Can Metabolic Flux Analysis provide a clearer indication of a person's well-being compared with gene expression-only analysis?

Jonathan Webb

1 School of Computing, Teesside University, Middlesbrough, North Yorkshire, United Kingdom

Abstract

Influenza viruses infect about 20% of children and 5% of adults worldwide each 18 year and is a globally important contagion. Influenza viruses take advantage of the host cell metabolism to replicate their genetic material and to synthesize viral proteins. Cellular response to influenza strains can vary largely depending on the person and the strain itself, therefore metabolism is an important field of study to elucidate the mechanism of action and the cellular outcome of individuals affected by influenza strains.

Flux balance analysis, FBA, is a mathematical approach for analysing the flow of metabolites through a metabolic network. By applying FBA to human bronchial epithelial cells infected with influenza viruses (H7N9 [A/Anhui/01/2013], H7N7 [A/Netherland/219/2003], H5N1 [A/Vietnam/1203/2004], and a human seasonal virus H3N2 [A/Panama/2007/1999]), we believe our results provide a clearer indication of the effects of the virus on the cells than gene expression-only analysis.

The metabolic rates of the infected cells show a clear degradation of cell biomass with an increase in virus potency from all influenza strands. The Strong negative correlation for 'Squalene and Cholesterol' synthesis for the H7N7 strain could be directly linked with a H7N7 vaccine suffering from poor immunogenicity. An area of particular interest shows a clear effect of viral infection on fatty acid synthesis (FAS). Results between H5N1 and H7N9 infected cells in FAS reactions show strong similarities.

The results from this study have shown clear metabolic effects of influenza viruses on human bronchial epithelial cells. Differences and similarities between virus strands have been outlined that gene expression-only analysis would not have been able to identify. From this it can be said metabolic flux analysis does provide a clearer indication of a person's well-being by defining specifically effected metabolic reactions.

Keywords: Influenza, Metabolic Flux Analysis, Flux Balance Analysis, Constraint Based Reconstruction Analysis, Bioinformatics

Introduction

One way to measure a person's well-being is by analysing their cells. Cells are described as the basic unit of all (biological) life [1]. There are two well-documented methods of measuring a cell. One method is by measuring the gene expression of a person's cells. This can help identify viral infection of a cell via viral protein expression, susceptibility to cancer via oncogene expression, and bacterial resistance to penicillin via

beta-lactamase expression. In silico gene expression analysis is common practice in cell analysis. [2] Flux balance analysis (FBA) can be used to measure the production and consumption rates of metabolites in a metabolic system. Although researchers usually focus on genes it could be argued that the metabolic rates of a cell are a clearer indication of the state of a person's cells and in turn the person themselves due to metabolism being closer to the observable phenome.

Metabolic fluxes provide measure of degree of engagement of the various pathways in the overall cellular functions and metabolic processes at specific genetic and environmental conditions, constituting a fundamental determinant of cell physiology [3]. Cell metabolism is dramatically altered by virus infection [31]. For this reason, human cells measured during viral infection will generate more interesting results.

Influenza viruses infect about 20% of children and 5% of adults worldwide each year [4] and is a globally important contagion. Many strands of influenza exist and evolve each year maintaining the high rates of yearly infection with the H5N1 strand also known as bird flu causing a global pandemic with 630 human cases confirmed resulting in the deaths of 375 people between 2003 and June 2013 [5]. As a virus, influenza acts quickly with the initial site of replication thought to be tracheobronchial ciliated epithelium detected in secretions usually within 24h [6]. Also, influenza affects a human host almost entirely with symptoms varying through respiratory syndromes, disorders affecting the lung, heart, brain, liver, kidneys, and muscles, to fulminant primary viral and secondary bacterial pneumonia [6]. Influenza viruses take advantage of the host cell metabolism to replicate their genetic material and to synthesize viral proteins [30]. The importance of the effects of influenza with the viruses fast acting properties and worldwide infection rates make influenza infected cells desirable to model.

In the study outlined by Josset et al [8], gene expression analysis is used to identify host response differences between the H7N9 strand avian influenza virus to H5N1, H7N7, and H3N2 avian and human influenza strands via transcriptomic characterisation. In Josset et al's study [8], a virus's potency is measured in PFU/ml (plaque forming units of virus used for infection divided by the number of cells per millilitre) as a method of quantifying the virus in volume in bronchial cells. Transcriptomic analysis would show how well the cells are functioning and how each virus is effecting the cells differently therefore, such modelling would generate a comprehensive and poly-omic insight into the response to the virus at a metabolic level.

In this study, we model the metabolic rates of cellular fluxes of human bronchial epithelial (HBE) cells outlined in the study by Josset et al [8]. Then, we compare and correlate them with the levels of PFU/ml. We find that, in most measurements, PFU/ml strongly correlates with the metabolic flux rates for certain subsystems. However, differences between subsystem specific cell metabolism, PFU/ml, and virus show metabolic flux analysis (MFA) identifying changes in a person's well-being greater than gene expression only analysis. Furthermore, our results show greater changes in flux metabolism between virus strands and HPI than shown in gene expression only analysis. Figure 1 outlines the process we aim to achieve in this paper.

Taken together, we demonstrate that metabolic modelling can provide mechanistic understanding of the different metabolic effects on a person based on the effect of each virus on their cells and determine in which way each virus can effect an individuals well-being.

To our knowledge, this is the first attempt at modelling effects of influenza innoculation on human bronchial epithelial cells at genome scale. Josset et al. [9] specifically states flux balance analysis is yet to be applied to model H5N1 pathogenesis.

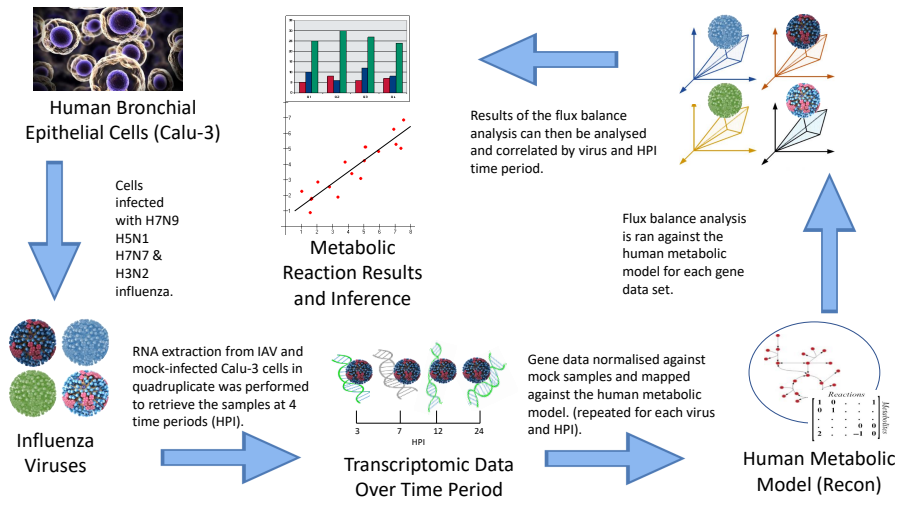

Figure 1. Sequential Diagram. This diagram shows the infection of human bronchial epithelial cells with viruses H7N9, H5N1, H7N7 and H3N2. Quadruplicate RNA extraction is then performed at 4 different hours post inoculation (HPI) times for each virus and the non-infected mock sample to gather the transcriptomic data. Each data set is mapped to the human metabolic model and flux balance analysis (FBA) ran against it. The results are then analysed and correlated to the PFU/ml results outlined by Josset et al. [8].

Methods

Model and Data

The idea behind the paper is to map gene expression data of influenza-affected cells onto a genome-scale metabolic model of human metabolism.

The gene expression data used by Josset et al [8] is to be used to perform MFA in MATLAB. Before being able to do this, the data was imported individually by HPI and virus strain from GEO (Gene Expression Omnibus) [29] under accession number GSE49840using the 'GEOImporter' module within GenePattern [11]. Three avian-origin IAVs isolated from fatal human cases, A/Anhui/01/2013 (H7N9) (Anhui01), A/Netherland/219/2003 (H7N7) (NL219), and A/Vietnam/1203/2004 (H5N1) (VN1203), and a seasonal human virus, A/Panama/2007/1999 (H3N2) (Pan99), were used in this study. The human bronchial epithelial cell line Calu-3 was used to grow the cells. Cells were infected with viruses with samples taken at 3, 7, 12 and 24 hours post inoculation (HPI). RNA extraction from IAV- and mock-infected Calu-3 cells in quadruplicate was performed to retrieve the samples. Non-infected cells were also measured using this technique at the same HPI times as mock values.

To use the data computationally with COBRA toolbox [12] and the human metabolic model Recon2.2 [13], only the 1765 genes referenced from the model would be required in the form of numerical data averaged and normalised against the mock samples. Gene conversion was performed through a custom script based on Biomart [14] to return the indices of the model specific genes from the dataset. Mock samples were then identified and used against the 'real' values programmatically to return a normalised data structure containing datasets that could be used to perform FBA using the human metabolic model.

Flux Balance Analysis

Flux-balance analysis (FBA) is a mathematical approach based on linear algebra that enables the analysis of the flow of metabolites through a chemical reaction network [15]. FBA can be applied using constraint based reconstruction analysis (COBRA) [17] via a MATLAB toolbox. According to FBA, each chemical reaction in a cell is associated with a flux in the model. Angione and Lio [16] defines the logic of metabolic constraint modelling as, when given m metabolites X_i, $i = 1, \ldots, m$ and n reactions with flux rates v_j, $j = 1, \ldots, n$, the constraints given are the lower and upper bound of the flux ranges (*capacity constraints*) and also by the balance of the concentration of the metabolites (*flux-balance constraints*). The balance that metabolites X_i must satisfy is

$$\frac{dX_i}{dt} = \sum_{j=1}^{n} S_{ij} v_j, \quad i = 1, \ldots, m \tag{1}$$

where S_{ij} is the stoichiometric coefficient of X_i in the jth reaction. Under steady state conditions $\frac{dX_i}{dt} = 0,\ \forall i$, the balance for the i$th$ metabolite is $\sum_{j=1}^{n} S_{ij} v_j = 0$. Therefore, at steady state, the balance equation is $Sv = 0$, where S is the stoichiometric matrix (m rows and n columns), and v is the vector of the fluxes (metabolic and transport fluxes) [16].

Each reaction is dependant on a single gene set, represented by a string of genes linked by AND/OR operators. For example, with two genes composing a gene set in an 'AND' relation, both are required to catalyze the corresponding reaction, and knocking out only one gene is sufficient to knock out the reaction. In this case, the gene set represents an *enzymatic complex*. Conversely, when two genes compose a gene set in an 'OR' relation, the two genes synthesize for *isozymes*, which differ in the amino acid sequence, but catalyze the same reaction. Therefore, one gene is sufficient to catalyze the reaction, and all genes must be knocked out in order to knock out the reaction.

With the aim of overcoming the limitations offered by the Boolean knockout approach, we need to formalize the AND/OR relation between genes using a real function that makes it possible to define the *"gene sets expression"* as function of the gene expression. Let $x_i^j,\ i = 1, \ldots, p$, be the gene expression levels of the genes $s_i^j,\ i = 1, \ldots, p$, and let $\wedge_i s_i^{(1)}$ and $\vee_i s_i^{(2)}$ be two basic gene sets modeling an enzymatic complex and an isozyme respectively. We adopt the following map τ between a gene set and its expression:

$$\bigwedge_{i=1,\ldots,p} s_i^{(1)} \overset{\tau}{\longmapsto} \min_{i=1,\ldots,p} \{x_i^{(1)}\}, \tag{2}$$

$$\bigvee_{i=1,\ldots,p} s_i^{(2)} \overset{\tau}{\longmapsto} \max_{i=1,\ldots,p} \{x_i^{(2)}\}. \tag{3}$$

Specifically, the expression level of a gene set that needs all its genes to work correctly (AND relation), is constrained to be equal to the lowest of the expression levels of its genes. Conversely, the expression level of a gene set that needs at least one of its genes (OR relation), is the highest of the expression levels of its genes. The bounds of a reaction catalyzed by an enzyme complex will be the function of the minimum expression level of its genes, while the bounds of a reaction catalyzed by isozymes will be the function of the maximum expression level of its genes. Nested gene sets are treated with the same methodology, i.e. applying (2) and (3) recursively.

The main difference in the FBA model is how the bounds of the fluxes depend on

the gene expression, Angione and Lio. [16] states the following bilevel linear program:

$$\begin{array}{lll} \max & g^\intercal v & \\ \text{such that} & \max & f^\intercal v \\ & \text{such that} & Sv = 0 \\ & & v_i \geq V_i^{min} h(y_i) \\ & & v_i \leq V_i^{max} h(y_i) \end{array} \tag{4}$$

where f and g are n-dimensional arrays of weights associated with the first and second objectives respectively, and indicate how much the reaction fluxes contribute to the objective function. In the present study, f and g are Boolean vectors. For instance, $g_j = 1$ if and only if the flux v_j has to be maximized as second objective. In order to define the function h, let y_i be the gene set expression of the ith gene set, responsible for the ith reaction of the model. To map the gene set expression value into a specific condition of the model, we use the following piecewise multiplicative function:

$$h(y_i) = \begin{cases} (1 + |log(y_i)|)^{\text{sgn}(y_i - 1)} & \text{if } y_i \in \mathbb{R}^+ \setminus \{1\} \\ 1 & \text{if } y_i = 1 \end{cases} \tag{5}$$

where $\text{sgn}(y_i - 1) = (y_i - 1)/\left|y_i - 1\right|$.

However, in this adaption:

$$h = 1 \tag{6}$$

Simplifying the program to:

$$\begin{array}{lll} \max & g^\intercal v & \\ \text{such that} & \max & f^\intercal v \\ & \text{such that} & Sv = 0 \\ & & v_i \geq V_i^{min} y_i \\ & & v_i \leq V_i^{max} y_i \end{array} \tag{7}$$

In the FBA model, we change the minimum and maximum flux of the ith reaction accordingly: $v_i^{min} = V_i^{min} y_i$, $v_i^{max} = V_i^{max} y_i$, where V_i^{min} and V_i^{max} are the minimum and maximum flux of the wild-type configuration of the model. This choice is consistent with the fact that all the gene expression values in various conditions are relative to those of the wild-type bacterium.

y_i is defined as the reaction expression. The reaction expression is the minimum flux of gene expressions with an 'AND' relation and the maximum flux of gene expressions with an 'OR' relation. These rules are applied recursively for nested gene sets.

The outer maximization problem in (4) is subject to the inner one. More specifically, the inner maximization finds the distributions of flux in the network such that the growth rate is maximized. In the outer maximization, all the unregulated fluxes are then distributed such that the second objective is maximized. The lower and upper bounds of the ith flux v_i depend on the expression level of the genes involved in the ith reaction. The bilevel problem is finally converted to a single-level problem [7], and solved using the GLPK solver. It is simple to check that all the approaches based on Boolean gene knockouts become a particular case of METRADE, being $h(y_i) \overset{y_i \to 0}{\longrightarrow} 0$ and $h(1) = 1$.

Correlation and Comparisons

As the data outlined in the study by Josset et al. [8] can not directly be compared with the cell metabolic flux rates from the aforementioned FBA methods, correlations would be an appropriate comparison method.

Correlation is a method of assessing a possible two-way linear association between two continuous variables. [18] Correlation is measured by the correlation coefficient, which represents the strength of the linear association between the variables in question. It takes a value in the range 1 to +1. A correlation coefficient of zero indicates that no linear relationship exists between two continuous variables, and a correlation coefficient of 1 or +1 indicates a perfect linear relationship. The strength of relationship can be anywhere between 1 and +1. The stronger the correlation, the closer the correlation coefficient comes to ±1. [19]

A positive correlation, closer towards +1, indicates the increase of two continuous variables at the same rate as the other. A negative correlation, close to -1, indicates a decrease in one variable at the same rate as the increase in the other. Both negative and positive correlations will show interesting results and the correlation strength will be ranked by the absolute value of the correlation.

The correlation would be made for each set of flux results from virus across the 4 normalised HPI measurement averages against data of the same structure from Figure 2 in Josset et al's [8] study which determines the viruses potency at each time. Due to the nature of the data, which assumes an increase in potency and decrease in metabolism over time, Pearsons and Spearman correlation were both used to compare flux rates.

Results

We generated results based on the correlation between the metabolic flux at each time interval HPI and the PFU/ml of the virus. The correlation of this shows the relation between potency of the virus and each specific flux or metabolic pathway, defined as a set of reactions with the same cellular goal. Table 1 shows the correlation for the generic human biomass reaction for each virus. Biomass is a proxy of the healthy status of the cell. The strong negative correlation shows a definite degradation in cell biomass with an increase in the potency of the virus. This can be seen as supportive evidence to the accuracy of the results.

Table 1. Spearman and Pearson correlation results from generic human biomass reaction and PFU/ml over the 4 HPI time periods.

	Spearman				**Pearson**			
Virus	H7N9	H7N7	H5N1	H3N2	H7N9	H7N7	H5N1	H3N2
RHO	-0.8	-1	-0.8	-1	-0.714959106	-0.663533688	-0.675942177	-0.684516852
PVAL	-0.8	-1	-0.8	-1	0.285040894	0.336466312	0.324057823	0.315483148

Figure 2 shows the top seven metabolic pathways, on average, from the Spearman correlation for each metabolic pathway by virus. 'Squalene and cholesterol synthesis' shows a very strong negative correlation from the H7N9 strain. Squalene is described as a cholesterol precursor, it is a natural product with immunostimulatory properties [21] and is used in influenza vaccines to enhance the immunogenicity [22]. Inactivated influenza vaccines have been available since the 1940s [23] but with new influenza subtypes constantly evolving, clinical trials are still relevant. A recent trial for a vaccine for the H7N7 influenza subtype found the inactivated subunit influenza A (H7N7) vaccine was safe but poorly immunogenic in humans [24]. Although evidence directly verifying the strong negative correlation between Squalene and cholesterol synthesis and the potency of the H7N7 could not be found, the in silico finding could explain the immunogenicity of the H7N7 vaccine [24].

Where most subsystems in Figure 2 show a correlation in one or more virus strain with the remaining viruses considered as not correlated, cysteine metabolism shows a moderate negative correlation for the H3N2 strain but a moderate to low positive

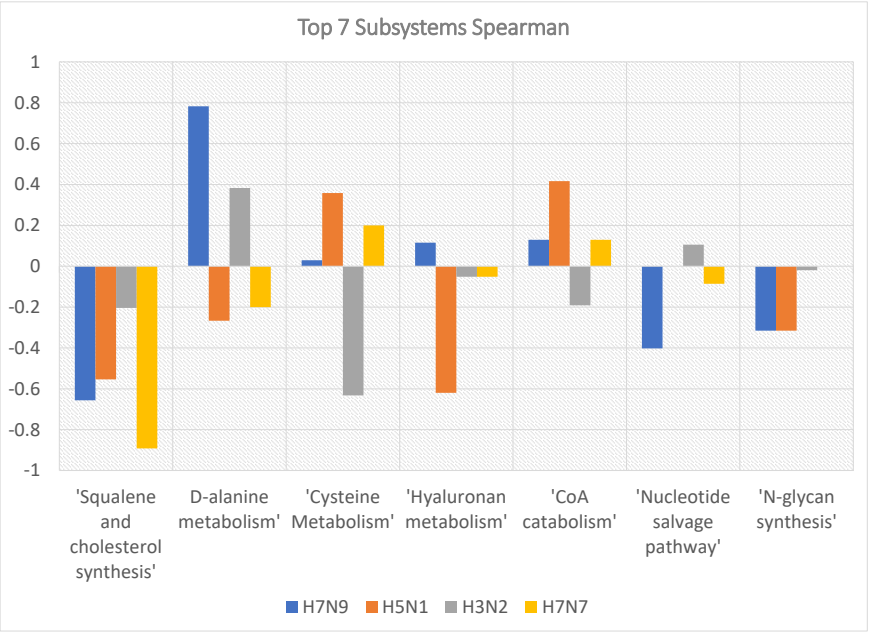

Figure 2. Top metabolic pathways correlation using Spearman. The sevem highest correlated metabolic pathways from all virus strains are shown alongside the correlations from the other virus strains. The variance between viruses shows differences in effects on cell metabolism.

correlation for the strain. Figure 3 confirms the correlation with a moderate positive and moderate negative correlation for H5N1 and H3N2 respectively. Human infections with highly pathogenic influenza A H5N1 viruses are associated with severe pneumonia, lymphopenia, high viral loads in the respiratory tract, and hyper-induction of cytokines and chemokines [25]. Chemokines contain mostly four cysteine residues [26] which explains the increase in cell cysteine metabolism. H3N2 infected cells have shown decreases in certain chemokine levels [27] confirming the negative correlation in cysteine metabolism.

A study by Janke et al. [30] into the effects of influenza d experimentally identified enzymes needed as a precursor for lipid biosynthesis. They concluded that fatty acid synthesis might play a crucial role for the replication of influenza viruses as they acquire lipid envelopes from their host cells. Figure 4 presents the very strong correlation of the lipid biomass chemical reaction with the potency of the virus. Although Janke et al's study [30] revolves around the H1N1 seasonal influenza A virus, our results show that the strains H7N9, H7N7, H5N1, and H3N2 also have an impact on the biomass of lipid.

Furthermore, Figure 4 outlines an important reaction containing the consumption of Acetyle-CoA and NADPH enzymes [30]. This reaction's main function is to catalyze the synthesis of palmitate from Acetyle-CoA and malonyl-CoA, in the presence of NADPH, into long-chain saturated fatty acids [32]. The results show that there is a decreased consumption of CoA with increasing PFU in strands H7N9 and H5N1, these are key precursors for fatty acid synthesis, which has been linked to influenza [30].

Figure 5 shows a consistency in correlations in fatty acid synthesis reaction between

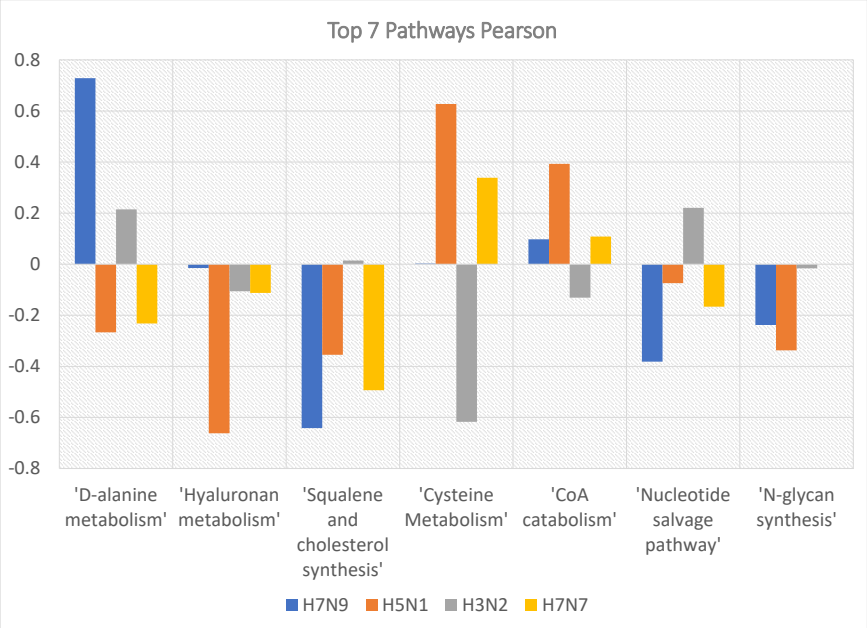

Figure 3. Top metabolic pathways correlation using Pearson correlation coefficient. The seven highest correlated metabolic pathways from all virus strains are shown alongside the correlations from the other virus strains. The variance between viruses shows differences in effects on cell metabolism.

H7N9 and H5N1. These results show similar previously undocumented effects on fatty acid synthesis metabolic reactions between these two strands while a general effect on fatty acid synthesis remains noticeable across all strands.

These results warrant potential further investigation into the effects on these enzymes by the documented influenza strains and the reasons for the differences and similarities between viruses and any further connection with the H1N1 strand.

Conclusion

This study set out to assess the level of indication into a persons well-being through Metabolic Flux Analysis compared with gene-expression only analysis. Cells infected with a virus would show a greater variance in cell metabolism. Each virus species also likely requires unique metabolic changes for successful spread and recent research has identified additional virus-specific metabolic changes induced by many virus species [28].

The multiple strands and fast infection rates of influenza make it the perfect virus for this study. Many studies have been carried out in recent years due to the bird and swine flu pandemics. The data from a study by Josset et al. [8] was selected to perform FBA against. Investigating virus specific reactions proved challenging due to the nature of influenza. Viral infections typically result in a variety of defence mechanisms within the host cell [29], leading to unexpected metabolic reactions.

The results show a very strong negative correlation between PFU/ml of infected cells and biomass. At that level of investigation, PFU/ml would prove to be an accurate

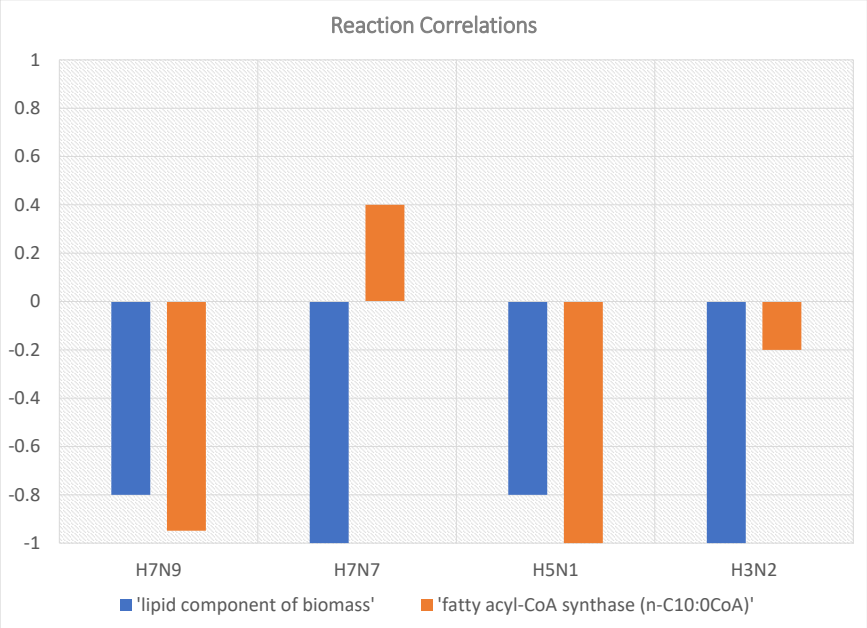

Figure 4. Lipid biomass correlation results (Spearman correlation).

measurement to the degradation of the cells and potency of the virus. However, when investigating the correlative data between individual flux reactions and metabolic pathways it becomes clear the differences in how each influenza virus strain effects the metabolism of human bronchial epithelial cells.

Because of the levels of proven differences between the cell metabolism via this in-silico modelling, it can be said that metabolic flux analysis can prove a more in depth indication to the reasons for a degradation in generic human biomass of a person's cells. Biomass can be used as a general indication of a person's well-being although the reasons behind any degradation to the cell biomass would need to be further investigated via the metabolic flux rates of specific reactions or metabolic pathways. Also, because of the very strong negative correlation to biomass, it can be said that PFU/ml can provide a clear indication of the general well-being of a person.

This paper will serve as a base for future studies and promote the benefits of MFA. The research has thrown up many questions in need of further investigation. Firstly, the effects of H7N7 strain of influenza on the metabolism of squalene raised questions surrounding the poor immuninity of a H7N7 vaccine. With potential reasoning for the poor immuninity, further research into this could help develop a more potent vaccine. Also, Ritter et al. [29] reports interesting results in lactate concentrations and glycolytic activity for human H1N1 infected cells. Further research, extending the results using the strains and techniques from this paper, could investigate differences between virus strains learning more about the potential H7N9 virus for pandemic prevention. Finally, with the knowledge that fatty acid synthesis might play a crucial role for the replication of influenza viruses as they acquire lipid envelopes from their host cells [30]. Further investigation into the effect on lipid biomass across all measured influenza infected strands heavily correlated with the potency of the virus and the similar effects on fatty

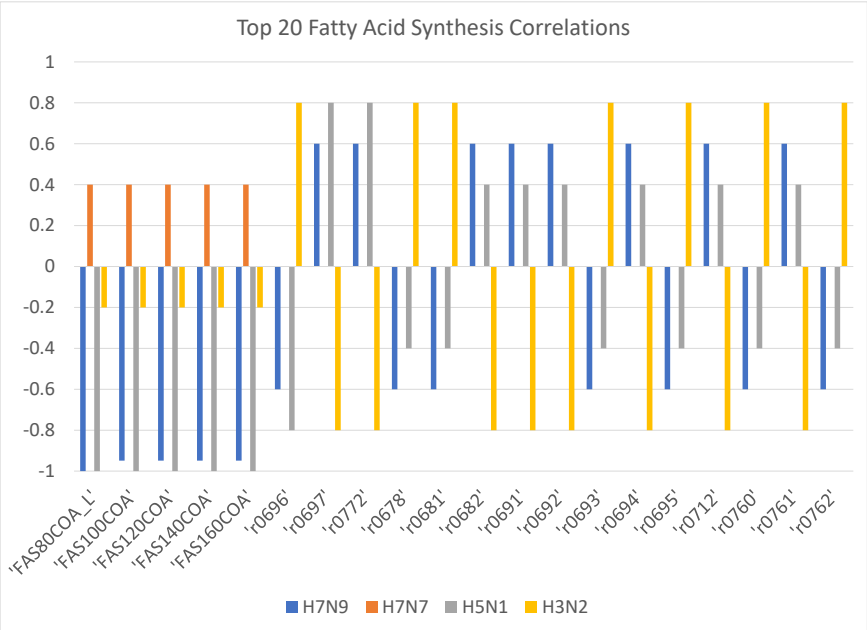

Figure 5. Highly correlated fatty acid synthesis reactions showing similarities between the H7N9 and H5N1 viruses (Spearman correlation).

acid synthesis reactions between the H7N9, H5N1 strand and, potentially the H1N1 strand could be made.

Supporting Information

Correlation Coefficients

The Pearson's correlation coefficient is a common measure of association between two continuous variables. It is defined as the ratio of the covariance of the two variables to the product of their respective standard deviations, commonly denoted by the Greek letter ρ (rho):

$$\rho = \frac{Cov(X,Y)}{\sigma_x, \sigma_y} \tag{8}$$

The sample correlation coefficient, r, can be obtaining by plugging-in the sample covariance and the sample standard deviations into the previous formula, i.e.:

$$r = \frac{\sum_n^{i=1}((x_i - \bar{x})(i_y - \bar{y}))}{\sqrt{\sum_n^{i=1}(x_i - \bar{x})^2 \sum_n^{i=1}(y_i - \bar{y})^2}} \tag{9}$$

where:

$$\bar{x} = \frac{\sum_n^{i=1} x_i}{n}; \bar{y} = \frac{\sum_n^{i=1} y_i}{n}; \tag{10}$$

The Pearson's correlation coefficient ranges from -1 to +1. A positive monotonic association (two variables tend to increase or decrease simultaneously) results in $\rho > 0$, and negative monotonic association (one variable tends to increase when the other decreases) results in $\rho < 0$. ρ of 0 corresponds to the absence of the monotonic association, or absence of any association in the case of bivariate normal data. However, for bivariate distributions other than bivariate normal distribution, the Pearson's correlation can be zero for dependent variables. For example, it can be '0' for the variables with non-monotonic relationship, e.g. $Y = X^2$, (x∈(-1, 1)). The absolute value of ρ indicates the strength of the monotonic relationship between the two variables. 2, 15, 17, 18, 19 ρ of 1 indicates a perfect linear relationship, i.e. Y = a+bX. [20]

The Spearman correlation evaluates the monotonic relationship between two continuous or ordinal variables. In a monotonic relationship, the variables tend to change together, but not necessarily at a constant rate. The Spearman correlation coefficient is based on the ranked values for each variable rather than the raw data. Spearman coefficient could be seen as more applicable if the change in one variable is proportional to the change in another.

Top 10 Spearman correlated data by subsystem

Index	Subsystem	H7N9	H5N1	H3N2	H7N7
1	'Transport, extracellular'	0.049867899	0.044306478	-0.028829976	0.056142462
2	'Transport, endoplasmic reticular'	0.057185723	0.028402248	-0.004540327	-0.020703643
3	'Transport, peroxisomal'	0.001587651	-0.027670187	-0.017105642	-0.018945988
4	'Nucleotide interconversion'	-0.019511766	0.005820299	-0.016208345	0.024689322
5	'Transport, mitochondrial'	-0.039270549	-0.016773424	-0.022205694	-0.03198595
6	'Purine catabolism'	0.14716324	0.058984722	0.009499198	0.019790596
7	'Exchange/demand reaction'	0.001747223	0.005050004	0.001030359	0.002707943
8	'Eicosanoid metabolism'	0.006044074	-0.000692169	-0.006794133	0.002551325
9	Unassigned'	-0.003118048	0.009777155	-0.009560603	0.009945378
10	'Vitamin B2 metabolism'	0.085714286	0.082060951	-0.057069994	-0.212926344

Top 10 Pearson correlated data by subsystem

Index	Subsystem	H7N9	H5N1	H3N2	H7N7
62	'Cysteine Metabolism'	0.002360717	0.62762125	-0.6172134	0.338681784
47	'Squalene and cholesterol synthesis'	-0.641891949	-0.354843702	0.014304409	-0.493281745
49	'D-alanine metabolism'	0.728155754	-0.266566545	0.214495703	-0.232228118
70	'Hyaluronan metabolism'	-0.014931696	-0.662088085	-0.105808011	-0.112431019
61	'Nucleotide salvage pathway'	-0.381646771	-0.073944546	0.220433357	-0.166330534
53	'CoA catabolism'	0.097474648	0.393076313	-0.130974379	0.108430665
50	'R group synthesis'	0.23999436	-0.068206675	0.105391948	-0.255751203
57	'N-glycan synthesis'	-0.238271362	-0.337174488	-0.016328397	0
33	'Citric acid cycle'	0.178608737	0.22102468	0.044018525	0.097873367
74	'Thiamine metabolism'	-0.172131433	-0.002900506	0.153932595	0.182078271

Top 20 Pearson correlated data by reaction

Reactions	SubSystem	H7N9	H7N7	H5N1	H3N2
'CYTDt2r'	Unassigned	0.966319696	0.804120183	0.887558263	-0.9866223
'r1556'	'Transport, extracellular'	0.964648344	-0.55723637	0.970679658	0.953160996
'r2006'	'Transport, extracellular'	-0.770490288	0.88829887	0.821279711	-0.923802156
'r0426'	'Exchange'	-0.857266288	-0.760484597	0.794687116	-0.966986979
'r2274'	'Transport, extracellular'	0.467259425	0.995561665	0.92857565	0.968403372
'r2360'	'Transport, extracellular'	0.755442919	-0.587420494	-0.99456214	-0.998527209
'TAURt4_2_r'	'Transport, extracellular'	0.963166505	0.741701057	0.610574497	0.977907456
'EX_dgsn(e)'	'Exchange/demand reaction'	0.858937233	0.634267031	0.78410701	-0.987999298
'FATP2t'	'Transport, extracellular'	0.899125808	0.918746928	0.479749484	0.959150929
'OCDCEAtr'	'Transport, extracellular'	0.899125808	0.918746928	0.479749484	0.959150929
'r0934'	'Transport, peroxisomal'	-0.963166505	-0.741701057	-0.573598815	-0.977907456
'r2525'	'Transport, extracellular'	-0.954175553	-0.790252663	-0.969568458	-0.531826066
'r2526'	'Transport, extracellular'	-0.954175553	-0.790252663	-0.969568458	-0.531826066
'r1587'	'Transport, extracellular'	0.964648344	-0.568359529	0.970679658	0.740082986
'HCO3_CLt'	'Transport, extracellular'	0.943293833	0.686856556	0.643399741	0.962943752
'r2508'	'Transport, endoplasmic reticular'	0.677680888	-0.890988381	0.674038436	0.993107694
'r0022'	'Glutathione metabolism'	0.58484789	0.993824772	0.992137718	0.661300071
'r2357'	'Transport, extracellular'	0.649648871	-0.772286156	-0.882397402	-0.915965178
'INSt'	'Transport, extracellular'	-0.901702453	-0.81794524	-0.772711356	-0.725852361
'r1683'	'Transport, extracellular'	-0.972123686	-0.848132719	0.770086944	-0.627360884

Top 7 Spearman correlated data by reaction

Reactions	SubSystem	H7N9	H7N7	H5N1	H3N2
3HKYNAKGAT	Tryptophan metabolism	-0.948683298	-0.948683298	-0.948683298	-0.948683298
KYNAKGAT	Tryptophan metabolism	-0.948683298	-0.948683298	-0.948683298	-0.948683298
KYNATESYN	Tryptophan metabolism	-0.948683298	-0.948683298	-0.948683298	-0.948683298
r0647	Tryptophan metabolism	0.948683298	0.948683298	0.948683298	0.948683298
RE1233C	Tryptophan metabolism	0.948683298	0.948683298	0.948683298	0.948683298
RE2349C	Tryptophan metabolism	0.948683298	0.948683298	0.948683298	0.948683298
ILETAm	Valine, leucine, and isoleucine metabolism	0.948683298	0.948683298	0.774596669	0.948683298
ILEt5m	Transport, mitochondrial	0.948683298	0.948683298	0.774596669	0.948683298

Acknowledgments

This research was supported by Teesside University. I thank my supervisor, Claudio Angione, who provided insight and expertise that greatly assisted the research.

References

1. Zien, A. (2004). A primer on molecular biology. In: A. Zien, ed., Computational Analysis of Gene Expression Data, 4th ed. Cambridge, MA, USA: MIT Press, pp.3-34.

2. Murray, D., Doran, P., MacMathuna, P. and Moss, A. (2007). In silico gene expression analysis – an overview. Molecular Cancer, [online] 6(1), p.50. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1964762/ [Accessed 5 Feb. 2017].

3. Klapa, M. and Stephanopoulos, G. (2000). Metabolic Flux Analysis. In: P. Schügerl and P. Bellgardt, ed., Bioreaction Engineering, 1st ed. [online] Springer Berlin Heidelberg, pp.106-124. Available at: http://link.springer.com/book/10.1007/978-3-642-59735-0 [Accessed 2 Feb. 2017].

4. Turner D, Wailoo A, Nicholson K, Couper N, Sutton A, Abrams K. Systematic review and economic decision modeling for the prevention and treatment of influenza A and B. http://www.nice.org.uk/pdf/influenzaassrep.pdf [Accessed Oct 15, 2003].

5. World Health Organisation. (2013). Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2013. [online] Available at: http://www.who.int/influenza/human_animal_interface/EN_GIP_20130604CumulativeNumberH5N1cases.pdf [Accessed 28 Apr. 2017].

6. Nicholson, K.G., Wood, J.M. & Zambon, M. 2003, "Influenza", The Lancet, vol. 362, no. 9397, pp. 1733-45. [online] Available at: http://search.proquest.com/docview/199011914/ [Accessed 28 Apr. 2017].

7. Burgard, A. P., Pharkya, P. & Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and Bioengineering 84, 647–657 (2003).

8. Josset, L., Zeng, H., Kelly, S., Tumpey, T. and Katze, M. (2014). Transcriptomic Characterization of the Novel Avian-Origin Influenza A (H7N9) Virus: Specific Host Response and Responses Intermediate between Avian (H5N1 and H7N7) and Human (H3N2) Viruses and Implications for Treatment Options. mBio, [online] 5(1), pp.e01102-13-e01102-13. Available at: https://www.ncbi.nlm.nih.gov/pubmed/?term=24496798 [Accessed 29 Mar. 2017].

9. Josset, L., Tisoncik-Go, J. and Katze, M. (2013). Moving H5N1 studies into the era of systems biology. Virus Research, 178(1), pp.151-167.

10. Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research, [online] 30(1), pp.207-210. Available at: https://www.researchgate.net/profile/Ron_Edgar/publication/11605619_Edgar_R_Domrachev_M_Lash_AEGene_Expression_Omnibus_NCBI_gene_expression_and_hybridization_array_data_repository_Nucl_Acids_Res_30_207-210/links/54d37a5b0cf28e0697284016/Edgar-R-Domrachev-M-Lash-AEGene-Expression-Omnibus-NCBI-gene-expression-and-hybridization-array-data-repository-Nucl-Acids-Res-30-207-210.pdf [Accessed 13 Feb. 2017].

11. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P. and Mesirov, J. (2006). GenePattern 2.0. Nature Genetics, 38(5), pp.500-501.

12. Schellenberger, J., Que, R., Fleming, R., Thiele, I., Orth, J., Feist, A., Zielinski, D., Bordbar, A., Lewis, N., Rahmanian, S., Kang, J., Hyduke, D. and Palsson, B. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols, [online] 6(9), pp.1290-1307. Available at: https://www.ncbi.nlm.nih.gov/pubmed/21886097 [Accessed 1 Feb. 2017].

13. Swainston, N., Smallbone, K., Hefzi, H., Dobson, P., Brewer, J., Hanscho, M., Zielinski, D., Ang, K., Gardiner, N., Gutierrez, J., Kyriakopoulos, S., Lakshmanan, M., Li, S., Liu, J., Martínez, V., Orellana, C., Quek, L., Thomas, A., Zanghellini, J., Borth, N., Lee, D., Nielsen, L., Kell, D., Lewis, N. and Mendes, P. (2016). Recon 2.2: from reconstruction to model of human metabolism. Metabolomics, [online] 12(7). Available at: http://link.springer.com/article/10.1007/s11306-016-1051-4 [Accessed 1 Feb. 2017].

14. Kinsella, R., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P. and Flicek, P. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database, [online] 2011(0), pp.bar030-bar030. Available at: https://academic.oup.com/database/article/doi/10.1093/database/bar030/465356/Ensembl-BioMarts-a-hub-for-data-retrieval-across [Accessed 12 Apr. 2017].

15. Orth, J., Thiele, I. and Palsson, B. (2010). What is flux balance analysis?. Nature Biotechnology, [online] 28(3), pp.245-248. Available at: http://www.nature.com/nbt/journal/v28/n3/abs/nbt.1614.html [Accessed 28 Apr. 2017].

16. Angione, C. and Lió, P. (2015). Predictive analytics of environmental adaptability in multi-omic network models. Scientific Reports, [online] 5, pp.3-4. Available at: https://www.nature.com/articles/srep15147 [Accessed 13 Apr. 2017].

17. O'Brien, E., Monk, J. and Palsson, B. (2015). Using Genome-scale Models to Predict Biological Capabilities. Cell, [online] 161(5), pp.971-987. Available at: https://doi.org/10.1016/j.cell.2015.05.019 [Accessed 29 Apr. 2017].

18. Altman, D. (1991). Practical statistics for medical research. 1st ed. London: Kluwer Academic Publishers, pp.278-300.

19. Mukaka, M. (2012). A guide to appropriate use of Correlation coefficient in medical research. Malawi Medical Journal: The Journal of Medical Association of Malawi, 24(3), pp.69-71.

20. Chok, N.S., 2010. Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data (Doctoral dissertation, University of Pittsburgh).

21. Iyer, V., Cayatte, C., Guzman, B., Schneider-Ohrum, K., Matuszak, R., Snell, A., Rajani, G., McCarthy, M. and Muralidhara, B. (2015). Impact of formulation and particle size on stability and immunogenicity of oil-in-water emulsion adjuvants. Human Vaccines & Immunotherapeutics, 11(7), pp.1853-1864.

22. Vesikari, T., Pepin, S., Kusters, I., Hoffenbach, A. and Denis, M. (2012). Assessment of squalene adjuvanted and non-adjuvanted vaccines against pandemic H1N1 influenza in children 6 months to 17 years of age. Human Vaccines & Immunotherapeutics, 8(9), pp.1283-1292.

23. Compans, R. and Orenstein, W. (2009). Vaccines for Pandemic Influenza. 1st ed. New York: Springer, pp.43-82.

24. Couch, R., Patel, S., Wade-Bowers, C. and Niño, D. (2012). A Randomized Clinical Trial of an Inactivated Avian Influenza A (H7N7) Vaccine. PLoS ONE, 7(12), p.e49704.

25. Geiler, J., Michaelis, M., Naczk, P., Leutz, A., Langer, K., Doerr, H. and Cinatl, J. (2010). N-acetyl-l-cysteine (NAC) inhibits virus replication and expression of pro-inflammatory molecules in A549 cells infected with highly pathogenic H5N1 influenza A virus. Biochemical Pharmacology, 79(3), pp.413-420.

26. Laing, K. (2004). Chemokines. Developmental & Comparative Immunology, [online] 28(5), pp.443-460. Available at: http://www.sciencedirect.com/science/article/pii/S0145305X03001800 [Accessed 1 May 2017].

27. Yang, S., Qu, J., Wang, C., Yu, X., Liu, Y. and Cao, B. (2013). Influenza pneumonia among adolescents and adults: a concurrent comparison between influenza A (H1N1) pdm09 and A (H3N2) in the post-pandemic period. The Clinical Respiratory Journal, [online] 8(2), pp.185-191. Available at: http://onlinelibrary.wiley.com/doi/10.1111/crj.12056/pdf [Accessed 1 May 2017].

28. Sanchez, E. and Lagunoff, M. (2015). Viral activation of cellular metabolism. Virology, 479-480, pp.609-618.

29. Ritter, J., Wahl, A., Freund, S., Genzel, Y. and Reichl, U. (2010). Metabolic effects of influenza virus infection in cultured animal cells: Intra- and extracellular metabolite profiling. BMC Systems Biology, 4(1), p.61.

30. Janke, R., Genzel, Y., Wetzel, M. and Reichl, U. (2011). Effect of influenza virus infection on key metabolic enzyme activities in MDCK cells. BMC Proceedings, 5(Suppl 8), p.P129.

31. Sanchez, E. and Lagunoff, M. (2015). Viral activation of cellular metabolism. Virology, 479-480, pp.609-618.

32. Nepokroeff, C., Lakshmanan, M. and Porter, J. (1975). [6] Fatty acid synthase from rat liver. Methods in Enzymology, pp.37-44.