implications on drug design," *PLoS Computational Biology*, vol. 5, no. 10, 2009.

[142] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, 1977.

[143] J. Ivanic, X. Yu, A. Wallqvist, and J. Reifman, "Influence of protein abundance on high-throughput protein-protein interaction detection," *PLoS ONE*, vol. 4, no. 6, 2009.

[144] B. De Chassey, V. Navratil, L. Tafforeau, M. S. Hiet, A. Aublin-Gex, S. Agaugué, G. Meiffren, F. Pradezynski, B. F. Faria, T. Chantier, M. Le Breton, J. Pellet, N. Davoust, P. E. Mangeot, A. Chaboud, F. Penin, Y. Jacob, P. O. Vidalain, M. Vidal, P. André, C. Rabourdin-Combe, and V. Lotteau, "Hepatitis C virus infection protein network," *Molecular Systems Biology*, 2008.

[145] S. Khadka, A. D. Vangeloff, C. Zhang, P. Siddavatam, N. S. Heaton, L. Wang, R. Sengupta, S. Sahasrabudhe, G. Randall, M. Gribskov, R. J. Kuhn, R. Perera, and D. J. LaCount, "A Physical Interaction Network of Dengue Virus and Human Proteins," *Molecular & Cellular Proteomics*, vol. 10, no. 12, p. M111.012187, 2011.

[146] L. Giot, J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L. Hao, C. E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. A. Stanyon, R. L. Finley, K. P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R. A. Shimkets, M. P. McKenna, J. Chant, and J. M. Rothberg, "A protein interaction map of Drosophila melanogaster.," *Science*, vol. 302, no. 5651, pp. 1727–1736, 2003.

[147] T. Andrews, F. Honti, R. Pfundt, N. d. Leeuw, J. Hehir-Kwa, A. V.-v. Silfhout, B. d. Vries, and C. Webber, "The clustering of functionally related genes contributes to CNV-mediated disease," *Genome Research*, vol. 25, no. 6, pp. 802–813, 2015.

[148] L. Hao, A. Sakurai, T. Watanabe, E. Sorensen, C. A. Nidom, M. A. Newton, P. Ahlquist, and Y. Kawaoka, "Drosophila RNAi screen identifies host genes important for influenza virus replication.," *Nature*, vol. 454, pp. 890–893, 8 2008.

[149] S. D. Shapira, I. Gat-Viks, B. O. Shum, A. Dricot, M. M. de Grace, L. Wu, P. B. Gupta, T. Hao, S. J. Silver, D. E. Root, D. E. Hill, A. Regev, and N. Hacohen, "A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection.," *Cell*, vol. 139, pp. 1255–1267, 12 2009.

[150] A. L. Brass, I. C. Huang, Y. Benita, S. P. John, M. N. Krishnan, E. M. Feeley, B. J. Ryan, J. L. Weyer, L. van der Weyden, E. Fikrig, D. J. Adams, R. J. Xavier, M. Farzan, and S. J. Elledge, "The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus.," *Cell*, vol. 139, pp. 1243–1254, 12 2009.

[151] B. Sui, D. Bamba, K. Weng, H. Ung, S. Chang, J. Van Dyke, M. Goldblatt, R. Duan, M. S. Kinch, and W. B. Li, "The use of Random Homozygous Gene Perturbation to identify novel host-oriented targets for influenza," *Virology*, vol. 387, no. 2, pp. 473–481, 2009.

[152] R. König, S. Stertz, Y. Zhou, A. Inoue, H. H. Hoffmann, S. Bhattacharyya, J. G. Alamares, D. M. Tscherne, M. B. Ortigoza, Y. Liang, Q. Gao, S. E. Andrews, S. Bandyopadhyay, P. De Jesus, B. P. Tu, L. Pache, C. Shih, A. Orth, G. Bonamy, L. Miraglia, T. Ideker, A. García-Sastre, J. A. Young, P. Palese, M. L. Shaw, and S. K. Chanda, "Human host factors required for influenza virus replication," *Nature*, 2010.

[153] J. Xia, J. Sun, P. Jia, and Z. Zhao, "Do cancer proteins really interact strongly in the human protein-protein interaction network?," *Computational Biology and Chemistry*, vol. 35, no. 3, pp. 121–125, 2011.

[154] J. F. Poyatos and L. D. Hurst, "Is optimal gene order impossible?," *Trends in Genetics*, vol. 22, no. 8, pp. 420–423, 2006.

[155] B. G. Hale, R. E. Randall, J. Ortín, and D. Jackson, "The multifunctional NS1 protein of influenza A viruses.," *The Journal of general virology*, vol. 89, pp. 2359–76, 10 2008.

[156] E.-K. Pauli, M. Schmolke, T. Wolff, D. Viemann, J. Roth, J. G. Bode, and S. Ludwig, "Influenza A Virus Inhibits Type I IFN Signaling via NF-κB-Dependent Induction of SOCS-3 Expression," *PLoS Pathogens*, vol. 4, p. e1000196, 11 2008.

[157] C. J. Echeverri, P. A. Beachy, B. Baum, M. Boutros, F. Buchholz, S. K. Chanda, J. Downward, J. Ellenberg, A. G. Fraser, N. Hacohen, W. C. Hahn, A. L. Jackson, A. Kiger, P. S. Linsley, L. Lum, Y. Ma, B. Mathey-Prévôt, D. E. Root, D. M. Sabatini, J. Taipale, N. Perrimon, and R. Bernards, "Minimizing the risk of reporting false

positives in large-scale RNAi screens," *Nature Methods*, vol. 3, no. 10, pp. 777–779, 2006.

[158] X. D. Zhang, "An effective method for controlling false discovery and false nondiscovery rates in genome-scale RNAi screens," *Journal of Biomolecular Screening*, vol. 15, no. 9, pp. 1116–1122, 2010.

[159] A. Birmingham, L. M. Selfors, T. Forster, D. Wrobel, C. J. Kennedy, E. Shanks, J. Santoyo-Lopez, D. J. Dunican, A. Long, D. Kelleher, Q. Smith, R. L. Beijersbergen, P. Ghazal, and C. E. Shamu, "Statistical methods for analysis of high-throughput RNA interference screens.," *Nature methods*, vol. 6, pp. 569–75, 8 2009.

[160] L. V. Gubareva, L. Kaiser, and F. G. Hayden, "Influenza virus neuraminidase inhibitors," 3 2000.

[161] "WHO Coronavirus (COVID-19) Dashboard — WHO Coronavirus (COVID-19) Dashboard With Vaccination Data."

[162] "Large clinical trial to study repurposed drugs to treat COVID-19 symptoms — National Institutes of Health (NIH)."

[163] S. Pushpakom, F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilliams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, and M. Pirmohamed, "Drug repurposing: Progress, challenges and recommendations," 2018.

[164] J. W. Scannell, A. Blanckley, H. Boldon, and B. Warrington, "Diagnosing the decline in pharmaceutical R&D efficiency," 2012.

[165] F. Touret, M. Gilles, K. Barral, A. Nougairède, E. Decroly, X. d. Lamballerie, and B. Coutard, "In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication," *bioRxiv*, 2020.

[166] L. Riva, S. Yuan, X. Yin, L. Martin-Sancho, N. Matsunaga, S. Burgstaller, L. Pache, P. D. Jesus, M. V. Hull, M. Chang, J. F. Chan, J. Cao, V. K.-M. Poon, K. Herbert, T.-T. Nguyen, Y. Pu, C. Nguyen, A. Rubanov, L. Martinez-Sobrido, W.-C. Lui, L. Miorin, K. White, J. R. Johnson, C. Benner, R. Sun, P. Schultz, A. I. Su, A. Garcia-Sastre, A. Chatterjee, K.-Y. Yuen, and S. Chanda, "A Large-scale Drug Repositioning Survey for SARS-CoV-2 Antivirals," *bioRxiv*, 2020.

[167] E. Bernhard, D. Bojkova, A. Zaliani, J. Cinatl, C. Claussen, S. Westhaus, J. Reinshagen, M. Kuzikov, M. Wolf, G. Geisslinger, P. Gribbon, and S. Ciesek, "Identification of inhibitors of SARS-CoV-2 in-vitro cellular toxicity in human (Caco-2) cells using a large scale drug repurposing collection," *Research Square*, 4 2020.

[168] D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Huettenhain, R. M. Kaake, A. L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B. J. Polacco, H. Braberg, J. M. Fabius, M. Eckhardt, M. Soucheray, M. J. Bennett, M. Cakir, M. J. Mcgregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. M. Kain, L. Miorin, E. Moreno, Z. Zar, C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I. T. Kirby, J. E. Melnyk, J. S. Chorba, K. Lou, S. A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C. J. P. Mathy, T. Perica, K. B. Pilla, S. J. Ganesan, D. J. Saltzberg, R. Rakesh, X. Liu, S. B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy-Lugo, Y. Lin, X.-P. Huang, Y. Liu, S. A. Wankowicz, M. Bohn, M. Safari, F. S. Ugur, C. Koh, N. Sadat Savar, Q. D. Tran, D. Shengjuler, S. J. Fletcher, M. C. O'neal, Y. Cai, J. C. J. Chang, D. J. Broadhurst, S. Klippsten, P. P. Sharp, N. A. Wenzell, D. Kuzuoglu, H.-Y. Wang, R. Trenker, J. M. Young, D. A. Cavero, J. Hiatt, T. L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R. M. Stroud, A. D. Frankel, O. S. Rosenberg, K. A. Verba, D. A. Agard, M. Ott, M. Emerman, N. Jura, M. Von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. D'enfert, S. Mukherjee, M. Jacobson, H. S. Malik, D. G. Fujimori, T. Ideker, C. S. Craik, S. N. Floor, J. S. Fraser, J. D. Gross, A. Sali, B. L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K. M. Shokat, B. K. Shoichet, and N. J. Krogan, "A SARS-CoV-2 protein interaction map reveals targets for drug repurposing A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature," *Nature*, pp. 1–13, 4 2020.

[169] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, and F. Cheng, "Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2," *Cell Discovery*, 2020.

[170] L. Perrin-Cocon, O. Diaz, C. Jacquemin, V. Barthel, E. Ogire, C. Ramière, P. André, V. Lotteau, and P. O. Vidalain, "The current landscape of coronavirus-host protein-protein interactions," 8 2020.

[171] C. Cava, G. Bertoli, and I. Castiglioni, "A protein interaction map identifies existing drugs targeting SARS-CoV-2," *BMC Pharmacology and Toxicology*, vol. 21, p. 65, 12 2020.

[172] H. D. Mitchell, A. J. Eisfeld, A. C. Sims, J. E. McDermott, M. M. Matzke, B. J. M. Webb-Robertson, S. C. Tilton, N. Tchitchek, L. Josset, C. Li, A. L. Ellis, J. H. Chang,

R. A. Heegel, M. L. Luna, A. A. Schepmoes, A. K. Shukla, T. O. Metz, G. Neumann, A. G. Benecke, R. D. Smith, R. S. Baric, Y. Kawaoka, M. G. Katze, and K. M. Waters, "A Network Integration Approach to Predict Conserved Regulators Related to Pathogenicity of Influenza and SARS-CoV Respiratory Viruses," *PLoS ONE*, 2013.

[173] M. A. Moni and P. Liò, "Network-based analysis of comorbidities risk during an infection: SARS and HIV case studies," *BMC Bioinformatics*, 2014.

[174] T. M. Murali, M. D. Dyer, D. Badger, B. M. Tyler, and M. G. Katze, "Network-based prediction and analysis of HIV dependency factors," *PLoS Computational Biology*, 2011.

[175] R. G. Ptak, W. Fu, B. E. Sanders-Beer, J. E. Dickerson, J. W. Pinney, D. L. Robertson, M. N. Rozanov, K. S. Katz, D. R. Maglott, K. D. Pruitt, and C. W. Dieffenbach, "¡i¿Short Communication:¡/i¿ Cataloguing the HIV Type 1 Human Protein Interaction Network," *AIDS Research and Human Retroviruses*, 2008.

[176] S. Shityakov, T. Dandekar, and C. Förster, "Gene expression profiles and protein-protein interaction network analysis in AIDS patients with HIV-associated encephalitis and dementia," *HIV/AIDS - Research and Palliative Care*, 2015.

[177] A. Vinayagam, T. E. Gibson, H.-J. Lee, B. Yilmazel, C. Roesel, Y. Hu, Y. Kwon, A. Sharma, Y.-Y. Liu, N. Perrimon, and A.-L. Barabási, "Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets," *Proceedings of the National Academy of Sciences*, vol. 113, no. 18, pp. 4976–4981, 2016.

[178] M. J. Korth, N. Tchitchek, A. G. Benecke, and M. G. Katze, "Systems approaches to influenza-virus host interactions and the pathogenesis of highly virulent and pandemic viruses.," *Seminars in immunology*, pp. 1–12, 12 2012.

[179] C. T. Lin, "Structural Controllability," *IEEE Transactions on Automatic Control*, vol. 19, no. 3, pp. 201–208, 1974.

[180] S. Wuchty, "Controllability in protein interaction networks," *Proceedings of the National Academy of Sciences*, 2014.

[181] J. E. Hopcroft and R. M. Karp, "An $n^{5/2} $ Algorithm for Maximum Matchings in Bipartite Graphs," *SIAM Journal on Computing*, 1973.

[182] T. Jia and A. L. Barabási, "Control capacity and a random sampling method in exploring controllability of complex networks," *Scientific Reports*, vol. 3, 2013.

[183] A. Vinayagam, U. Stelzl, R. Foulle, S. Plassmann, M. Zenkner, J. Timm, H. E. Assmus, M. A. Andrade-Navarro, and E. E. Wanker, "A directed protein interaction network for investigating intracellular signal transduction," *Science Signaling*, 2011.

[184] J. Yu and R. L. Finley, "Combining multiple positive training sets to generate confidence scores for protein-protein interactions," *Bioinformatics*, 2009.

[185] X. Zhang, T. Lv, and Y. Pu, "Input graph: The hidden geometry in controlling complex networks," *Scientific Reports*, 2016.

[186] S. A. Samarajiwa, S. Forster, K. Auchettl, and P. J. Hertzog, "INTERFEROME: The database of interferon regulated genes," *Nucleic Acids Research*, 2009.

[187] A. Krämer, J. Green, J. Pollard, and S. Tugendreich, "Causal analysis approaches in ingenuity pathway analysis," *Bioinformatics*, vol. 30, no. 4, pp. 523–530, 2014.

[188] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. MacIejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, and M. Wilson, "DrugBank 5.0: A major update to the DrugBank database for 2018," *Nucleic Acids Research*, 2018.

[189] S. Koyama, K. J. Ishii, C. Coban, and S. Akira, "Innate immune response to viral infection," *Cytokine*, vol. 43, pp. 336–341, 9 2008.

[190] M. R. Thompson, J. J. Kaminski, E. A. Kurt-Jones, and K. A. Fitzgerald, "Pattern recognition receptors and the innate immune response to viral infection," 2011.

[191] A. Iwasaki and R. Medzhitov, "Toll-like receptor control of the adaptive immune responses," 2004.

[192] G. N. Barber, "Host defense, viruses and apoptosis," 2001.

[193] B. J. Thomson, "Viruses and apoptosis," 2001.

[194] M. Gale Jr, S.-L. Tan, and M. G. Katze, "Translational control of viral gene expression in eukaryotes.," *Microbiology and Molecular Biology Reviews*, 2000.

[195] N. Sonenberg and A. G. Hinnebusch, "Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets," 2009.

[196] D. Walsh, M. B. Mathews, and I. Mohr, "Tinkering with translation: Protein synthesis in virus-infected cells," *Cold Spring Harbor Perspectives in Biology*, 2013.

[197] A. T. Pavia, "Viral infections of the lower respiratory tract: Old viruses, new viruses, and the role of diagnosis," *Clinical Infectious Diseases*, 2011.

[198] M. Jamaluddin, J. E. Wiktorowicz, K. V. Soman, I. Boldogh, J. D. Forbus, H. Spratt, R. P. Garofalo, and A. R. Brasier, "Role of Peroxiredoxin 1 and Peroxiredoxin 4 in Protection of Respiratory Syncytial Virus-Induced Cysteinyl Oxidation of Nuclear Cytoskeletal Proteins," *Journal of Virology*, 2010.

[199] J. Dubois, O. Terrier, and M. Rosa-Calatrava, "Influenza viruses and mRNA splicing: Doing more with less," 2014.

[200] N. Kumar, Z.-T. Xin, Y. Liang, H. Ly, and Y. Liang, "NF-kappaB signaling differentially regulates influenza virus RNA synthesis.," *Journal of virology*, vol. 82, no. 20, pp. 9880–9, 2008.

[201] S. Ludwig and O. Planz, "Influenza viruses and the NF-κB signaling pathway - Towards a novel concept of antiviral therapy," 2008.

[202] M. H. Schaefer, T. J. S. Lopes, N. Mah, J. E. Shoemaker, Y. Matsuoka, J.-F. Fontaine, C. Louis-Jeune, A. J. Eisfeld, G. Neumann, C. Perez-Iratxeta, Y. Kawaoka, H. Kitano, and M. a. Andrade-Navarro, "Adding protein context to the human protein-protein interaction network to reveal meaningful interactions.," *PLoS computational biology*, vol. 9, p. e1002860, 1 2013.

[203] H. Noh and R. Gunawan, "Inferring gene targets of drugs and chemical compounds from gene expression profiles," *Bioinformatics*, 2016.

[204] H. Noh, H. Ziyi, and R. Gunawan, "Inferring Causal Gene Targets from Time Course Expression Data," *IFAC-PapersOnLine*, 2016.

[205] H. Noh, J. E. Shoemaker, and R. Gunawan, "Network perturbation analysis of gene transcriptional profiles reveals protein targets and mechanism of action of drugs and influenza A viral infection," *Nucleic Acids Research*, 2018.

[206] S. Jain, J. Arrais, N. J. Venkatachari, V. Ayyavoo, and Z. Bar-Joseph, "Reconstructing the temporal progression of HIV-1 immune response pathways," *Bioinformatics*, 2016.

[207] P. D. Uchil, A. Hinz, S. Siegel, A. Coenen-Stass, T. Pertel, J. Luban, and W. Mothes, "TRIM Protein-Mediated Regulation of Inflammatory and Innate Immune Signaling and Its Association with Antiretroviral Activity," *Journal of Virology*, 2013.

[208] M. X. Jiang, X. Hong, B. B. Liao, S. Z. Shi, X. F. Lai, H. Y. Zheng, L. Xie, Y. Wang, X. L. Wang, H. B. Xin, M. Fu, and K. Y. Deng, "Expression profiling of TRIM protein family in THP1-derived macrophages following TLR stimulation," *Scientific Reports*, 2017.

[209] A. L. Totura, A. Whitmore, S. Agnihothram, A. Schäfer, M. G. Katze, M. T. Heise, and R. S. Baric, "Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection," *mBio*, 2015.

[210] S. Bauer, V. Groh, J. Wu, A. Steinle, J. H. Phillips, L. L. Lanier, and T. Spies, "Activation of NK cells and T cells by NKG2D, a receptor for stress- inducible MICA," *Science*, 1999.

[211] H. Jung, B. Hsiung, K. Pestal, E. Procyk, and D. H. Raulet, "RAE-1 ligands for the NKG2D receptor are regulated by E2F transcription factors, which controlcell cycle entry," *Journal of Experimental Medicine*, 2012.

[212] C. Bottino, R. Castriconi, D. Pende, P. Rivera, M. Nanni, B. Carnemolla, C. Cantoni, J. Grassi, S. Marcenaro, N. Reymond, M. Vitale, L. Moretta, M. Lopez, and A. Moretta, "Identification of PVR (CD155) and Nectin-2 (CD112) as cell surface ligands for the human DNAM-1 (CD226) activating molecule," *Journal of Experimental Medicine*, 2003.

[213] K. E. Sloan, B. K. Eustace, J. K. Stewart, C. Zehetmeier, C. Torella, M. Simeone, J. E. Roy, C. Unger, D. N. Louis, L. L. Ilag, and D. G. Jay, "CD155/PVR plays a key role in cell motility during tumor cell invasion and migration," *BMC Cancer*, 2004.

[214] M. B. Oldstone and H. Rosen, "Cytokine storm plays a direct role in the morbidity and mortality from influenza virus infection and is chemically treatable with a single sphingosine-1-phosphate agonist molecule," *Current Topics in Microbiology and Immunology*, 2014.

[215] Y. Fu, Y. Cheng, and Y. Wu, "Understanding SARS-CoV-2-Mediated Inflammatory Responses: From Mechanisms to Potential Therapeutic Tools," 2020.

[216] E. Scarselli, H. Ansuini, R. Cerino, R. M. Roccasecca, S. Acali, G. Filocamo, C. Traboni, A. Nicosia, R. Cortese, and A. Vitelli, "The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus," *EMBO Journal*, 2002.

[217] B. Bartosch, A. Vitelli, C. Granier, C. Goujon, J. Dubuisson, S. Pascale, E. Scarselli, R. Cortese, A. Nicosia, and F. L. Cosset, "Cell Entry of Hepatitis C Virus Requires a Set of Co-receptors that Include the CD81 Tetraspanin and the SR-B1 Scavenger Receptor," *Journal of Biological Chemistry*, 2003.

[218] Y. Kawasaki, A. Nakagawa, K. Nagaosa, A. Shiratsuchi, and Y. Nakanishi, "Phosphatidylserine binding of class B scavenger receptor type I, a phagocytosis receptor of testicular sertoli cells," *Journal of Biological Chemistry*, 2002.

[219] W. Witt, I. Kolleck, H. Fechner, P. Sinha, and B. Rüstow, "Regulation by vitamin E of the scavenger receptor BI in rat liver and HepG2 cells," *Journal of Lipid Research*, 2000.

[220] M. Hayek, S. Taylor, B. Bender, S. Han, M. Meydani, D. Smith, S. Eghtesada, and S. Meydani, "Vitamin E Supplementation Decreases Lung Virus Titers in Mice Infected with Influenza," *The Journal of Infectious Diseases*, 1997.

[221] A. van Nieuwenhuijze, O. Burton, P. Lemaitre, A. E. Denton, A. Cascalho, R. E. Goodchild, B. Malengier-Devlies, B. Cauwe, M. A. Linterman, S. Humblet-Baron, and A. Liston, "Mice Deficient in Nucleoporin Nup210 Develop Peripheral T Cell Alterations," *Frontiers in immunology*, 2018.

[222] Y. Yang, M. Wislez, N. Fujimoto, L. Prudkin, J. G. Izzo, F. Uno, L. Ji, A. E. Hanna, R. R. Langley, D. Liu, F. M. Johnson, I. Wistuba, and J. M. Kurie, "A selective small molecule inhibitor of c-Met, PHA-665752, reverses lung premalignancy induced by mutant K-ras," *Molecular Cancer Therapeutics*, 2008.

[223] T. Hayashi and T. P. Su, "Sigma-1 Receptor Chaperones at the ER- Mitochondrion Interface Regulate Ca2+ Signaling and Cell Survival," *Cell*, 2007.

[224] L. X. Liao, X. M. Song, L. C. Wang, H. N. Lv, J. F. Chen, D. Liu, G. Fu, M. B. Zhao, Y. Jiang, K. W. Zeng, and P. F. Tu, "Highly selective inhibition of IMPDH2 provides the basis of antineuroinflammation therapy," *Proceedings of the National Academy of Sciences of the United States of America*, 2017.

[225] M. Morita, L. W. Ler, M. R. Fabian, N. Siddiqui, M. Mullin, V. C. Henderson, T. Alain, B. D. Fonseca, G. Karashchuk, C. F. Bennett, T. Kabuta, S. Higashi, O. Larsson, I. Topisirovic, R. J. Smith, A.-C. Gingras, and N. Sonenberg, "A Novel 4EHP-GIGYF2 Translational Repressor Complex Is Essential for Mammalian Development," *Molecular and Cellular Biology*, 2012.

[226] J. Kindrachuk, B. Ork, B. J. Hart, S. Mazur, M. R. Holbrook, M. B. Frieman, D. Traynor, R. F. Johnson, J. Dyall, J. H. Kuhn, G. G. Olinger, A. L. E. Hensley, and P. B. Jahrlinga, "Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome analysis," *Antimicrobial Agents and Chemotherapy*, 2015.

[227] R. L. Stead and C. G. Proud, "Rapamycin enhances eIF4E phosphorylation by activating MAP kinase-interacting kinase 2a (Mnk2a)," *FEBS Letters*, 2013.

[228] E. E. Ackerman, J. F. Alcorn, T. Hase, and J. E. Shoemaker, "A dual controllability analysis of influenza virus-host protein-protein interaction networks for antiviral drug target discovery," *BMC Bioinformatics*, vol. 20, p. 297, 12 2019.

[229] R. A. Saenz, M. Quinlivan, D. Elton, S. MacRae, A. S. Blunden, J. A. Mumford, J. M. Daly, P. Digard, A. Cullinane, B. T. Grenfell, J. W. McCauley, J. L. N. Wood, and J. R. Gog, "Dynamics of Influenza Virus Infection and Pathology," *Journal of Virology*, vol. 84, no. 8, pp. 3974–3983, 2010.

[230] P. Baccam, C. Beauchemin, C. A. Macken, F. G. Hayden, and A. S. Perelson, "Kinetics of Influenza A Virus Infection in Humans," *Journal of Virology*, vol. 80, pp. 7590–7599, 8 2006.

[231] A. Handel, I. M. Longini, and R. Antia, "Towards a quantitative understanding of the within-host dynamics of influenza A infections," *Journal of the Royal Society Interface*, vol. 7, pp. 35–47, 6 2009.

[232] G. A. Bocharov and A. A. Romanyukha, "Numerical treatment of the parameter identification problem for delay-differential systems arising in immune response modelling," *Applied Numerical Mathematics*, vol. 15, no. 3, pp. 307–326, 1994.

[233] A. Isaacs and J. Lindenmann, "Virus interference. I. The interferon," *Journal of Interferon Research*, vol. 7, no. 5, pp. 429–438, 1987.

[234] G. Trinchieri, "Type I interferon: Friend or foe?," 9 2010.

[235] K. Y. Lee, J. W. Rhim, and J. H. Kang, "Hyperactive immune cells (T cells) may be responsible for acute lung injury in influenza virus infections: A need for early immune-modulators for severe cases," *Medical Hypotheses*, vol. 76, pp. 64–69, 1 2011.

[236] J. S. M. Peiris, C. Y. Cheung, C. Y. H. Leung, and J. M. Nicholls, "Innate immune responses to influenza A H5N1: friend or foe?," 12 2009.

[237] B. J. Zheng, K. W. Chan, Y. P. Lin, G. Y. Zhao, C. Chan, H. J. Zhang, H. L. Chen, S. S. Wong, S. K. Lau, P. C. Woo, K. H. Chan, D. Y. Jin, and K. Y. Yuen, "Delayed antiviral plus immunomodulator treatment still reduces mortality in mice infected by high inoculum of influenza A/H5N1 virus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 8091–8096, 6 2008.

[238] C. Cilloniz, M. J. Pantin-Jackwood, C. Ni, A. G. Goodman, X. Peng, S. C. Proll, V. S. Carter, E. R. Rosenzweig, K. J. Szretter, J. M. Katz, M. J. Korth, D. E. Swayne, T. M. Tumpey, and M. G. Katze, "Lethal dissemination of H5N1 influenza virus is associated with dysregulation of inflammation and lipoxin signaling in a mouse model of infection.," *Journal of virology*, vol. 84, pp. 7613–24, 8 2010.

[239] V. D. Menachery, A. J. Eisfeld, A. Schäfer, L. Josset, A. C. Sims, S. Proll, S. Fan, C. Li, G. Neumann, S. C. Tilton, J. Chang, L. E. Gralinski, C. Long, R. Green, C. M. Williams, J. Weiss, M. M. Matzke, B. J. Webb-Robertson, A. A. Schepmoes, A. K. Shukla, T. O. Metz, R. D. Smith, K. M. Waters, M. G. Katze, Y. Kawaoka, and R. S. Baric, "Pathogenic influenza viruses and coronaviruses utilize similar and contrasting approaches to control interferon-stimulated gene responses," *mBio*, vol. 5, 5 2014.

[240] K. B. Walsh, J. R. Teijaro, H. Rosen, and M. B. Oldstone, "Quelling the storm: Utilization of sphingosine-1-phosphate receptor signaling to ameliorate influenza virus-induced cytokine storm," *Immunologic Research*, vol. 51, pp. 15–25, 10 2011.

[241] R. J. Snelgrove, A. Godlee, and T. Hussell, "Airway immune homeostasis and implications for influenza-induced inflammation," 7 2011.

[242] M. D. de Jong, C. P. Simmons, T. T. Thanh, V. M. Hien, G. J. D. Smith, T. N. B. Chau, D. M. Hoang, N. V. V. Chau, T. H. Khanh, V. C. Dong, P. T. Qui, B. V. Cam, D. Q. Ha, Y. Guan, J. S. M. Peiris, N. T. Chinh, T. T. Hien, and J. Farrar, "Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia.," *Nature medicine*, vol. 12, pp. 1203–7, 10 2006.

[243] M. E. Falagas, E. K. Vouloumanou, E. Baskouta, P. I. Rafailidis, K. Polyzos, and J. Rello, "Treatment options for 2009 H1N1 influenza: Evaluation of the published evidence," 5 2010.

[244] M. J. Carter, "A rationale for using steroids in the treatment of severe cases of H5N1 avian influenza," 7 2007.

[245] K. Shinya, T. Okamura, S. Sueta, N. Kasai, M. Tanaka, T. E. Ginting, A. Makino, A. J. Eisfeld, and Y. Kawaoka, "Toll-like receptor pre-stimulation protects mice against lethal infection with highly pathogenic influenza viruses," *Virology Journal*, vol. 8, 2011.

[246] C. W. Cluff, J. R. Baldridge, A. G. Stöver, J. T. Evans, D. A. Johnson, M. J. Lacy, V. G. Clawson, V. M. Yorgensen, C. L. Johnson, M. T. Livesay, R. M. Hershberg, and D. H. Persing, "Synthetic toll-like receptor 4 agonists stimulate innate resistance to infectious challenge," *Infection and Immunity*, vol. 73, pp. 3044–3052, 5 2005.

[247] A. Tanaka, S. Nakamura, M. Seki, K. Fukudome, N. Iwanaga, Y. Imamura, T. Miyazaki, K. Izumikawa, H. Kakeya, K. Yanagihara, and S. Kohno, "Toll-like receptor 4 agonistic antibody promotes innate immunity against severe pneumonia induced by coinfection with influenza virus and Streptococcus pneumoniae," *Clinical and Vaccine Immunology*, vol. 20, pp. 977–985, 7 2013.

[248] J. Wong, M. Christopher, S. Viswanathan, X. Dai, A. Salazar, L.-Q. Sun, and M. Wang, "Antiviral Role of Toll-Like Receptor-3 Agonists Against Seasonal and Avian Influenza Viruses," *Current Pharmaceutical Design*, vol. 15, pp. 1269–1274, 3 2009.

[249] C. Cillóniz, K. Shinya, X. Peng, M. J. Korth, S. C. Proll, L. D. Aicher, V. S. Carter, J. H. Chang, D. Kobasa, F. Feldmann, J. E. Strong, H. Feldmann, Y. Kawaoka, and M. G. Katze, "Lethal influenza virus infection in macaques is associated with early dysregulation of inflammatory related genes.," *PLoS pathogens*, vol. 5, no. 10, 2009.

[250] D. Kobasa, S. M. Jones, K. Shinya, J. C. Kash, J. Copps, H. Ebihara, Y. Hatta, J. H. Kim, P. Halfmann, M. Hatta, F. Feldmann, J. B. Alimonti, L. Fernando, Y. Li, M. G. Katze, H. Feldmann, and Y. Kawaoka, "Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus," *Nature*, vol. 445, pp. 319–323, 1 2007.

[251] G. A. Bocharov and A. A. Romanyukha, "Mathematical model of antiviral immune response III. Influenza a virus infection," *Journal of Theoretical Biology*, vol. 167, no. 4, pp. 323–360, 1994.

[252] J. Guarner and R. Falcón-Escobedo, "Comparison of the pathology caused by H1N1, H5N1, and H3N2 influenza viruses," 2009.

[253] C. Korteweg and J. Gu, "Pandemic influenza a (H1N1) virus infection and avian influenza a (H5N1) virus infection: A comparative analysis," 2010.

[254] C. Xia, M. Vijayan, C. J. Pritzl, S. Y. Fuchs, A. B. McDermott, and B. Hahm, "Hemagglutinin of Influenza A Virus Antagonizes Type I Interferon (IFN) Responses by Inducing Degradation of Type I IFN Receptor 1," *Journal of Virology*, vol. 90, pp. 2403–2417, 3 2016.

[255] F. Ma, B. Li, Y. Yu, S. S. Iyer, M. Sun, and G. Cheng, " Positive feedback regulation of type I interferon by the interferonstimulated gene STING ," *EMBO reports*, vol. 16, pp. 202–212, 2 2015.

[256] C. Y. Wu, H. Y. Chuang, and C. H. Wong, "Influenza virus neuraminidase regulates host CD8+ T-cell response in mice," *Communications Biology*, vol. 3, pp. 1–10, 12 2020.

[257] W. Li, G. Wang, H. Zhang, G. Xin, D. Zhang, J. Zeng, X. Chen, Y. Xu, Y. Cui, and K. Li, "Effects of NS1 variants of H5N1 influenza virus on interferon induction, TNFα response and p53 activity," *Cellular and Molecular Immunology*, vol. 7, pp. 235–242, 5 2010.

[258] I. Meunier and V. von Messling, "NS1-mediated delay of type I interferon induction contributes to influenza A virulence in ferrets," *Journal of General Virology*, 2011.

[259] A. S. Perelson, F. G. Hayden, C. A. Macken, C. Beauchemin, and P. Baccam, "Kinetics of Influenza A Virus Infection in Humans," *Journal of Virology*, 2006.

[260] L. Pinky and H. M. Dobrovolny, "The impact of cell regeneration on the dynamics of viral coinfection," *Chaos*, 2017.

[261] A. Handel, I. M. Longini, and R. Antia, "Antiviral resistance and the control of pandemic influenza: The roles of stochasticity, evolution and model details," *Journal of Theoretical Biology*, vol. 256, no. 1, pp. 117–125, 2009.

[262] P. Woo, E. Tung, K. Chan, C. Lau, S. Lau, and K. Yuen, "Cytokine Profiles Induced by the Novel SwineOrigin Influenza A/H1N1 Virus: Implications for Treatment Strategies," *The Journal of Infectious Diseases*, 2009.

[263] K. P. Hui, H. S. Li, M. C. Cheung, R. W. Chan, K. M. Yuen, C. K. Mok, J. M. Nicholls, J. S. Peiris, and M. C. Chan, "Highly pathogenic avian influenza H5N1 virus delays apoptotic responses via activation of STAT3," *Scientific Reports*, 2016.

[264] WHO, "FAQs: H5N1 influenza," 2011.

[265] K. L. Lin, Y. Suzuki, H. Nakano, E. Ramsburg, and M. D. Gunn, " CCR2 + Monocyte-Derived Dendritic Cells and Exudate Macrophages Produce Influenza-Induced Pulmonary Immune Pathology and Mortality ," *The Journal of Immunology*, vol. 180, pp. 2562–2572, 2 2008.

[266] S. L. Deshmane, S. Kremlev, S. Amini, and B. E. Sawaya, "Monocyte chemoattractant protein-1 (MCP-1): An overview," 2009.

[267] Z. S. Singer, P. M. Ambrose, T. Danino, and C. M. Rice, "Quantitative measurements of early alphaviral replication dynamics in single cells reveals the basis for superinfection exclusion," *Cell Systems*, 2021.

[268] T. Yoshimura, E. A. Robinson, S. Tanaka, E. Appella, and E. J. Leonard, "Purification and amino acid analysis of two human monocyte chemoattractants produced by phytohemagglutinin-stimulated human blood mononuclear leukocytes.," *Journal of immunology (Baltimore, Md. : 1950)*, 1989.

[269] J. N. Weiss, "The Hill equation revisited: uses and misuses," *The FASEB Journal*, vol. 11, pp. 835–841, 9 1997.

[270] S. L. Waldrop, K. A. Davis, V. C. Maino, and L. J. Picker, "Normal Human CD4+ Memory T Cells Display Broad Heterogeneity in Their Activation Threshold for Cytokine Synthesis," *The Journal of Immunology*, vol. 161, no. 10, 1998.

[271] Y. Itoh and R. N. Germain, "Single cell analysis reveals regulated hierarchical T cell antigen receptor signaling thresholds and intraclonal heterogeneity for individual cytokine responses of CD4+ T cells," *Journal of Experimental Medicine*, vol. 186, pp. 757–766, 8 1997.

[272] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[273] W. Pan, "Akaike's information criterion in generalized estimating equations," *Biometrics*, vol. 57, pp. 120–125, 3 2001.

[274] S. Portet, "A primer on model selection using the Akaike Information Criterion," *Infectious Disease Modelling*, vol. 5, pp. 111–128, 1 2020.

[275] W. Sakamoto, "Bias-reduced marginal Akaike information criteria based on a Monte Carlo method for linear mixed-effects models," *Scandinavian Journal of Statistics*, vol. 46, pp. 87–115, 3 2019.

[276] A. Saltelli, R. B. C. S. . D. Analysis, and u. 1998, "An alternative way to compute Fourier amplitude sensitivity test (FAST)," *Elsevier*.

[277] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology," 9 2008.

[278] L. Hao, Q. He, Z. Wang, M. Craven, M. A. Newton, and P. Ahlquist, "Limited Agreement of Independent RNAi Screens for Virus-Required Host Genes Owes More to False-Negative than False-Positive Factors," *PLoS Computational Biology*, vol. 9, 9 2013.

[279] Y. Gong and Z. Zhang, "Alternative signaling pathways: When, where and why?," *FEBS Letters*, vol. 579, pp. 5265–5274, 10 2005.

[280] G. Weng, U. S. Bhalla, and R. Iyengar, "Complexity in biological signaling systems.," *Science (New York, N.Y.)*, vol. 284, pp. 92–6, 4 1999.

[281] H. Kitano, "Biological robustness," *Nature Reviews Genetics*, vol. 5, pp. 826–837, 11 2004.

[282] M. Csete and J. Doyle, "Bow ties, metabolism and disease.," *Trends in biotechnology*, vol. 22, pp. 446–50, 9 2004.

[283] N. Kumar, Z.-t. Xin, Y. Liang, H. Ly, and Y. Liang, "NF-B Signaling Differentially Regulates Influenza Virus RNA Synthesis," *Journal of Virology*, 2008.

[284] M. Uhart, G. Flores, and D. M. Bustos, "Controllability of protein-protein interaction phosphorylation-based networks: Participation of the hub 14-3-3 protein family," *Scientific Reports*, 2016.

[285] K. P. Y. Hui, S. M. Y. Lee, C.-y. Cheung, I. H. Y. Ng, L. L. M. Poon, Y. Guan, N. Y. Y. Ip, A. S. Y. Lau, and J. S. M. Peiris, "Induction of Proinflammatory Cytokines in Primary Human Macrophages by Influenza A Virus (H5N1) Is Selectively Regulated by IFN Regulatory Factor 3 and p38 MAPK," *The Journal of Immunology*, vol. 182, pp. 1088–1098, 1 2009.