The M^2 Pipeline: A Novel Tool for Rapidly Analyzing Host Metabolic Profiles from RNA-seq Data

Ryan Baker
NIMML Institute
Blacksburg, USA
rbaker@nimml.org

Josep Bassaganya-Riera
NIMML Institute
Blacksburg, USA
jbassaga@nimml.org

Nuria Tubau-Juni
NIMML Institute
Blacksburg, USA
ntubau@nimml.org

Andrew J. Leber
NIMML Institute
Blacksburg, USA
ajleber@nimml.org

Raquel Hontecillas
NIMML Institute
Blacksburg, USA
rmagarzo@nimml.org

2023 11th International Conference on Bioinformatics and Computational Biology (ICBCB) | 979-8-3503-9787-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICBCB57893.2023.10246649

***Abstract*— A novel bioinformatics and data analytics pipeline that utilizes global gene expression datasets in combination with advanced computational modeling of metabolic pathways has been applied to predicting the metabolic profiles during viral infections. The Modeling Metabolism (M^2) pipeline was used to analyze and compare complex metabolic profiles of four viral respiratory infections (rhinovirus infection, flu, COVID-19, and MERS-CoV infection) based on host RNA-seq datasets. During infection, multiple factors can affect the metabolic state of the host. Metabolic changes are biomarkers of host antiviral responses and viral pathogenesis. Therefore, analyzing the metabolic response to infectious diseases might provide valuable mechanistic insights into host-pathogen interactions and highlight metabolic changes that play important roles in disease pathogenesis. By using the M^2 pipeline, we predicted that influenza virus, SARS-CoV-2, and MERS-CoV infections resulted in increased glycolytic activities and a reduced capacity for oxidative phosphorylation. Furthermore, SARS-CoV-2 and MERS-CoV infections both presented dysregulated pentose phosphate pathways (PPP), while influenza infection elicited an increase in PPP activity in correlation with disease severity and mortality. Notably, rhinovirus infection, the mildest respiratory infection studied, had little effect on the overall perturbation of host cellular metabolism. The M^2 pipeline provides a rapid, comprehensive, and systems-wide analysis of metabolic profiles from host response RNA-seq datasets during emerging and re-emerging infections.**

Keywords—Bioinformatics, Computational Biology, Metabolism, Infectious Disease, Immunology, RNA-seq

I. Introduction

Basal metabolic rates found in homeostasis can experience significant shifts during disease. The upregulation or downregulation of metabolic pathways leads to measurable changes in intermediate metabolites and end products which have a drastic impact on cellular phenotype and function. These metabolic changes are the result of cells responding to environmental cues and adapting to modifications of energetic demands to increase the synthesis of biomolecules. For instance, in homeostasis, when energy demand within the cell is low, pyruvate is transported into the mitochondria to fuel oxidative phosphorylation, producing ATP efficiently. Increases in energetic demands during disease lead to glucose being metabolized to lactate to facilitate higher rates of ATP production via glycolytic flux. Here, we present a computational model of metabolism that is comprised of massively and dynamically interacting metabolic pathways associated with infection and immunity[1].

Viral infections result in significant alterations of cellular metabolism to meet the increased energetic demands for the *de novo* synthesis of biomolecules during viral replication to produce new viral progeny[2]. More severe respiratory infections can cause respiratory distress, leading to hypoxia and increased production of reactive oxygen species[3]. To maintain cellular energy production during periods of low oxygen supply, HIF1-a causes an upregulation of glycolytic activity, increasing the concentration of lactate[4]. Often, increased tissue concentrations of lactate are associated with inflammation and greater oxidative stress[5]. The pentose phosphate pathway (PPP) is critical for the synthesis of precursors for nucleotides and amino acids, as well as to produce NADPH, a key neutralizer of reactive oxygen species (ROS). During viral infections, increased PPP activity can be an indicator of disease severity, as the PPP can be hijacked to increase the production of nucleotides and amino acids for virus replication[2]. Fatty acid oxidation involves short chain (< 16) fatty acids entering the mitochondria to be broken down to produce acetyl-coA, driving the TCA cycle and oxidative phosphorylation. Infection with SARS-CoV-2 and Influenza viruses decrease the ability of fatty acids to enter the mitochondria via CPT1 and CPT2-mediated transport, respectively. Decreased access to energy obtained from fatty acids, makes cells rely more heavily on glycolysis, further favoring viral replication[4, 6]. Glutaminolysis and the malate aspartate shuttle regulate the concentration of metabolites within the mitochondria and are necessary for producing precursors for gluconeogenesis. Glutaminolysis utilizes glutamine to resupply the TCA cycle via the production of a-KG and is required for influenza infection[7]. During SARS-CoV-2 infection, increased citrate export from the mitochondria to fuel fatty acid synthesis demands increased TCA cycle metabolite replenishment via glutaminolysis[8]. The malate-aspartate shuttle is involved in the transfer of metabolites from the TCA cycle into gluconeogenesis, as well as the transfer of energy to the mitochondria by shuttling energy stored in cytosolic NADH to its mitochondrial counterpart, also an important regulating pathway for the balance of TCA cycle metabolites. The analysis of these pathways as a massively and dynamically interacting network is important to understanding alterations of mitochondrial metabolism and how these changes can lead to mitochondrial dysfunction during viral infections[9, 10].

This work was funded by a grant from the Defense Threat Reduction Agency (HDTRA1-20-1-0021) to JB-R and RH

Given the importance of metabolic changes in viral infections, we have developed a computational model that enables a quick, system-wide analysis of metabolic responses by using RNA-seq data. We have applied our novel Modeling Metabolism (M^2) bioinformatic pipeline to perform a comprehensive comparative evaluation of metabolic shifts during infection with various viral strains, including rhinovirus, Influenza A virus, SARS-CoV-2, and MERS-CoV from RNA-seq host response data. The concentration and fluxes of metabolites are key values for understanding preferred metabolic pathways and the overall metabolic status of an infected host. These values are difficult to obtain via experimental methods and measuring all metabolites cannot be done at the same time[11]. Computational modeling uses well-established rate laws to model chemical reactions[12, 13]. Concentrations of biochemical species can be estimated given a system of reactions and experimentally derived reaction parameters. Several previous models have successfully estimated the kinetics associated with pathways in metabolism [14, 15]. The complexity and interplay between metabolic pathways, requires an approach that considers the dynamics of the system, as opposed to a static analysis, to provide a global, system-wide metabolic profile during infection. We identified several gene expression datasets for infectious respiratory diseases, and ran them through our high throughput, advanced bioinformatic and computational modeling M^2 pipeline. As a result, we obtained large amounts of output data in the form of predicted concentrations of enzymes and metabolites in the analyzed biosamples, that were subsequently filtered to generate novel hypotheses about the host immunometabolic responses, including potential metabolic hijacking, for each infection. Here, we show multiple model outputs to demonstrate how this M^2 pipeline can be used to shed new light on how host metabolism is affected by infectious respiratory disease.

Infectious disease epidemics or pandemics caused by newly emerging viruses are open-ended threats to public health systems, most recently emphasized by COVID-19. Only in the past 20 years, the WHO declared epidemic transmission of newly identified strains of coronavirus SARS-CoV, MERS-CoV, SARS-CoV-2, swine-origin Influenza H1N1, Ebola and Zika virus [16]. Despite strategic planning by international organizations and states, with their unpredictable presentation, these viruses have caused significant disruption. In addition to emerging and re-emerging viruses on the surveillance radar, it is estimated that there are half million unidentified mammalian viruses in wildlife reservoirs that could potentially jump the species barrier into humans [17]. The chances of these rare events are increasing due to over-stretching of ecological niches by human activity**.** A robust bioinformatics and data analytics pipeline such as the M^2 allow for a rapid, system-wide and comprehensive analysis of new epidemiological threats, increases the amount of biological information that can be extracted from time course RNA-seq experiments, and provides a platform for performing additional *in silico* experiments for obtaining preliminary data or hypothesis building.

II. Methods

A. Data Accession and Processing

The M^2 pipeline was developed as a tool for predicting the host metabolic responses to various stimuli. We gathered, developed, and layered multiple methodologies together to create a pipeline that transformed gene expression data into a simulation of metabolic reactions. Gene expression data for four respiratory infectious diseases were obtained from archives found on GEO database [18]. The viruses included in the analysis are Rhinovirus, Influenza virus, SARS-CoV-2, and MERS-CoV. Information pertaining to the strains and experimental design can be found on Table I. A transcripts per million (TPM) transformation was performed to normalize the data, accounting for gene length and provided a basis to compare samples from the same study. The TPM values were then plotted and modified such that the area under the curve of TPM values was equal to the total concentration of proteins within a cell. This transformation allowed for protein concentrations to be assumed for each gene of interest. Using a study outline that listed multiple isoforms associated with each metabolic gene of interest, each study with time-course data had the enzymes of interest plotted against polynomial equations of orders 0 through 5 using the optimize package within Scipy for Python. The processing of the RNA-seq data to polynomial equations involved matching the appropriate naming conventions for each data for the selection of each enzyme (e.g., RefSeq or Ensembl). Analyst intervention was necessary at this portion of the pipeline to weigh the R^2 values of the fits against potential overfitting of the data. Isoforms were ranked by their total expression by the pipeline and if no intervention based on tissue sample or specific function is performed by the analyst, the highest concentration was selected to be input to the model. After the polynomial equations were chosen for each enzyme, they were implemented into the base model of metabolism using python scripts that replaced constant values with polynomial assignments of concentration based on the internal time of the simulation. This process created a standalone SBML compliant model that was used to run a time-course simulation of metabolism.

TABLE I. Virus Studies Information

Virus	Strain	Time Points (days post-infection)	Tissue	Model Organism	GSE #
Rhinovirus	RV-A1	0, 1, 2, 3, 4	Bronchial epithelial cells	Homo sapiens	GSE146532
Influenza (H5N1)	A/Vietnam/1203/2004	0, 1, 2, 4, 7	Lungs	Macaca mulatta	GSE33351
Influenza (H1N1)	A/Texas/36/91	0, 1, 2, 4, 7	Lungs	Macaca mulatta	GSE33351
Influenza (1918 HANA)	A/Texas/36/91 + (HA, NA)	0, 1, 2, 4, 7	Lungs	Macaca mulatta	GSE33351
Influenza (1918 HANANS)	A/Texas/36/91 + (HA, NA, NS)	0, 1, 2, 4, 7	Lungs	Macaca mulatta	GSE33351
SARS-CoV-2	USA-WA1/2020)	0, 1, 2, 4, 6, 8, 14	Lungs	Mesocricetus auratus	GSE161200
MERS-CoV	HCoV-EMC/2012	0, 3, 4, 6	Lungs	Callithric jacchus	GSE55023

B. The Metabolic Model

The base model of metabolism was a system of differential equations with 131 species and 57 reactions (Figure 1)[1]. Many highly conserved mammalian metabolic pathways were incorporated into the model using a systems biology approach. Glycolysis, gluconeogenesis, the pentose phosphate pathway, the TCA cycle, fatty acid oxidation, oxidative phosphorylation, the malate-aspartate shuttle, and glutaminolysis were all included in the network of the model. Glucose, fatty acid, and glutamine concentrations were considered the model inputs in addition to the enzyme concentration equations, which were estimated from experimental values obtained from the RNA-seq data. The concentration of each metabolite and fluxes of each reaction were considered the model outputs. To estimate the rate

of a reaction, we considered the dissociation constant of the enzyme and substrate, the maximum enzymatic activity, and the substrate and enzyme concentrations. The dissociation constant (kd) and maximum enzymatic activity (kcat) was derived from experimental data that had been obtained over the course of decades of experimental biological research, and compiled into the SABIO-h, Biochemical Reaction Kinetics Database[19]. The reaction enzyme commission (EC) number was used to identify the appropriate rate laws on the SABIO-h database. External substrate concentrations were estimated using experimentally derived values[20, 21] and enzyme concentrations were derived from gene expression data, which required an experimentally derived dataset (RNA-seq) and the transformation described in the previous section. Reactions within the model used rate law equations to determine the flux of each reaction. The rate laws used were Michaelis Menten[12], ping-pong bi-bi, or Hill-like equations[13]. Time-course simulations were run using PyCoTools package for python. The duration of the simulation corresponded to the length of each of the time-course RNA-seq studies and resolution time associated with each disease (Rhinovirus: 96 hours, Influenza: 160 hours, Covid-19: 320 hours, MERS: 140 hours). Each species within the model was plotted separately and compared to the uninfected control simulation and other strains if applicable, which allowed for direct comparisons between infected simulations and the uninfected control.

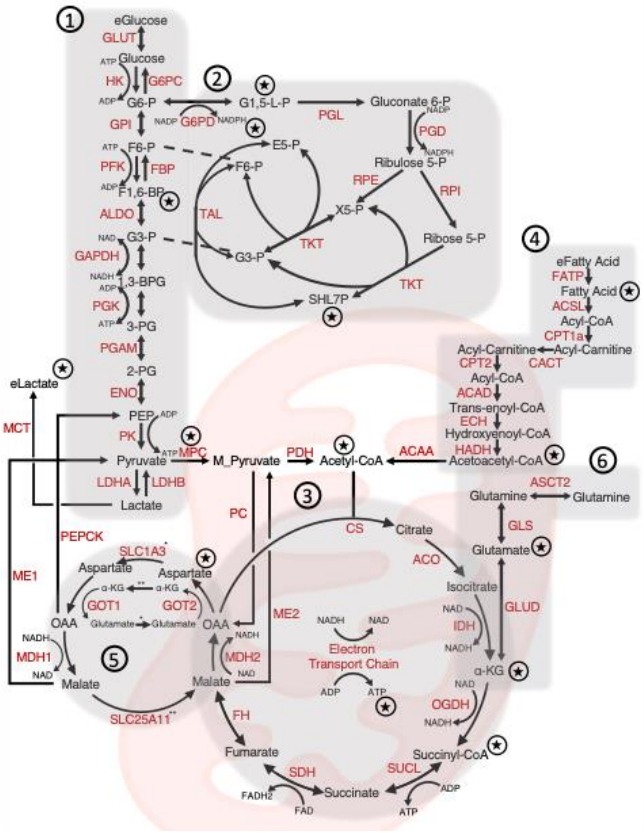

Fig. 1. Network diagram of metabolic pathways included in the computational model of metabolism. The metabolic pathways represented in the model are glycolysis (1), pentose phosphate pathway (2), citric acid cycle and oxidative phosphorylation (3), fatty acid oxidation (4), malate-aspartate shuttle (5) and glutaminolysis (6). Species for which simulation results are reported are marked with asterisks.

III. Results

A. SARS-CoV-2 and MERS-CoV Infection Result in Highly Elevated Lactate Production in Host Cells

To analyze how glycolysis was altered by the four viral infections, we highlight the glycolytic pathway within our model and focus on three key portions, F1,6-BP, MPC, and extracellular lactate. The concentration of F1,6-BP shows how metabolites are progressing past the major glycolytic regulator, PFK. MPC is the enzyme responsible for transporting pyruvate into the mitochondria, allowing for higher activity of the TCA cycle and fewer pyruvate molecules being reduced to form lactate. Extracellular lactate shows how much complete glycolysis has occurred.

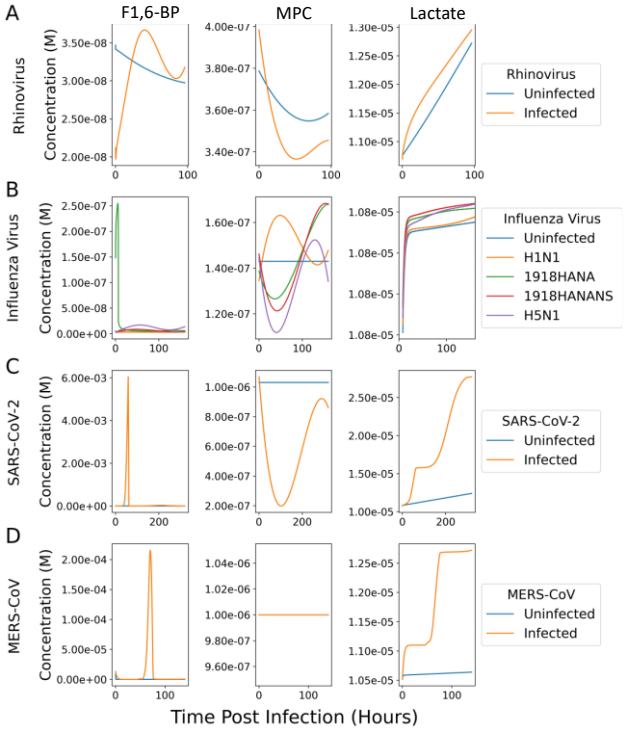

Fig. 2. Analysis of glycolysis during infectious respiratory diseases. Time-course plots comparing the simulated concentrations of F1,6-BP, MPC, and Lactate (Extracellular) during infection by (a) Rhinovirus (b) Influenza virus strains (c) SARS-CoV-2, and (d) MERS-CoV. For each virus type analyzed, an uninfected control was included for reference of baseline metabolic level.

The rhinovirus infection simulation shows a slight increase in glycolytic flux, indicated by increased lactate production (Figure 2a). The Influenza virus infection simulations showed little change between lactate production of the control host and the host infected by the A/Texas/36/91 strain, while infections by the A/Vietnam/1203/2004, A/Texas/36/91 + (HA, NA), and A/Texas/36/91 + (HA, NA, NS) strains all showed more elevated lactate production (Figure 2b).

SARS-CoV-2 infection experienced a burst-like increase in F1,6-BP concentration at 45 hours post-infection, followed by a lower, less abrupt increase at 200 hours or approximately 8 days post infection (DPI). Simulated concentration of MPC was lower in SARS-CoV-2 infected lungs but recovered to control levels by the end of the simulation. This response is consistent with an initial response to disease, followed by a recovery phase.

The decrease in transport capacity for pyruvate into the mitochondria was significant enough that it likely contributed to the increased lactate production. Concentration of extracellular lactate has two major time frames in which it increased drastically, centered around 45 hours and 200 hours (2 and 8 DPI). These increases corresponded to the upstream bursts of glycolytic metabolite production shown by F1,6-BP (Figure 2c). MERS-CoV also experienced a burst-like increase in F1,6-BP concentration at approximately 65 hours (2 DPI). MPC concentration for both the infected and control were not available in the dataset, so a default value of 1mM was selected. Extracellular lactate concentration increased drastically at 0 hours and 65 hours (2 DPI) following MERS-CoV infection, with little change outside of these timepoints (Figure 2d).

Overall, our simulation results suggest a direct correlation between glycolytic activity and disease severity, with SARS-CoV-2 and MERS-CoV showing the largest differences between infected and control, particularly a sharp increase in the amount of extracellular lactate being produced due to upstream glycolytic activity.

B. Pentose Phosphate Pathway Activity Is Associated with Increased Severity of Infection

For the analysis of the PPP, we focused on NADPH produced from G6PD, G1,5L6P, and SHL7P. NADPH produced from G6PD was selected for analysis because it is an important neutralizer of reactive oxygen species and a representative value for how many metabolites are entering the PPP. G1,5L6P and SHL7P, are the first and one of the last metabolites of the pathway, respectively, and were selected to assess metabolite flow through the PPP.

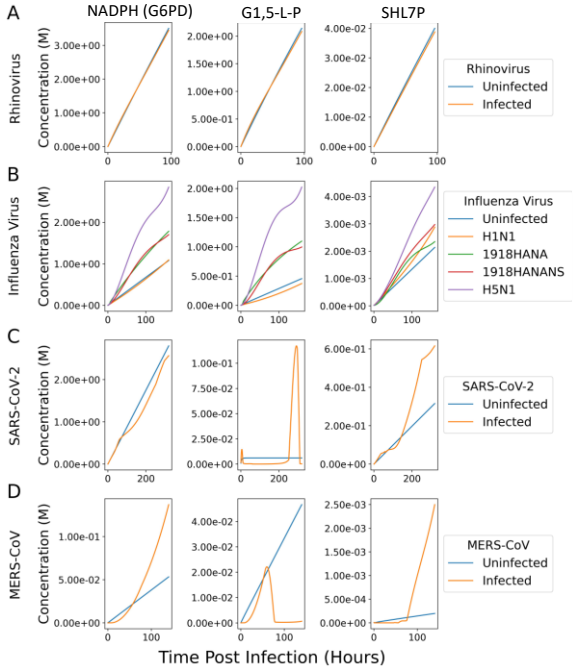

Fig. 3. Analysis of pentose phosphate pathway during infectious respiratory diseases. Time-course plots comparing the simulated concentrations of NADPH produced by G6PD, G1,5L6P, and SHL7P during infection by (a) Rhinovirus (b) Influenza virus strains (c) SARS-CoV-2, and (d) MERS-CoV. For each virus type analyzed, an uninfected control was included for reference of baseline metabolic level.

Rhinovirus infection did not alter the PPP based on the 3 metabolites analyzed in infected versus uninfected control groups (Figure 3a). During influenza virus infections, the production of NADPH was increased the most in H5N1. Both 1918 HANA and 1918 HANANS strains followed the same approximate pattern, with moderate elevation of NADPH. The H1N1 strain showed little difference in production of NADPH over the uninfected control. The concentration of G1,5L6P was highest in infection by the H5N1 strain, followed by infection by 1918 HANA and 1918 HANANS, with the H1N1 infection concentration being slightly below the uninfected control. SHL7P production was highest in the H5N1 strain, with all other strains being only slightly elevated above the control simulation (Figure 3b). During SARS-CoV-2 infection, the production of NADPH via G6PD was slightly decreased from approximately 100 hours to 150 hours (4 and 6 DPI). The concentration of G1,5L6P experienced a burst-like spike in concentration near the end of the simulation. The concentration of SH7LP during SARS-CoV-2 infection increased drastically at 150 hours (6 DPI) (Figure 3c). In MERS-CoV infection, the production of NADPH started off slower than the control, but it increased on the second half of the simulation period. G1,5L6P concentration remained similar in infected and in the control until approximately 55 hours (2 DPI), when it was quickly depleted, remaining low for the remainder of the simulation. SH7LP concentration was lower in the infected simulation until hour 75 (3 DPI) when it began to increase at a high constant rate (Figure 3d).

Simulations of the PPP showed mostly a strain severity-dependent response within the influenza infections, with more severe infections (i.e., H5N1) having a greater increase in the concentrations of metabolites of the PPP. SARS-CoV-2 infection showed decreased NADPH production, and MERS-CoV infection had a drastic increase in NADPH production at 60 hours post infection (2 DPI). Overall, PPP dysregulation increased with disease severity, and may be an indicator of disease severity in emergent disease.

C. Fatty Acid Oxidation Increased in Influenza Depending on Strain Lethality and Decreased in SARS-CoV-2

Fatty acid oxidation leads to the production of acetyl-coA and is an important driver of the TCA cycle. We focused on two main metabolites within fatty acid oxidation: fatty acids in the cytosol, and acetoacetyl-coA. Like the PPP, we highlighted metabolites at the start (Fatty Acid) and finish (Acetoacetyl-CoA) of fatty acid oxidation to obtain a snapshot of how the pathway was functioning. Acetoacetyl-CoA showed how many metabolites were directly available to be converted into acetyl-coA to continue fueling mitochondrial metabolism.

During rhinovirus infection, fatty acid concentration within the cytosol spiked slightly at the beginning of the simulation but quickly returned to background levels like the uninfected control. The concentration of acetoacetyl-coA mostly mirrored this behavior, with a spike in the early portion of the simulation, and evening out with the uninfected control by 75 hours (3 DPI) (Figure 4a). Infection with influenza virus mostly led to decreased concentration of fatty acids in the cytosol, except for the H5N1 strain, which showed a biphasic response with a suppression of fatty acids followed by increased concentration above uninfected and the rest of strains analyzed. The 1918

HANANS strain of Influenza also caused a biphasic response, with increased in fatty acid concentration though of lesser magnitude than the response to H5N1 infection. This effect seems to correlate negatively with the amount of acetoacetyl-coA produced, where the concentration of acetoacetyl-coA was increased during infection by most strains except for H5N1 and 1918 HANANS (Figure 4b). During SARS-CoV-2 infection, the concentration of fatty acids in the cytosol had a sharp increase initially, returning to control levels by 200 hours (8 DPI). The acetoacetyl-coA concentration was initially low, but after approximately 150 hours (6 DPI), increased to a higher concentration than the control (Figure 4c). In MERS-CoV, fatty acid concentration in the cytosol was increased much higher than that of the uninfected control, and the concentration of acetoacetyl-coA experienced a burst of concentration between 90 hours and 130 hours (3 and 5 DPI) (Figure 4d).

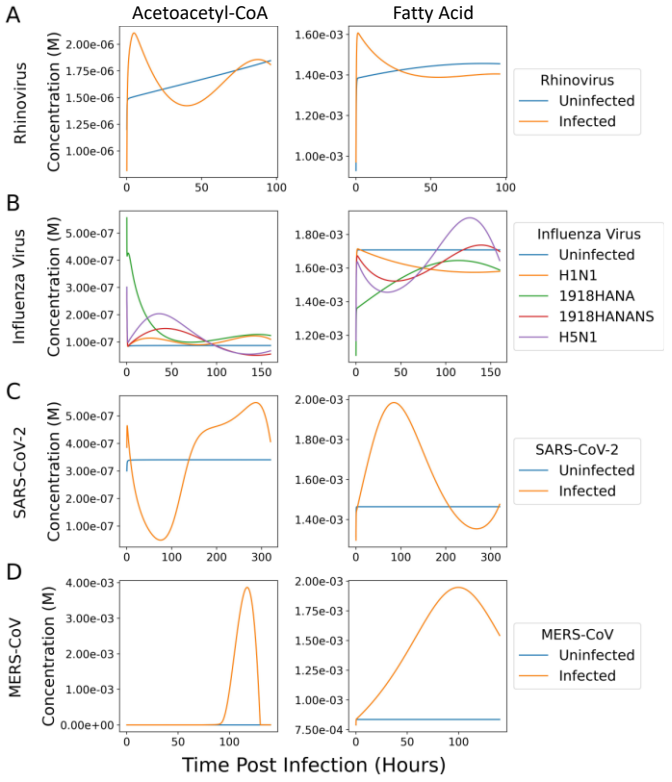

Fig. 4. Analysis of fatty acid oxidation during infectious respiratory diseases. Time-course plots comparing the simulated concentrations of acetoacetyl-coA and fatty acid in the cytosol during infection by (a) Rhinovirus (b) Influenza virus strains (c) SARS-CoV-2, and (d) MERS-CoV. For each virus type analyzed, an uninfected control was included for reference of baseline metabolic level.

Our simulation results point to differential responses depending on the virus type with minimum impact during rhinovirus infection, and infection with less pathogenic strains of influenza virus. In contrast, results from influenza virus H5N1 and 1918 HANANS suggest that higher fatty acid oxidation could be associated with increased morbidity and mortality. With regards to the coronavirus strains, both SARS-CoV2 and MERS-CoV show changes in FA metabolism with the former resulting in a sharp increase of FA and Acetoacetyl-CoA concentrations towards the end of the simulations.

D. SARS-CoV-2 and MERS-CoV Infections are Characterized by a Decrease in Host Mitochondrial Activity Early Post-Infection, Followed by Increased Activity During Recovery

Oxidative phosphorylation in the mitochondria is an important source of energy that is fueled by the TCA cycle, in which metabolites from glycolysis, fatty acid oxidation, and glutaminolysis converge. When infections or other stimuli disrupt the flow of metabolites into the mitochondria or suppress the expression of mitochondrial enzymes, this source of energy can be impeded, which can lead to cell apoptosis. Our analyses highlighted acetyl-coA concentration, ATP produced from oxidative phosphorylation, and succinyl-coA as indicators of TCA and oxidative phosphorylation activities. Acetyl-CoA is crucial for the continued function of the TCA cycle and is produced by both glycolysis and fatty acid oxidation. ATP production from oxidative phosphorylation is an important indicator of mitochondrial function, as it is affected by the activity of the entire TCA cycle and the complex concentrations within the electron transport chain.

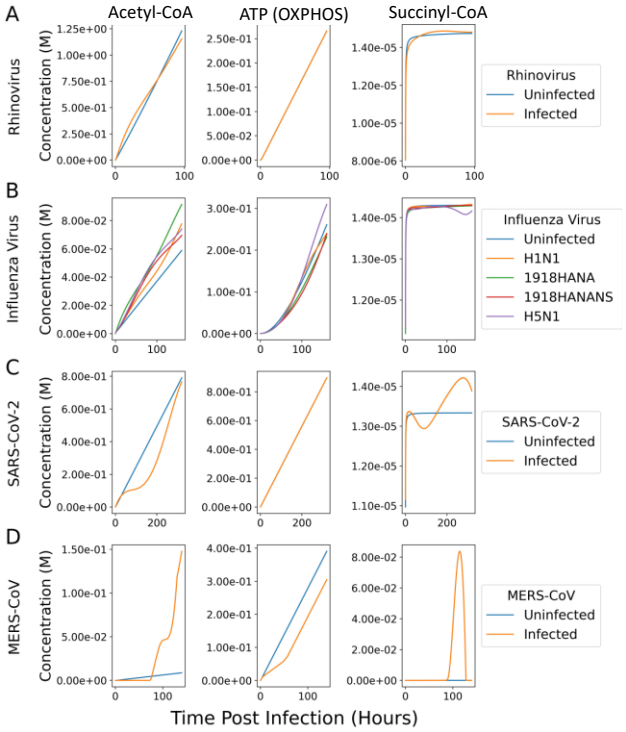

Fig. 5. Analysis of citrate cycle during infectious respiratory diseases. Time-course plots comparing the simulated concentrations of acetyl-coA, ATP produced from oxidative phosphorylation, and succinyl-coA during infection by (a) Rhinovirus (b) Influenza virus strains (c) SARS-CoV-2, and (d) MERS-CoV. For each virus type analyzed, an uninfected control was included for reference of baseline metabolic level.

Rhinovirus infection simulation shows little change from the uninfected control concentrations of acetyl-coA, ATP produced from oxidative phosphorylation, and succinyl-coA (Figure 5a). Influenza virus infection appears to have modest impact on the parameters included in the analysis. With regards to Acetyl-CoA, there was an increase over the control. ATP produced from oxidative phosphorylation was suppressed in infections by both 1918 HANA and 1918 HANANS strains and slightly increased in the H5N1 infection over the control. Succinyl-coA concentration was unchanged in all strains analyzed, except for

a slight suppression towards the end of the simulation seen for the H5N1. (Figure 5b). The most remarkable effects of SARS-CoV-2 infection were a decrease in Acetyl-CoA and an initial decrease followed by an increase in concentration over the control in succinyl-CoA. MERS-CoV showed a significant increase in Acetyl-CoA and a spike of succinyl-CoA overlapping in time. ATP was suppressed in comparison to the control by MERS-CoV with no detectable effects in SARS-CoV-2 (Figure 5c).

Mitochondrial metabolism in influenza virus infections shows slightly increased concentrations of acetyl-coA. Initially, all strains of Influenza have decreased production of ATP from oxidative phosphorylation, but through the infection, H5N1 recovers to a higher level than the control, whereas the other strains are not able to recover to the control amount.

E. Glutaminolysis Activity Increased in H5N1, SARS-CoV-2, and MERS-CoV Infections

We highlighted glutaminolysis and the malate aspartate shuttle to assess if metabolites are flowing into or out of the TCA cycle (Figure 6a). To do so, our analysis was focused on the concentration of aspartate, α-KG, and glutamate. The malate aspartate shuttle anti-ports multiple metabolites over the mitochondrial membrane, facilitating gluconeogenesis and transport of energy into the mitochondria. Glutaminolysis produces glutamate which is converted into α-KG and utilized in the TCA cycle.

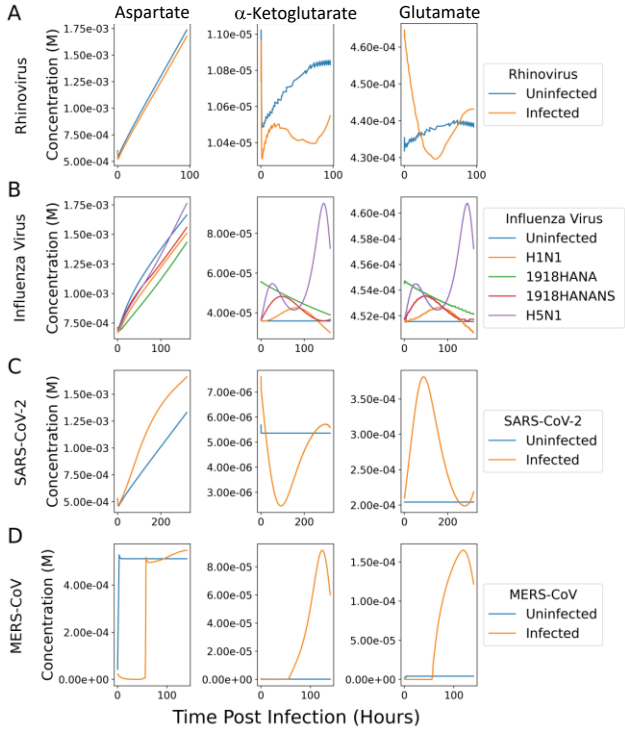

Fig. 6. Analysis of malate aspartate shuttle and glutaminolysis during infectious respiratory diseases. Time-course plots comparing the simulated concentrations of mitochondrial aspartate, mitochondrial a-ketoglutarate, and mitochondrial glutamate during infection by (a) Rhinovirus (b) Influenza virus strains (c) SARS-CoV-2, and (d) MERS-CoV. For each virus type analyzed, an uninfected control was included for reference of baseline metabolic level.

During rhinovirus infection, aspartate concentration in the mitochondria did not change compared to the uninfected control. α-KG and glutamate concentration both remained like the uninfected control (Figure 6b). Aspartate concentrations during influenza virus infection, were mostly lower than in the control except for H5N1 in which levels recovered towards the end. All infections by strains of Influenza virus except H5N1 caused α-KG and glutamate concentrations to be elevated initially and then return to similar levels to the control. H5N1 infection diverged from this pattern and increased these metabolites in the second half of the simulation (Figure 6c). SARS-CoV-2 infection shows increased aspartate concentration in the mitochondria, diminished levels of α-KG, and a large increase in glutamate concentration compared to the control (Figure 6d). During infection with MERS-CoV, aspartate stayed low before a burst of metabolites increased the concentration very quickly to match that of the control. a-KG increased at 50 hours (2 DPI), peaking just before the end of the simulation at 125 hours (5 DPI). Glutamate concentration followed a similar trend, with an increase over the uninfected control starting 50 hours (2 DPI) into the simulation and peaking also at 125 hours (5 DPI) (Figure 6e).

The differences between the rhinovirus infection and control are slightly higher in glutaminolysis than in any other pathway analyzed in the study. In general, glutaminolysis was slightly decreased in the infected simulation, reducing the number of metabolites within the TCA cycle, and very slightly decreasing the number of metabolites moved into the malate-aspartate shuttle and gluconeogenesis. In influenza, glutaminolysis was initially increased, pushing more metabolites into the TCA cycle, and most strains except for the H5N1, which is the most pathogenic of all the strains included in the analysis, decreased the activity of glutaminolysis as the infections cleared. H5N1 had increased amounts of α-KG brought into the TCA, and it leads to increased gluconeogenesis via the malate aspartate shuttle. SARS-CoV-2 infection showed increased gluconeogenesis, which decreased all metabolites within the TCA cycle, including α-KG. Increased glutaminolysis caused the recovery of these concentrations and brings α-KG back to the control concentration. During MERS-CoV infection, TCA cycle was supplied with many metabolites through glutaminolysis after approximately 50 hours post infection (2 DPI). This had no effect on ATP produced from oxidative phosphorylation, and is likely associated with the pathogenesis according to Bharadwaj et al.[8].

F. Disease Severity Correlates with Higher Perturbation of Metabolism

Using our M^2 pipeline we have been able to compare metabolic responses in infected versus an uninfected tissue specimen. By normalizing each final concentration value for the species within the model, we can compare various infections based on fold change in the infected with respect to the uninfected control (Figure 7). After performing this analysis, the data show that infection with rhinovirus was easily identified as having very slight changes from its control for all metabolic pathways included in the model. With regards to flu, there was similar behavior within strains for all the pathways analyzed, although the highly pathogenic H5N1 strain stood out as having a slightly more perturbed metabolic profile with the PPP and glycolysis being differentially upregulated in comparison to the rest of influenza virus strains. SARS-CoV-2 infection showed a

similar magnitude of perturbation as the H5N1, although the PPP and some parts of glycolysis were in this case suppressed. Additionally, species associated with the TCA cycle were largely decreased, and gluconeogenesis species were increased. In MERS-CoV infection, glycolysis species were upregulated, TCA cycle was highly dysregulated, and fatty acid oxidation was upregulated. Overall, the severity of the infection correlated strongly with the level of fold change from the controls. MERS-CoV shows indications that the TCA cycle was broken, with some portions being highly increased and others suppressed.

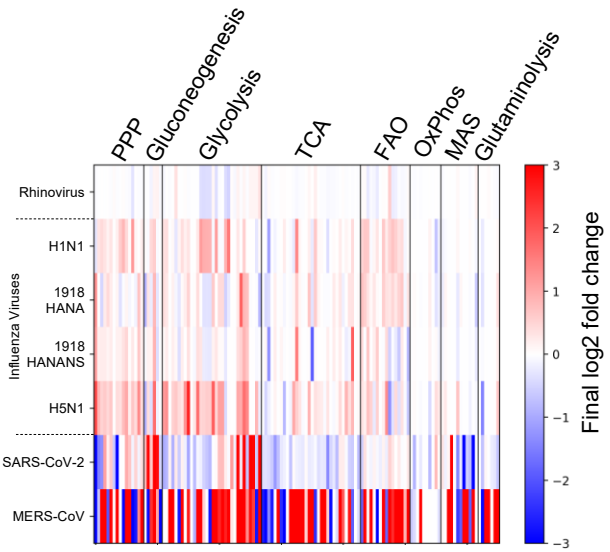

Fig. 7. Summary heatmap showing final log2 fold change of various infections to their respective internal uninfected controls. Sections of the heatmap were labeled according to the portion of metabolism to which each species belonged with red representing upregulated and blue downregulated metabolic responses.

IV. Discussion

Metabolism is a massively and dynamically interconnected network of chemical reactions facilitated by a collection of enzymes unique to each species. We created the M^2 pipeline representing large portions of metabolic pathways that are highly conserved in most eukaryotic organisms while accounting for multiple possible isoforms in our workflow. The M^2 pipeline integrates theoretical and data-driven approaches to create a metabolic profile that allows for host metabolic response to various perturbations to be identified quickly after the experimental data (i.e., RNA-seq) is obtained. With emergent diseases being introduced to the human population, tools for a rapid analysis of the metabolic alterations during infection could be used to predict the metabolic fingerprint of a virus, which could provide initial insights into pathogenesis and severity of that particular biological threat. A tool such as our M^2 pipeline could increase our level of preparedness to rapidly respond to critical events like the sudden emergence of a previously unknown viral disease. The modularity of the M^2 pipeline's design allows for any gene expression dataset to be analyzed, although the quality of data is critical to allow for a more accurate predictive power in the analysis.

The model of metabolism that we have created includes glycolysis, gluconeogenesis, the pentose phosphate pathway, fatty acid oxidation, the TCA cycle, oxidative phosphorylation, glutaminolysis, and the malate aspartate shuttle. This differential model encompasses many different pathways of metabolism that have not previously been combined in a mathematical model of metabolism. The wide range of pathways within the model allows for a detailed analysis for any gene expression input. The reactions where larger pathways converge, like the TCA cycle. offer ample opportunity to characterize the flow of metabolites through multiple pathways in a wide variety of contexts (i.e., infectious disease, drug discovery, ionizing radiation, or cancer).

Glycolysis is associated with lactate production and supplies pyruvate for mitochondrial metabolism. In immunology, increased rates in glycolysis and lactate production are associated with the pro-inflammatory response and differentiation of M1 macrophages, Th1, and Th17 cell types[22]. Glycolysis is also often hijacked by viruses to produce building blocks for additional viral particles and is associated with increased viral activity[6, 23]. In our metabolic analysis of infectious respiratory diseases, we found glycolysis to be upregulated in a disease-severity dependent manner. SARS-CoV-2 and MERS-CoV both exported much more lactate than their respective control groups. This increase in glycolysis is likely due to viral intervention, and the identification of mechanisms that limit glycolysis and other viral-producing pathways may prove to have beneficial responses in disease outcome. Codo et al[4] showed that glucose-rich environments favored SARS-CoV-2 replication and suggested a glycolysis-limiting therapeutic to suppress SARS-CoV-2 replication. Additionally, hyperglycemia and diabetes are predisposing factors for increased mortality from COVID-19[24, 25], providing further evidence that the metabolic environment on the host can have a significant impact on the outcome of the disease. Our modeling efforts can reproduce the increase in glycolytic flux seen experimentally and reported in the literature[4, 23, 26] using only gene expression data as input. From these analyses, the driving forces for glycolytic changes were determined to be mainly caused by upstream glycolytic activity and reduced pyruvate transport into the mitochondria.

The PPP is another glucose consuming pathway that is necessary for normal cell function, its production of NADPH is required for antioxidant activities and nucleotide production is downstream of R5-P. Unlike glycolysis, it is unclear if upregulation of the PPP is universally beneficial or detrimental in viral infections. Keshavarz et al. identified the PPP as being detrimental during influenza infections due to the increase in nucleotide production for the viral particles[26]. The influenza virus infection simulations caused a much greater change in the activity of the PPP compared to rhinovirus. H5N1 is a highly pathogenic influenza virus strain, followed by 1918 HANA, 1918 HANANS, and H1N1 being the least lethal[27-29]. The level of dysregulation in the PPP found in our analysis correlates well with the expected morbidity and lethality of the strain, suggesting that more virulent strains cause increased PPP activity. The TCA cycle, oxidative phosphorylation, fatty acid oxidation, glutaminolysis, and the malate aspartate shuttle all effect the metabolic profile within the mitochondria. Glycolysis and fatty acid oxidation feed the TCA cycle and the number of metabolites within the cycle are balanced by glutaminolysis and the malate aspartate shuttle. We observed fatty acid oxidation increasing initially in influenza infection, which has been linked

to the induction of innate anti-viral responses[30]. Our results show significant differences in fatty acid oxidation, but further investigation would be necessary to determine if the increased activity correlates with disease-or-strain-mediated morbidity.

The increased oxidative stress seen in SARS-CoV-2 infections may be playing a role in the decreased fatty acid oxidation we observe here. During the recovery phase of the SARS-CoV-2 infection, fatty acid oxidation increases beyond that of the uninfected, bringing acetyl-coA level nearly back to that of the control. The gene expression data associated with oxidative phosphorylation was missing from the dataset obtained for the SARS-CoV-2 study, so changes to ATP production due to gene expression changes in the electron transport chain are unable to be observed. MERS-CoV infection shows a strong decrease in oxidative phosphorylation during the initial portion of the infection, with the rate of production returning to normal levels afterwards. This decrease in ATP production has been observed experimentally: Pan et al. describe how the nsp1 protein inhibits oxidative phosphorylation in host cells during MERS-CoV infections[31]. Additionally, the MERS-CoV infection had a burst-like replenishment of the TCA cycle via glutaminolysis after the peak of the infection, which would suggest that glutamine metabolism may be important in the recovery of infection by MERS-CoV, whereas in SARS-CoV-2 infections, glutamine metabolism is increased drastically during the initial phases. Bharadwaj et al has demonstrated the increase in glutamine metabolism as contributing to the pathogenesis of SARS-CoV-2, and being directly activated by SARS-CoV-2 during infection[8] which strongly matches the results that we obtained from our pipeline. The M^2 pipeline produces the most accurate results with high quality RNA-seq datasets that have a high coverage, but lower quality gene expression datasets or datasets missing genes still provide insight into the metabolic response to disease, allowing for greatly increased knowledge extraction from costly biological experiments.

V. Conclusion

The analyses presented here were the result of our advanced M^2 bioinformatics pipeline. From the outputs of the model, both the similarities that emerged between the infections as well as the differences represent useful comparative insights that can help develop novel treatment modalities against these viral infections. Many biologically relevant predictions on the metabolic response have been experimentally validated as found in published literature, showing how the data-driven calibration of the model accurately represents the biological pathways and provides a valuable tool for the fast and precise analysis of global metabolic networks. Tools like the one presented here are important due to the constant threat of emerging or re-emerging infections becoming pandemic events. The short time required between sequencing data acquisition to host metabolic modeling outputs improves the readiness against potential public health threats. The M^2 pipeline can quickly take time course host gene expression datasets as input and return detailed predictions of metabolic profiles as outputs that can serve as initial estimates of severity of disease as well as provide insights into potential metabolic-related mechanisms of pathogenesis and potential therapeutic interventions.

References

[1] Baker, R.H.R.T.-J.N.L.A.J.K.S.B.-R.J., *Computational modeling of complex bioenergetic mechanisms that modulate CD4+ T cell effector and regulatory functions.* npj Systems Biology and Applications, 2022.

[2] Thaker, S.K., J. Ch'ng, and H.R. Christofk, *Viral hijacking of cellular metabolism.* BMC Biol, 2019. 17(1): p. 59.

[3] Khomich, O.A., et al., *Redox Biology of Respiratory Viral Infections.* Viruses, 2018. 10(8).

[4] Codo, A.C., et al., *Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1alpha/Glycolysis-Dependent Axis.* Cell Metab, 2020. 32(3): p. 437-446 e5.

[5] Pucino, V., et al., *Lactate at the crossroads of metabolism, inflammation, and autoimmunity.* Eur J Immunol, 2017. 47(1): p. 14-21.

[6] Ren, L., et al., *Influenza A Virus (H1N1) Infection Induces Glycolysis to Facilitate Viral Replication.* Virol Sin, 2021. 36(6): p. 1532-1542.

[7] Smallwood, H.S., et al., *Targeting Metabolic Reprogramming by Influenza Infection for Therapeutic Intervention.* Cell Rep, 2017. 19(8): p. 1640-1653.

[8] Bharadwaj, S., et al., *SARS-CoV-2 and Glutamine: SARS-CoV-2 Triggered Pathogenesis via Metabolic Reprograming of Glutamine in Host Cells.* Front Mol Biosci, 2020. 7: p. 627842.

[9] Gibellini, L., et al., *Altered bioenergetics and mitochondrial dysfunction of monocytes in patients with COVID-19 pneumonia.* EMBO Mol Med, 2020. 12(12): p. e13001.

[10] Saleh, J., et al., *Mitochondria and microbiota dysfunction in COVID-19 pathogenesis.* Mitochondrion, 2020. 54: p. 1-7.

[11] Lu, W., et al., *Metabolite Measurement: Pitfalls to Avoid and Practices to Follow.* Annu Rev Biochem, 2017. 86: p. 277-304.

[12] Michaelis, L., et al., *The original Michaelis constant: translation of the 1913 Michaelis-Menten paper.* Biochemistry, 2011. 50(39): p. 8264-9.

[13] Goutelle, S., et al., *The Hill equation: a review of its capabilities in pharmacological modelling.* Fundam Clin Pharmacol, 2008. 22(6): p. 633-48.

[14] Heiske, M., T. Letellier, and E. Klipp, *Comprehensive mathematical model of oxidative phosphorylation valid for physiological and pathological conditions.* FEBS J, 2017. 284(17): p. 2802-2828.

[15] Konig, M., S. Bulik, and H.G. Holzhutter, *Quantifying the contribution of the liver to glucose homeostasis: a detailed kinetic model of human hepatic glucose metabolism.* PLoS Comput Biol, 2012. 8(6): p. e1002577.

[16] Morens, D.M. and A.S. Fauci, *Emerging Pandemic Diseases: How We Got to COVID-19.* Cell, 2020. 182(5): p. 1077-1092.

[17] Vigant, F., N.C. Santos, and B. Lee, *Broad-spectrum antivirals against viral fusion.* Nat Rev Microbiol, 2015. 13(7): p. 426-37.

[18] Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets--update.* Nucleic Acids Res, 2013. 41(Database issue): p. D991-5.

[19] Wittig, U., et al., *SABIO-RK--database for biochemical reaction kinetics.* Nucleic Acids Res, 2012. 40(Database issue): p. D790-6.

[20] Daly, M.E., et al., *Acute effects on insulin sensitivity and diurnal metabolic profiles of a high-sucrose compared with a high-starch diet.* Am J Clin Nutr, 1998. 67(6): p. 1186-96.

[21] Cruzat, V., et al., *Glutamine: Metabolism and Immune Function, Supplementation and Clinical Translation.* Nutrients, 2018. 10(11).

[22] Soto-Heredero, G., et al., *Glycolysis - a key player in the inflammatory response.* FEBS J, 2020. 287(16): p. 3350-3369.

[23] Icard, P., et al., *The key role of Warburg effect in SARS-CoV-2 replication and associated inflammatory response.* Biochimie, 2021. 180: p. 169-177.

[24] Khunti, K., et al., *COVID-19, Hyperglycemia, and New-Onset Diabetes.* Diabetes Care, 2021. 44(12): p. 2645-2655.

[25] Wang, J. and W. Meng, *COVID-19 and diabetes: the contributions of hyperglycemia.* J Mol Cell Biol, 2020. 12(12): p. 958-962.

[26] Keshavarz, M., et al., *Metabolic host response and therapeutic approaches to influenza infection.* Cell Mol Biol Lett, 2020. 25: p. 15.

[27] Dawood, F.S., et al., *Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study.* Lancet Infect Dis, 2012. 12(9): p. 687-95.

[28] Johnson, N.P. and J. Mueller, *Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic.* Bull Hist Med, 2002. 76(1): p. 105-15.

[29] Organization, W.H. *Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2022*. 2022 [cited 2022 July 14, 2022]; Available from: https://cdn.who.int/media/docs/default-source/influenza/human-animal-interface-risk-assessments/2022_june_tableh5n1.pdf?sfvrsn=172efb9e_1&download=true.

[30] van Liempd, S., et al., *Impaired beta-oxidation increases vulnerability to influenza A infection.* J Biol Chem, 2021. 297(5): p. 101298.

[31] Pan, Z., et al., *MERS-CoV nsp1 impairs the cellular metabolic processes by selectively downregulating mRNAs in a novel granules.* Virulence, 2022. 13(1): p. 355-369.