



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Integration of omics data to generate and analyse COVID-19 specific genome-scale metabolic models

Tadeja Režen<sup>a</sup>, Alexandre Martins<sup>b</sup>, Miha Mraz<sup>c</sup>, Nikolaj Zimic<sup>c</sup>, Damjana Rozman<sup>a</sup>, Miha Moškon<sup>c,\*</sup>

<sup>a</sup> Centre for Functional Genomics and Bio-Chips, Institute for Biochemistry and Molecular Genetics, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

<sup>b</sup> ESIGLEEC Graduate School of Engineering, Rouen, France

<sup>c</sup> Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia

### ARTICLE INFO

#### Keywords:

COVID-19  
Genome-scale metabolic models  
Model extraction methods  
Context-specific models  
Metabolic enrichment analysis

### ABSTRACT

COVID-19 presents a complex disease that needs to be addressed using systems medicine approaches that include genome-scale metabolic models (GEMs). Previous studies have used a single model extraction method (MEM) and/or a single transcriptomic dataset to reconstruct context-specific models, which proved to be insufficient for the broader biological contexts. We have applied four MEMs in combination with five COVID-19 datasets. Models produced by GIMME were separated by infection, while tINIT preserved the biological variability in the data and enabled the best prediction of the enrichment of metabolic subsystems. Vitamin D3 metabolism was predicted to be down-regulated in one dataset by GIMME, and in all by tINIT. Models generated by tINIT and GIMME predicted downregulation of retinol metabolism in different datasets, while downregulated cholesterol metabolism was predicted only by tINIT-generated models. Predictions are in line with the observations in COVID-19 patients. Our data indicated that GIMME and tINIT models provided the most biologically relevant results and should have a larger emphasis in further analyses. Particularly tINIT models identified the metabolic pathways that are a part of the host response and are potential antiviral targets. The code and the results of the analyses are available to download from [https://github.com/CompBioLj/COVID\\_GEMs\\_and\\_MEMs](https://github.com/CompBioLj/COVID_GEMs_and_MEMs).

### 1. Introduction

The novel coronavirus (SARS-CoV-2) and the COVID-19 disease pandemic have impacted many aspects of our lives since its initial emergence and spread in the early 2020s. Various computational approaches have quickly been developed and have already been used to aid diagnosis, predict outcomes, analyse dynamics, and improve treatment of this disease. For example, Zoabi et al. [1] applied machine learning approaches to predict COVID-19 diagnosis based on the presence and severity of clinical symptoms [2]. used artificial intelligence methods to diagnose and predict the COVID-19 patient's response to treatment. The COPASI tool [3] was used to model the dynamics of SARS-CoV-2 using differential equations [4]. A recent review of computational approaches dedicated to combating the COVID-19 pandemic has been presented by Hufsky et al. [4].

Among recent computational approaches aimed to address the so-far limited understanding of COVID-19 disease and dynamics are also Genome-scale metabolic models (GEMs). These models are gaining

relevance in various scientific fields from bioengineering and production of biologics to the understanding of complex diseases [5]. Manual reconstruction of an integrated host-virus GEM has already been used to identify and confirm potential antiviral targets [6,7]. However, because this model only incorporated a SARS-CoV-2 virus biomass target function (VBOF) into a GEM of human alveolar macrophages, its predictive power could be further improved by fitting metabolic constraints to the diseased state of a metabolic network. A similar approach of modelling the infection in the hepatocyte-derived cellular carcinoma cell line Huh7 using the manual extension of the Recon 2.2 model was reported by Yaneske et al. [8].

GEMs can be used in a combination with different model extractions methods (MEMs), which allow us to automatically adapt a model and its constraints to a specific context using omics data [9]. One of the main problems is that different MEMs produce variable, sometimes contradictory results [10]. Different datasets gathered either from COVID-19 patients or cell lines have been applied to the analysis of COVID-19 metabolic signatures by genome-scale metabolic modelling and

\* Corresponding author.

E-mail address: [miha.moskon@fri.uni-lj.si](mailto:miha.moskon@fri.uni-lj.si) (M. Moškon).

<https://doi.org/10.1016/j.combiomed.2022.105428>

Received 4 February 2022; Received in revised form 17 March 2022; Accepted 19 March 2022

Available online 23 March 2022

0010-4825/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

automatically generated context-specific GEMs [11,12]. However, only selected MEM approaches have been applied in the present studies. Namely, Nanda and Ghosh [11] applied iMAT and Cheng et al. [12] applied tINIT extraction method to variable scope of COVID-19 datasets.

In this study, we describe the integration of various publicly available COVID-19 datasets using different MEMs in combination with the Human-GEM metabolic model [13]. We reconstruct an individual metabolic model for each of the datasets and healthy groups and each of the MEMs, namely iMAT [14], INIT [15], tINIT [16] and GIMME [17]. We perform a thorough comparison of obtained models and discuss their biological relevance. Moreover, we generate a set of flux samples [18] using each of these models, which are used to identify enriched metabolic subsystems in each of the models between the infected and healthy groups. These are then used to perform a comparative analysis of the obtained results, which we discuss in a broader biological context.

## 2. Background

The coronavirus causing COVID-19, SARS-CoV-2, belongs to the family *Coronaviridae* and is a single-stranded RNA virus surrounded by a lipid envelope containing several viral proteins [19]. Viruses do not possess enzymes to produce the necessary ingredients and energy to build new viruses. Therefore, upon entry into the host cell, the virus hijacks host cell machinery, modifies many metabolic pathways and even the composition of membranes [20]. Antivirals targeting the host metabolic factors are emerging as an alternative therapeutic strategy [21]. Predicting and understanding the changes in the host metabolism is a necessary step towards the development of new antivirals. Computational approaches, such as GEMs, enabling the reconstruction of host metabolism after viral infection can aid in the search for new antiviral targets.

GEMs have gained large importance in different fields ranging from biotechnology to systems biology and systems medicine [5,22]. For example, GEMs have been applied in cancer research [23,24], drug targeting in pathogens [25], prediction of drug side effects [26], and optimisation of production of recombinant proteins using Chinese hamster ovary (CHO) cells [27]. Reconstruction and analysis of GEMs are mostly based on constraint-based approaches. The main representatives of these approaches are flux balance analysis (FBA) and its extensions [28,29]. FBA performs a steady-state analysis of the metabolic network. This is based on the constraints of the observed metabolic reactions (i.e., feasible metabolic flux ranges), and a predefined objective function. FBA yields a solution describing metabolic flux values that maximise the objective function within the given constraints. A limitation of FBA is that it yields a single non-unique solution, which describes only one of the optimal steady-states of the system. This can be addressed with the application of different FBA extensions, such as flux variability analysis (FVA), which assesses the metabolic flux ranges that maintain a specific state of the metabolic network [30]. Another alternative is to apply parsimonious FBA (pFBA) that yields a unique solution by maximising the objective function and simultaneously minimising the overall fluxes through the metabolic network [31]. However, these approaches still require a specification of one or more optimisation functions. These might strongly affect the obtained results and are hard to define in a general setting [32,33]. Even though using a biomass accumulation function as an optimisation criterion is plausible in certain scenarios (e.g., for studies involving microorganisms or cancer cells), its general applicability is questionable [34]. An alternative approach to the analysis of GEMs without the requirement to specify a cellular objective is flux sampling. This is based on a generation of representative metabolic fluxes covering the feasible solution space [35]. Flux sampling generates a possible range of metabolic flux values as in the case of FVA, however, in an unbiased manner.

Available GEM reconstructions usually describe a biological system in a generic state. For example, the Human-GEM model describes a reference reconstruction of a human cell [13]. This needs to be adapted

to a specific cell type in a specific context before further analyses are performed [23]. Model extraction methods (MEMs) allow the automatic integration of different omics (usually transcriptomics) data to adapt a generic model to a given context using gene-protein-reaction (GPR) rules encoded in a model [36]. For example, GIMME aims to remove the reactions catalysed by the products of genes with expression levels below a predefined threshold [17,37]. iMAT additionally aims to pertain to reactions catalysed by the products of highly expressed genes [9,14,37]. The latter is also the case of INIT, which does not impose a strict steady-state assumption and thus allows the accumulation of certain metabolites [15]. tINIT presents an extension of INIT, in which the resulting model needs to be able to perform a given set of metabolic tasks. Context-specific GEMs have already been applied to the analysis of omics data describing different conditions and for understanding complex systems disorders (e.g., see Ref. [38]). Such analyses have also been conducted in the domain of the analysis of COVID-19 metabolic signatures [11,12]. However, present studies have only applied a single MEM to the analysed datasets. Since each MEM is based on certain assumptions different MEMs might produce significantly different or even contradictory results [10]. The predictive power of extracted models can be increased if different MEMs are analysed in a given context to guide the selection of a method that yields models with the largest significance (e.g., models preserving the separation between the observed experimental groups) [39]. Furthermore, the selection of pre-processing and omics data integration steps also has a significant effect on obtained results and the use of various tools, approaches and datasets significantly increase the accuracy, precision and robustness of analyses [40–43]. Setting up workflows combining the integration of different datasets and tools enabling systematic assessment of omics data is one of the main goals of systems medicine [44,45].

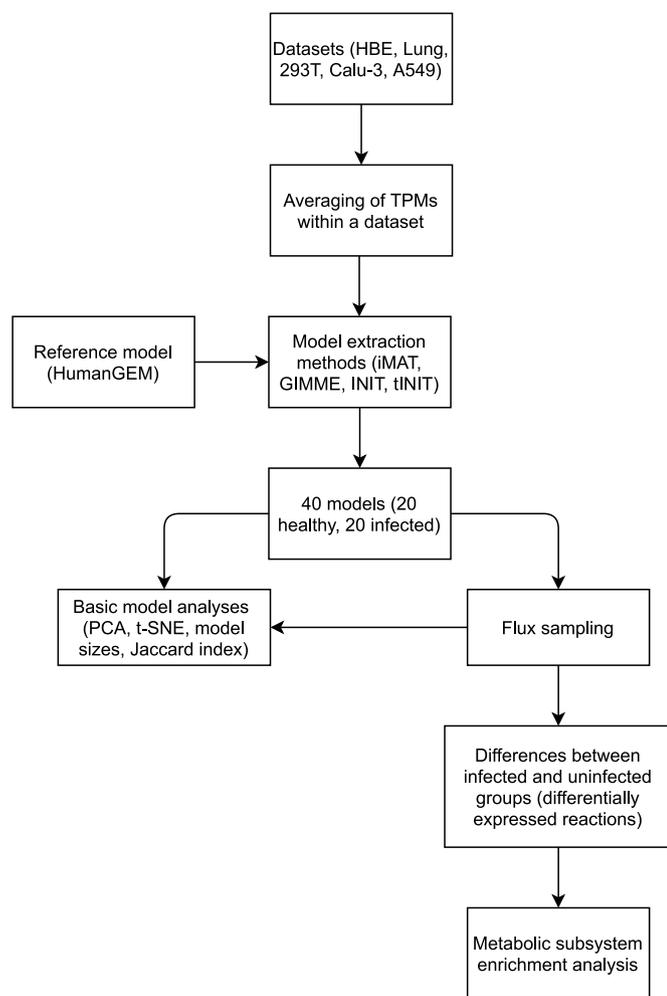
In this work, we describe an approach to perform a selection of the most suitable MEMs in a combination with different COVID-19-specific datasets. The proposed selection is guided by different analytical approaches including principal component analysis (PCA), evaluation of model distances based on Jaccard index metrics, and analysis of the reactions and their flux values observed in different settings. Moreover, using the selected MEMs we perform detailed analyses of obtained context-specific GEMs to compare the metabolic states between healthy and infected cells. There are two main contributions of this study. Firstly, we propose a set of analyses that can be used to identify the most suitable MEM for an exact GEM-based analysis of observed datasets. Secondly, we analyse the models produced by selected MEMs in a wider biological context and identify the metabolic pathways that are a part of the host response and thus present potential antiviral targets.

## 3. Methods and data

The whole process of data acquisition and preprocessing, integration of data into a reference model with the extraction of context-specific models and their analysis is illustrated in Fig. 1 and described in the following section.

### 3.1. COVID-19 datasets

We collected the data needed to reconstruct the models from publicly available repositories. We identified the repositories with transcriptomes of healthy and COVID-19-specific human bronchial epithelial (HBE), lung biopsy cells, human embryonic kidney (293T), Calu-3, and adenocarcinoma human alveolar basal epithelial (A549) cell lines as described by Blanco-Melo et al. [46] and Weingarten-Gabbay et al. [47]. Since each dataset belongs to a different cell type, it also defines the biological context in which the measurement of the transcriptome was performed. Next, we used the SRA toolkit [48,49] to download the raw transcriptomic data, from each of the repositories. We used the Kallisto tool [50] to obtain the transcripts-per-million (TPM) values for the observed transcripts. However, to extract the models, we had to evaluate



**Fig. 1.** Model extraction and analysis pipeline as applied in this study. Abbreviations and symbols: HBE–human bronchial epithelial cells; 293T–human embryonic kidney cells; A549–human alveolar basal epithelial cells; TPM–transcripts-per-million; PCA–principal component analysis; t-SNE–t-distributed stochastic neighbour embedding.

the expression levels of each of the observed genes. Since a gene can have multiple transcripts, its expression level was evaluated by summing the levels of all of the transcripts belonging to a gene. Finally, we calculated the average TPM values for each gene for each of the observed experiments and each of the observed conditions, namely healthy and infected.

### 3.2. Extraction of COVID-19 specific models

We used the Human-GEM model version 1.6.0 [13] as a scaffold model for the extraction process.<sup>1</sup> Before reconstructing infected models, the model was augmented with the virus biomass objective function (VBOF) as described by Nanda and Ghosh [11]. Different model extraction methods (MEMs) were used in the reconstruction, namely iMAT [14], INIT, tINIT [16], GIMME [17]. Models were reconstructed for both the infected and the healthy group. 75th percentile of the average TPM values within a dataset and within a condition (infected or healthy) was used as a threshold value for each of the algorithms. During the reconstruction, the default Human-GEM model biomass function was used for the healthy state and VBOF for the infected state.

<sup>1</sup> Available at <https://github.com/SysBioChalmers/Human-GEM/releases/tag/v1.6.0>.

After the reconstruction, we verified that a generated model could perform the essential metabolic tasks. For healthy cells, these were obtained from the Human-GEM model repository [13]. However, for infected cells, the metabolic task file was acquired from the repository accompanying the paper [11]. If the models could not perform all the specified metabolic tasks, they were augmented with necessary reactions as described by Robinson et al. [13].

### 3.3. Analysis of extracted models

We evaluated how the differences in model sizes depend on different factors. The significance of model size differences in dependence on a factor was evaluated using the Kruskal-Wallis H-test in cases where the number of groups was larger than 2 and the two-sample Kolmogorov-Smirnov test elsewhere. Moreover, we analysed the specificity of active reactions in dependence on each factor using the Jaccard index metric, which can be used to assess the similarity between a pair of models as

$$J(M_i, M_j) = \frac{|M_i \cap M_j|}{|M_i \cup M_j|} \quad (1)$$

Here  $M_i$  presents a set of reactions in model  $i$  and  $M_j$  a set of reactions in model  $j$  [51]. Jaccard indices were assessed for all pairs of models and then used to compare the reaction specificity within each of the observed groups of models, and between the observed group of models and the remaining models (boxplot visualisation and two-sample Kolmogorov-Smirnov test).

The reference model, which was used as a scaffold model in the extraction process, has more than 13,000 reactions. Extracted models eliminate inactive reactions for a given context but can still preserve up to more than 3,000 active reactions (see 2). Dimensionality reduction techniques such as principal component analysis (PCA) allow us to project high-dimensional data to a space with a manageable level of dimensions. This projection can be used to guide further analyses. For example, PCA plots in two-dimensional space can be used to assess how well obtained models cluster together in dependence on different factors. In our case, we aimed to find a configuration of the extraction process (namely, selection of the most suitable MEM) that would be able to separate the healthy models from the infected models. Moreover, PCA can be used to assess the relative impact of each of the observed factors (namely MEM, dataset, and infection) on a model content [10]. We thus opted to analyse the obtained models using PCA, the results of which were also confirmed by t-SNE (t-distributed stochastic neighbour embedding) plots [52,53]. The analysis was conducted in a similar manner as described by Opdam et al. [9] and later applied by Walakira et al. [39]. We constructed a matrix describing the presence or absence of each metabolic reaction in each of the reconstructed models. We evaluated the amount of the variability between the models that is explained by each of the factors of extraction, namely MEM, dataset, and presence or absence of the infection. We observed the clustering of obtained models and their compliance with predefined groups using PCA and t-SNE plots.

To analyse the dynamics of extracted models and to identify an active set of reactions for a given context, we generated 1000 flux samples for each of the reconstructed models using the Artificial Centering Hit-and-Run (ACHR) sampler [54]. These samples were then used to evaluate the effects of different factors on model sizes, specificity of reactions, and variability between the generated models. Moreover, we used the generated flux samples to identify the enriched metabolic reaction. The two-sample Kolmogorov-Smirnov test was used to identify the reactions that were differentially expressed between healthy and infected models. Within the enrichment analysis additional criteria to identify the significantly changed reactions needs to be established. If flux sampling is used, any number of samples and thus an arbitrary low p-value for a statistical test can be obtained [12]. We used fold-change criteria (FC) between the healthy and infected groups, as described

previously by Nanda and Ghosh [11]:

$$FC = \frac{\bar{R}_i - \bar{R}_h}{|\bar{R}_i + \bar{R}_h|}, \quad (2)$$

where  $\bar{R}_i$  is the average reaction activity in infected and  $\bar{R}_h$  is the average reaction activity in healthy models. We used the FC criteria to further identify the significantly altered reactions, namely to detect up- and down-regulated reactions after the infection. For the former, we used the FC value cutoff value of at least 0.82 (10-fold activation), and for the latter the cutoff value of at most  $-0.82$  (10-fold inhibition) between the infected and the healthy groups [11].

GEMs are composed of different metabolic subsystems containing metabolic reactions with specific functionalities. For example, Human-GEM, which was used as a reference model in our analysis, is composed of 142 subsystems ranging from Glycolysis/Gluconeogenesis to Cholesterol Metabolism. We used the generated lists of significantly up- and down-regulated reactions for each of the MEMs and each of the datasets to perform the enrichment analysis of the metabolic subsystems with the hypergeometric test as described by Walakira et al. [39]. We identified the metabolic subsystems that were consistently enriched in the same dataset using different MEMs and/or consistently enriched in different datasets using the same MEMs. These results were then used to perform the biological interpretation of the analysed datasets and compared with related COVID-19 studies.

Reconstruction of context-specific models, their curation, and sampling of model fluxes were performed with the COBRA [55] and RAVEN Toolboxes [56] in Matlab R2019b (MathWorks Inc., Natick, Massachusetts, USA) using the Gurobi solver (Gurobi Optimisation, LLC, Beaverton, Oregon, USA). Further analyses of reconstructed models, visualisation, and enrichment analysis were performed in Python 3.<sup>2</sup>

## 4. Results

### 4.1. Models vary by size in dependence on different factors

We analysed how the size of the extracted models depends on different factors, namely on a MEM, a dataset, and infection. The distributions of model sizes in dependence on different factors are presented in Fig. 2(A). Differences in model sizes were significant only between different MEMs ( $p < 10^{-5}$ , Kruskal-Wallis H-test). Models produced with GIMME pertained to the largest number of reactions (mean model size was 7134.6 reactions) and models produced with iMAT the smallest number of reactions (mean model size was 3035.4 reactions).

To identify the number of reactions that carry nonzero flux in simulations, we generated a set of flux samples for each of the models. We evaluated the size for each of the models by counting the reactions that had nonzero flux in at least one of the samples. The distributions of model sizes after flux sampling is presented in Fig. 2(B). This analysis indicated that the majority of reactions pertained in the GIMME-produced model were inactive in simulations (mean model size was 2751.8 reactions). On the other hand, iMAT-produced models pertained to the majority of the reactions even after the flux sampling analysis (mean model size was 2903.9 reactions). Even though differences between model sizes were smaller in this experiment, these differences were significant between MEMs ( $p < 10^{-6}$ , Kruskal-Wallis H-test) and between the infected and uninfected group of models ( $p = 0.023$ , two-sample Kolmogorov–Smirnov test). This indicates that the reaction sets preserved after the flux sampling reflected larger biological significance in a given context than the full sets of reactions identified solely by a MEM.

### 4.2. Metabolic reactions should be analysed on the quantitative level

We proceeded with the analysis of the obtained models based on the generated flux samples. We analysed the specificity of active reactions in dependence on each factor using the Jaccard index metric [51]. We analysed the distribution of this metric within each factor and between the factors (see Fig. 3).

We further analysed the significance of specificity of active reactions of each of the factors using the two-sample Kolmogorov–Smirnov test. We adjusted the calculated p-values for multiple testing using the Benjamini-Hochberg procedure. Active reactions were significantly specific for each of the MEMs and all the datasets except for the HBE data. Reactions were not significantly specific for neither healthy nor infected models. This indicated that differences between healthy and infected models should be investigated further based on quantitative flux values through the observed reactions and/or within a specific MEM/dataset. Complete results of the reaction specificity analysis are available in Supplementary Table 1.

### 4.3. Infection and cell lines separate GIMME- and tINIT-produced models

Principal component analysis (PCA) can be used to evaluate the degree of variability in the models explained by each of the factors [9, 39]. As in the case of the reaction specificity analysis, we conducted the PCA analysis based on generated flux samples. Namely, reactions that had zero fluxes in all generated samples were removed from a model.

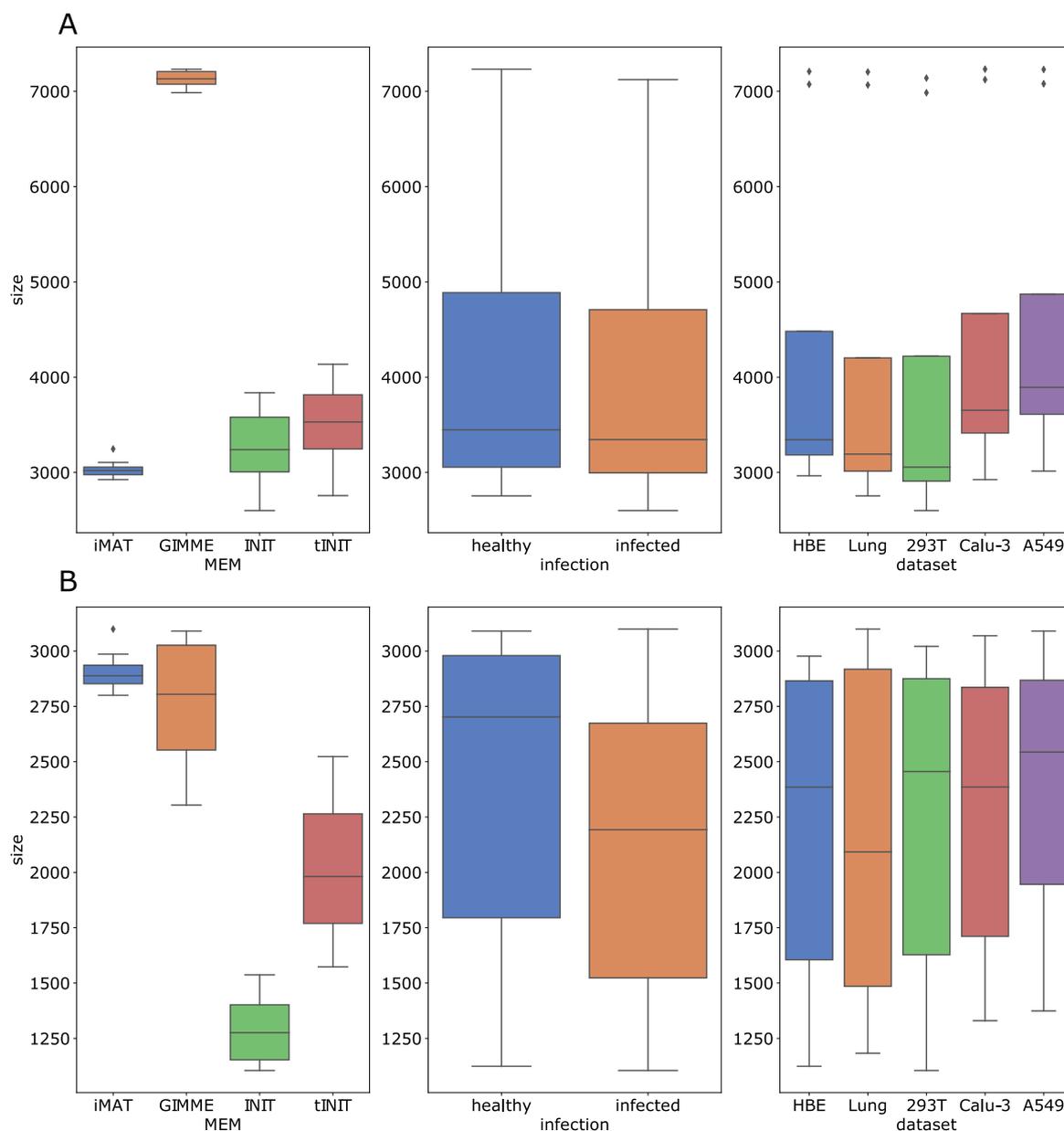
As expected, the model extraction method (MEM) described the largest amount of variability in the data. More precisely, the first principal component (PC1) was able to explain around 30% of the variability and MEM described almost 89% of the variability in the PC1. The separation of models observed in the PCA plot using the first two principal components (see Supplementary Fig. 1 and Supplementary Table 2) complied with the separation observed in the t-SNE plot (see Supplementary Fig. 2). We further analysed the models produced with each MEM to evaluate the amount of variability explained by the remaining factors. In all cases, a dataset explained a significantly larger amount of variability than infection (see Supplementary Table 3). Within the PC1 the former equalled almost 94% for iMAT (see Supplementary Figure 3), 82.3% for GIMME (see Supplementary Figure 4), 54.6% for tINIT (see Supplementary Figure 5), and 71.5% for tINIT (see Supplementary Fig. 6). However, even though a dataset explained a larger amount of variability, models extracted using GIMME could be separated by infection using the PC2 explaining 21.4% of variability in the data, which was almost as much as PC1, which explained 25.3% of variability (see Supplementary Fig. 4).

Models were also partially separated by infection using tINIT and PC1, which explained around 23.3% of variability in the data. The only outliers were healthy models obtained using lung biopsy cell data (see Supplementary Fig. 6). However, models produced by tINIT were separated also by the datasets, which represent separate biological entities. These were cells or tissue infected by the SARS-CoV virus, which differed in cell-specific metabolic steady-states, levels of viral load and host response to the virus. GIMME models did not separate by the dataset, while PCA analysis of the original data showed strong separation by the dataset/cell type except for lung biopsy cell data. Separation by infection and cell type in GIMME and tINIT models indicated that these two MEMs can generate models which reflect large compliance with predefined groups of data and should have a larger emphasis in further analysis.

### 4.4. Results of the enrichment analysis vary with a MEM

We performed the enrichment analysis using the two-sample Kolmogorov–Smirnov test to identify up- and down-regulated metabolic reactions. Full results of the enrichment analysis are available at the supplementary GitHub repository. Furthermore, we used the lists of

<sup>2</sup> Available at [https://github.com/CompBioLj/COVID\\_GEMs\\_and\\_MEMs](https://github.com/CompBioLj/COVID_GEMs_and_MEMs).



**Fig. 2.** Extracted model sizes in dependence on different factors. The data visualised in Figure (A) were obtained by counting the reactions that can carry nonzero flux in each of the reconstructed models. The data visualised in Figure (B) were obtained by generating flux samples and counting the reactions for which nonzero flux was obtained in at least one of the samples for a given model. Abbreviations and symbols: MEM–model extraction method; HBE–human bronchial epithelial cells; 293T–human embryonic kidney cells; A549–human alveolar basal epithelial cells.

enriched metabolic reactions in a combination with metabolic subsystems as defined in the reference model [13] to perform the hypergeometric test, and to identify up- and down-regulated metabolic subsystems in each of the models (see also Section 3).

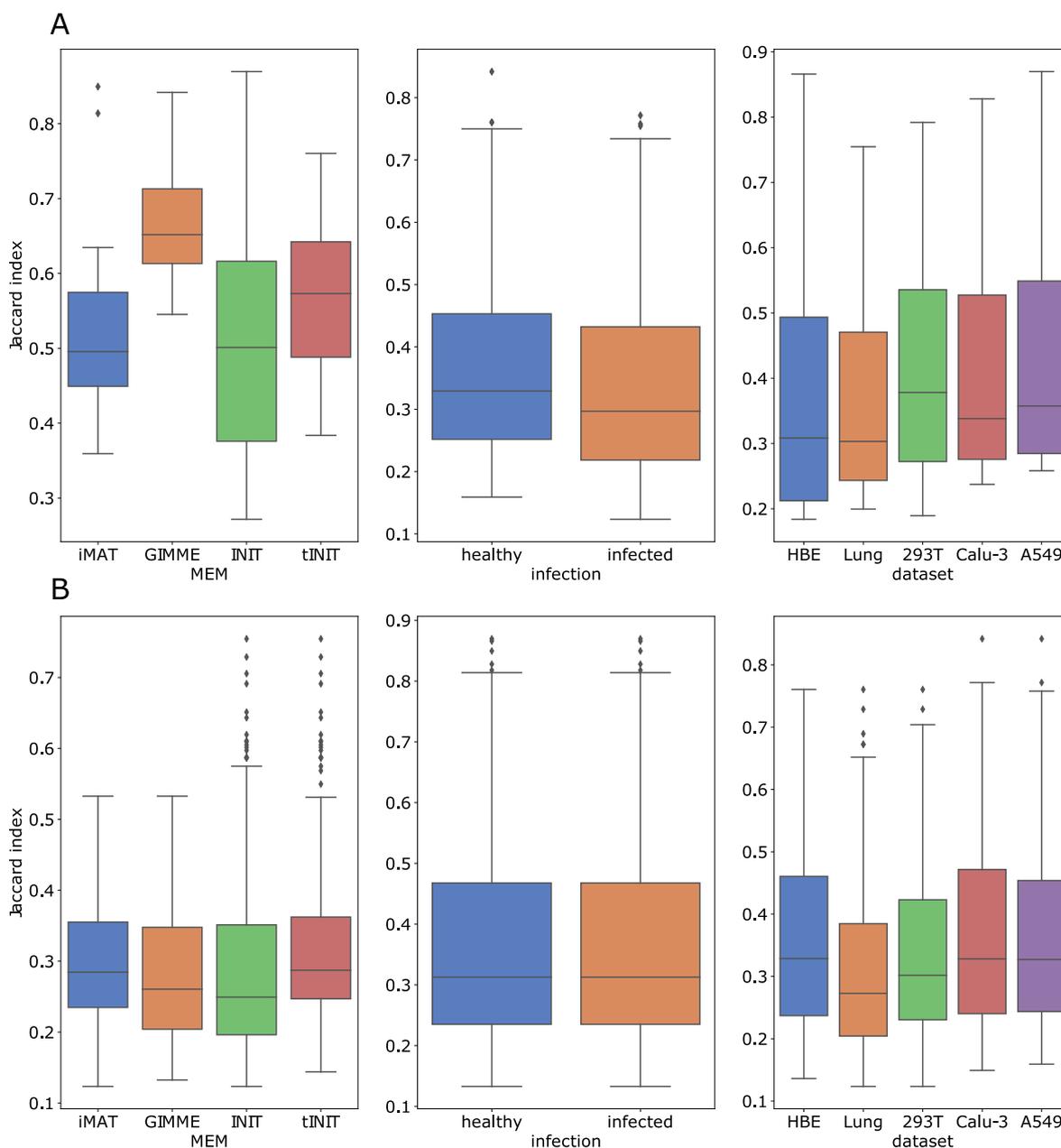
Different MEMs and different datasets produce different, in some cases contradictory results [9]. We analysed a specific MEM by extracting the metabolic subsystems that were consistently enriched (either up- or down-regulated) in a predefined number of models obtained with all of the observed datasets and with a given MEM. Fig. 4 visualises the 23 metabolic subsystems that were consistently enriched in at least two out of five datasets for a MEM. All enriched subsystems are available as [Supplementary Fig. 7](#).

tINIT has identified more than twice the number of enriched metabolic subsystems than GIMME combined for all datasets. Interestingly, GIMME found the lowest number of enriched subsystems in comparison to all MEMs. Enriched metabolic subsystems were mainly from the lipid-

related metabolism, from fatty acids, glycerophospholipids to cholesterol and their metabolites, ubiquinone, vitamins, etc. GIMME and tINIT had few commonly enriched subsystems in the same dataset. These were downregulation of vitamin E and D metabolism, upregulation of transport reactions, and opposite direction of fatty acid biosynthesis (odd chain). These results confirm that the selection of MEM has a higher effect on results than the dataset itself.

#### 4.5. Results of the enrichment analysis vary with a dataset

We opted to analyse the subsystems which were consistently enriched in at least a predefined number of MEMs for each dataset in a similar way as described in Section 4.4. Fig. 5 visualises the 12 metabolic subsystems that were consistently enriched in at least two out of four MEMs for a dataset. All enriched subsystems are available as [Supplementary Fig. 8](#).



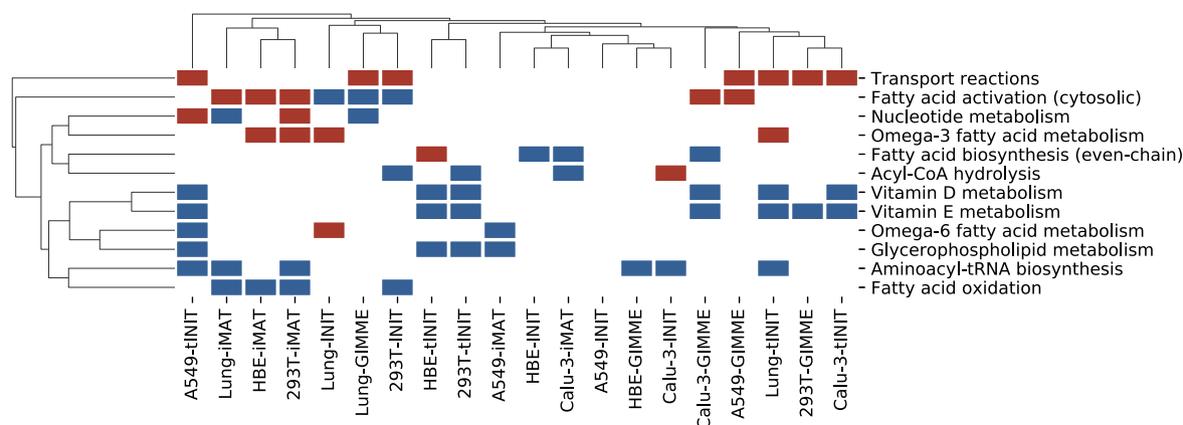
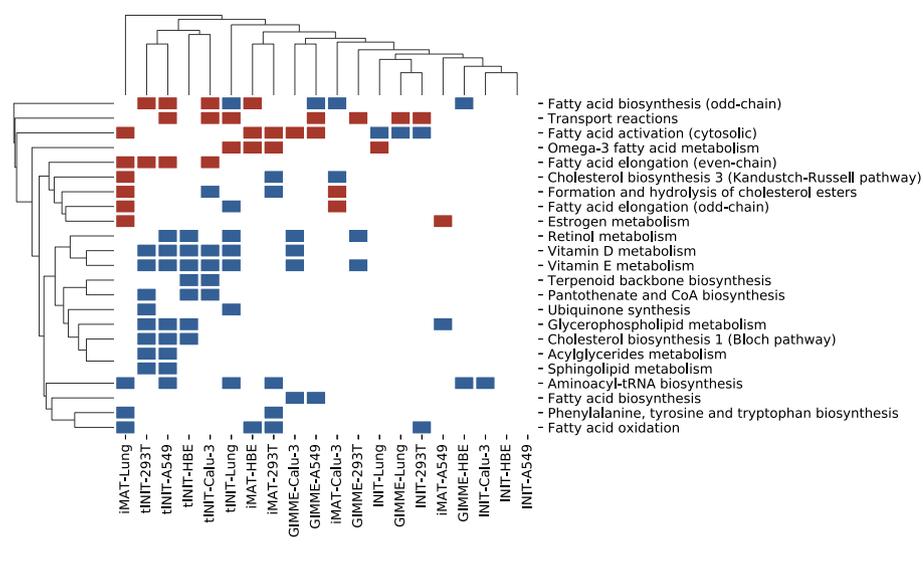
**Fig. 3.** The distributions of Jaccard indices within a factor (A) and between the factors (B). Jaccard indices were assessed for all pairs of models and then used to compare the reaction specificity within each of the observed group of models, and between each of the observed group of models and the remaining models. Figure (A) presents the distributions of Jaccard indices between pairs of models belonging to the indicated group. Figure (B) presents the distributions of Jaccard indices between pairs of models where exactly one of the models belongs to the indicated group. Abbreviations and symbols: MEM—model extraction method; HBE—human bronchial epithelial cells; 293T—human embryonic kidney cells; A549—human alveolar basal epithelial cells.

Commonly enriched subsystems were relatively specific for each of the datasets, similarly as observed in the case of MEM-based enrichment analysis. This is in line with the fact that each dataset represents a different cell type, which differs in the level of susceptibility to the infection and host response. There were only a few enriched metabolic subsystems common between three MEMs, while there is not even one subsystem commonly predicted by all MEMs in the same dataset. Therefore, we can confirm that the selection of MEM had the highest effect on the results and can significantly affect the biological interpretation.

## 5. Discussion

Survivors of severe or critical COVID-19 have long-lasting metabolic

abnormalities [57]. Multiple studies confirmed changes in metabolism in COVID-19 patients with significant changes in serum lipids [58]. Moreover, serum metabolome could be used as a predictive and diagnostic biomarker [59,60]. Serum cholesterol and fatty acids were among the metabolites, which can predict progression to severe COVID-19. Our analysis correctly identified that modulation of lipid metabolic pathways in cells infected by SARS-CoV-2 is in line with the observations in COVID-19 patients. Several of the enriched pathways have been also indicated for antiviral therapy to decrease the severity of the disease, and have been confirmed to have an active role in the SARS-CoV-2 infection and replication in preclinical models. Taking into account the PCA results and the data from COVID-19 patients, tINIT best preserved the biological variability in the data and enabled the best prediction of enrichment of metabolic subsystems in comparisons to other



**Fig. 5.** Common metabolic subsystems between model extraction methods. Represented are metabolic subsystems that were consistently enriched in at least two out of four different model extraction methods for a dataset after the infection. Blue colour represents downregulation and red colour upregulation of a metabolic subsystem. White colour represents no significant change. Abbreviations and symbols: HBE–human bronchial epithelial cells; 293T–human embryonic kidney cells; A549–human alveolar basal epithelial cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

#### MEMs.

Vitamin D3 is produced in the skin by UV irradiation from 7-dehydrocholesterol. It is biologically active after it is converted by enzymes in the liver and kidney to 1,25-dihydroxycholecalciferol. Vitamin D3 metabolism was predicted to be down-regulated in one dataset by GIMME models, and in all by tINIT models. This is in line with several meta-analyses of published data, which found significantly lower 25 (OH)D concentration, and significant relation between the low concentration and infection, severity or mortality of COVID-19 patients [61–66]. Vitamin D3 supplementation has been hypothesised to affect clinical outcomes in patients with COVID-19, however; this was not successfully confirmed [61,67]. Currently, vitamin D3 supplementation is not recommended [68]. However, all studies agree that the lack of randomised controlled studies, small cohorts, different dosages and vitamin D formulations, non-standardised study conditions and large heterogeneity between the studies hinders the evaluation and conclusions. Especially since studies in preclinical models showed that vitamin D3 inhibits SARS-CoV-2 virus replication in cells [69,70], but affect also other aspects of COVID-19 disease [71]. Additionally, the analysis of tINIT models proposed that the metabolism of vitamin D3 molecules is changed in different cell types and this could affect the efficacy and availability of vitamin D3 in the body. Therefore, different dosages and

**Fig. 4.** Common metabolic subsystems between datasets. Represented are metabolic subsystems that were consistently enriched in at least two out of five datasets for a model extraction method after the infection. Blue colour represents downregulation and red colour upregulation of a metabolic subsystem. White colour represents no significant change. Abbreviations and symbols: HBE–human bronchial epithelial cells; 293T–human embryonic kidney cells; A549–human alveolar basal epithelial cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

vitamin D formulations would need to be tested in clinical studies.

Retinol, a biologically active form of vitamin A, maintains the innate and adaptive immunity in viral infections [72]. Both tINIT and GIMME models predicted downregulation of retinol metabolism in different datasets, which is in line with the observations in COVID-19 patients. Low levels of vitamin A in plasma of COVID-19 patients was observed and this was significantly associated with the severity of the disease [73, 74]. High-throughput screening of natural compounds found that all-trans retinoic acid inhibited 3C-like protease of SARS-CoV-2 and by this exhibited antiviral effect [75]. Additionally, an AM580 compound that has inhibited viral replication of MERS-CoV (Middle East respiratory syndrome) *in vitro* and *in vivo* in mice, is a retinoid derivative and RAR $\alpha$  (Retinoic Acid Receptor alpha) agonist [76]. However, a meta-analysis of studies on vitamin A supplements in children did not confirm the effect on respiratory tract infection [77]. Both tINIT and GIMME models predicted also downregulation of vitamin E metabolism in different datasets, however, only one study so far measured the vitamin E level in COVID-19 patients, which was found unchanged [73].

Modulation of different pathways involved in fatty acid metabolism was predicted by GEMs in applied datasets or cells. Increased levels of plasma free fatty acids and triglycerides were observed in multiple COVID-19 patient studies. Meta-analysis of omics data confirmed

alterations in fatty acid metabolism [58]. The viral infections are known to hijack the lipid biosynthesis to enable viral reproduction [78]. The SARS-CoV-2 spike protein undergoes palmitoyl modification [79] and needs new fatty acids for membrane formation. FASN (Fatty Acid Synthase) is one of the rate-limiting enzymes in fatty acid synthesis and the key enzyme in the synthesis of palmitate. Knockdown of FASN in cell lines resulted in lower SARS-CoV-2 infection and lower quantity of viral RNA [80]. The same study also showed that inhibiting fatty acid synthesis by drugs, such as orlistat, lowered viral levels in the lung and increased survival in a mouse model. Interestingly, a study in elderly patients showed a lower metabolic flux in the fatty acid pathway in survivors vs the deceased [81]. A more detailed analysis at the level of a single fatty acid showed opposite changes in levels of different fatty acids in serum of patients with severe COVID-19 depending on the type of the fatty acid and its desaturation index [82]. For example, the palmitic and stearic acid, needed for the formation of viral membranes, were decreased. A decrease in serum 2-palmitoyl-glycerol in COVID-19 patients was one of the three potential diagnostic markers of the disease [59]. The retinoid derivative, AM580, is also an SREBF (Sterol Regulatory Element Binding Transcription Factor) inhibitor. SREBF is the major regulator of lipid metabolism and could express its antiviral function through this mechanism [76].

Downregulation of cholesterol and terpenoid biosynthesis was predicted by tINIT models in almost all datasets. Numerous studies have compared the plasma cholesterol level in COVID-19 patients, and two meta-analyses of this data concluded that total, HDL and LDL cholesterol are significantly lower in hospitalized severe patients and non-survivors [83,84]. Although there are some discrepancies between the studies in terms of the direction of change and which serum/plasma parameter is changed, they all confirm that cholesterol metabolism is affected by the SARS-CoV-2 infection. Moreover, factors involved in cholesterol homeostasis are actively involved in viral infection [85]. Cholesterol-rich domains are relevant for viral budding [86]. Membrane cholesterol was found to be essential for SARS-CoV-2 entry into the cell [87]. SARS-CoV-2 nucleoprotein interacts with NPC1 (Niemann-Pick type C1), an intracellular cholesterol transporter, and inhibitors of this interaction were able to reduce infection *in vitro* [33]. Moreover, functional interrogation of host factors required for SARS-CoV-2 infection identified several major regulators of cholesterol biosynthesis and homeostasis [88–90]. For example, HDL scavenger receptor SCARB1 (Scavenger Receptor Class B Member 1) facilitates virus entry via cholesterol-binding site [91]. The 25-hydroxycholesterol was identified as an antiviral host factor that inhibits SARS-CoV-2 infection [92,93].

## 6. Conclusion

We applied different MEMS in a combination with various scopes of publicly available COVID-19 specific datasets to analyse the metabolic signatures and potentially propose novel antiviral targets. We proposed a set of analyses that can be used to identify the most suitable model extraction method that can be used in the precise analysis of observed datasets. We propose that PCA and/or tSNE analyses can help explore the data and select the MEM that is the best in preserving the biological variability of the data, which seems to be crucial for generating precise predictions. When we paralleled the enrichment of perturbed metabolic pathways with patient data, we confirmed that tINIT-produced models have identified several fatty acid and cholesterol metabolic pathways, which were modulated also in COVID-19 patients. Moreover, these pathways were identified as factors of host response and potential drug targets. Most interestingly, tINIT models have predicted lower metabolism of several vitamins and thus a potential lack of vitamins needed to fight the virus.

## Authors' contributions

**Tadeja Režen:** Methodology, Formal analysis, Investigation, Writing

– Original Draft, Writing – Review & Editing **Alexandre Martins:** Methodology, Software, Formal analysis **Miha Mráz:** Writing – Review & Editing, Funding acquisition **Nikolaj Zimic:** Writing – Review & Editing, Funding acquisition **Damjana Rozman:** Writing – Review & Editing, Funding acquisition **Miha Moskon:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualisation, Supervision.

## Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

## Acknowledgements

The research was partially supported by the scientific-research program Pervasive Computing (P2-0359), by the scientific-research program Functional Genomics and Biotechnology for Health (P1-0390), by the basic research project CholesteROR in metabolic liver diseases (J1-9176), and by the Network of infrastructure Centres of University of Ljubljana (MRIC-UL-CFGBC, IP-022), all financed by the Slovenian Research Agency. AM was supported by the Erasmus+ exchange programme of the European Union. We would also like to acknowledge the support by the infrastructure project ELIXIR-SI RI-SI-2 financed by the European Regional Development Fund and by the Ministry of Education, Science and Sport of Republic of Slovenia. The funding sources had no role in the design of the study and collection, analysis, and interpretation of data nor in writing the manuscript.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compbimed.2022.105428>.

## References

- [1] Y. Zoabi, S. Deri-Rozov, N. Shomron, Machine learning-based prediction of COVID-19 diagnosis based on symptoms, *npj Digit. Med.* 4 (2021) 1–5, <https://doi.org/10.1038/s41746-020-00372-6>.
- [2] S.M. ELGhamrawy, A.E. Hassanien, Diagnosis and prediction model for COVID-19 patients response to treatment based on convolutional neural networks and whale optimization algorithm using CT images, *medRxiv* (2020), <https://doi.org/10.1101/2020.04.16.20063990>.
- [3] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, U. Kummer, COPASI—a complex pathway simulator, *Bioinformatics* 22 (2006) 3067–3074, <https://doi.org/10.1093/bioinformatics/btl485>.
- [4] F. Hufsky, K. Lamkiewicz, A. Almeida, A. Aouacheria, C. Arighi, A. Bateman, J. Baumbach, N. Beerenwinkel, C. Brandt, M. Cacciabue, et al., Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research, *Brief. Bioinform.* 22 (2021) 642–663, <https://doi.org/10.1093/bib/bbaa232>.
- [5] C. Gu, G.B. Kim, W.J. Kim, H.U. Kim, S.Y. Lee, Current status and applications of genome-scale metabolic models, *Genome Biol.* 20 (2019) 1–18, <https://doi.org/10.1186/s13059-019-1730-3>.
- [6] A. Renz, L. Widerspich, A. Dräger, FBA reveals guanylate kinase as a potential target for antiviral therapies against SARS-CoV-2, *Bioinformatics* 36 (2020) i813–i821, <https://doi.org/10.1093/bioinformatics/btaa813>.
- [7] A. Renz, L. Widerspich, A. Dräger, Genome-scale metabolic model of infection with SARS-CoV-2 mutants confirms guanylate kinase as robust potential antiviral target, *Genes* 12 (2021) 796, <https://doi.org/10.3390/genes12060796>.
- [8] E. Yaneske, G. Zampieri, L. Bertoldi, G. Benvenuto, C. Angione, Genome-scale metabolic modelling of SARS-CoV-2 in cancer cells reveals an increased shift to glycolytic energy production, *FEBS Lett.* 595 (2021) 2350–2365, <https://doi.org/10.1002/1873-3468.14180>.
- [9] S. Opdam, A. Richelle, B. Kellman, S. Li, D.C. Zielinski, N.E. Lewis, A systematic evaluation of methods for tailoring genome-scale metabolic models, *Cell Syst.* 4 (2017) 318–329, <https://doi.org/10.1016/j.cels.2017.01.010>.
- [10] A. Richelle, A.W. Chiang, C.-C. Kuo, N.E. Lewis, Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions, *PLoS Comput. Biol.* 15 (2019), e1006867, <https://doi.org/10.1371/journal.pcbi.1006867>.
- [11] P. Nanda, A. Ghosh, Genome scale-differential flux analysis reveals deregulation of lung cell metabolism on SARS-CoV-2 infection, *PLoS Comput. Biol.* 17 (2021), e1008860, <https://doi.org/10.1371/journal.pcbi.1008860>.

- [12] K. Cheng, L. Martin-Sancho, L.R. Pal, Y. Pu, L. Riva, X. Yin, S. Sinha, N.U. Nair, S. K. Chanda, E. Ruppín, Genome-scale metabolic modeling reveals SARS-CoV-2-induced metabolic changes and antiviral targets, *Mol. Syst. Biol.* 17 (2021), e10260, <https://doi.org/10.15252/msb.202110260>.
- [13] J.L. Robinson, P. Kocabaş, H. Wang, P.-E. Cholley, D. Cook, A. Nilsson, M. Anton, R. Ferreira, I. Domenzain, V. Billa, et al., An atlas of human metabolism, *Sci. Signal.* 13 (2020), <https://doi.org/10.1126/scisignal.aaz1482> eaaz1482.
- [14] H. Zur, E. Ruppín, T. Shlomi, iMAT: an integrative metabolic analysis tool, *Bioinformatics* 26 (2010) 3140–3142, <https://doi.org/10.1093/bioinformatics/btq602>.
- [15] R. Agren, S. Bordel, A. Mardinoglu, N. Pornputtpong, I. Nookaew, J. Nielsen, Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT, *PLoS Comput. Biol.* 8 (2012), e1002518, <https://doi.org/10.1371/journal.pcbi.1002518>.
- [16] R. Agren, A. Mardinoglu, A. Asplund, C. Kampf, M. Uhlen, J. Nielsen, Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling, *Mol. Syst. Biol.* 10 (2014) 721, <https://doi.org/10.1002/msb.14512>.
- [17] S.A. Becker, B.O. Palsson, Context-specific metabolic networks are consistent with experiments, *PLoS Comput. Biol.* 4 (2008), e1000082, <https://doi.org/10.1371/journal.pcbi.1000082>.
- [18] W. Megchelenbrink, M. Huynen, E. Marchiori, optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks, *PLoS One* 9 (2014), e86587, <https://doi.org/10.1371/journal.pone.0086587>.
- [19] Q. Han, Q. Lin, S. Jin, L. You, Coronavirus 2019-nCoV: a brief perspective from the front line, *J. Infect.* 80 (2020) 373–377, <https://doi.org/10.1016/j.jinf.2020.02.010>.
- [20] M.R. Denison, Seeking membranes: positive-strand RNA virus replication complexes, *PLoS Biol.* 6 (2008), e270, <https://doi.org/10.1371/journal.pbio.0060270>.
- [21] S. Mahajan, S. Choudhary, P. Kumar, S. Tomar, Antiviral strategies targeting host factors and mechanisms obliging +ssRNA viral pathogens, *Bioorg. Med. Chem.* 46 (2021) 116356, <https://doi.org/10.1016/j.bmc.2021.116356>.
- [22] C. Zhang, Q. Hua, Applications of genome-scale metabolic models in biotechnology and systems medicine, *Front. Physiol.* 6 (2016) 413, <https://doi.org/10.3389/fphys.2015.00413>.
- [23] A. Nilsson, J. Nielsen, Genome scale metabolic modeling of cancer, *Metab. Eng.* 43 (2017) 103–112, <https://doi.org/10.1016/j.ymben.2016.10.022>.
- [24] T. Barata, V. Vieira, R. Rodrigues, R.P. das Neves, M. Rocha, Reconstruction of tissue-specific genome-scale metabolic models for human cancer stem cells, *Comput. Biol. Med.* 142 (2022), 105177, <https://doi.org/10.1016/j.cmpbiomed.2021.105177>.
- [25] A.M. Abdel-Haleem, H. Hefzi, K. Mineta, X. Gao, T. Gojbori, B.O. Palsson, N. E. Lewis, N. Jamshidi, Functional interrogation of plasmodium genus metabolism identifies species- and stage-specific differences in nutrient essentiality and drug targeting, *PLoS Comput. Biol.* 14 (2018), e1005895, <https://doi.org/10.1371/journal.pcbi.1005895>.
- [26] I. Shaked, M.A. Oberhardt, N. Atias, R. Sharan, E. Ruppín, Metabolic network prediction of drug side effects, *Cell Syst.* 2 (2016) 209–213, <https://doi.org/10.1016/j.cels.2016.03.001>.
- [27] Ž. Rejc, L. Magdevska, T. Tršelič, T. Osolin, J. Mraz, E. Pavliha, N. Zimic, T. Cvitanović, D. Rozman, M. Moškon, M. Mraz, Computational modelling of genome-scale metabolic networks and its application to CHO cell cultures, *Comput. Biol. Med.* 88 (2017) 150–160, <https://doi.org/10.1016/j.cmpbiomed.2017.07.005>.
- [28] J.D. Orth, I. Thiele, B.O. Palsson, What is flux balance analysis? *Nat. Biotechnol.* 28 (2010) 245–248, <https://doi.org/10.1038/nbt.1614>.
- [29] K. Raman, N. Chandra, Flux balance analysis of biological systems: applications and challenges, *Brief. Bioinform.* 10 (2009) 435–449, <https://doi.org/10.1093/bib/bbp011>.
- [30] S. Gudmundsson, I. Thiele, Computationally efficient flux variability analysis, *BMC Bioinform.* 11 (2010) 1–3, <https://doi.org/10.1186/1471-2105-11-489>.
- [31] N.E. Lewis, K.K. Hixson, T.M. Conrad, J.A. Lerman, P. Charusanti, A.D. Polpitiya, J. N. Adkins, G. Schramm, S.O. Purvine, D. Lopez-Ferrer, et al., Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models, *Mol. Syst. Biol.* 6 (2010) 390, <https://doi.org/10.1038/msb.2010.47>.
- [32] C.E. García Sánchez, R.G. Torres Sáez, Comparison and analysis of objective functions in flux balance analysis, *Biotechnol. Prog.* 30 (2014) 985–991, <https://doi.org/10.1002/btpr.1949>.
- [33] I. García-Dorival, M.Á. Cuesta-Geijo, L. Barrado-Gil, I. Galindo, U. Garaigorta, J. Urquiza, A. Del Puerto, N.-E. Campillo, A. Martínez, P. Gastaminza, et al., Identification of Niemann-Pick C1 protein as a potential novel SARS-CoV-2 intracellular target, *Antivir. Res.* 194 (2021), 105167, <https://doi.org/10.1016/j.antiviral.2021.105167>.
- [34] A. Bordbar, J.M. Monk, Z.A. King, B.O. Palsson, Constraint-based models predict metabolic and associated cellular functions, *Nat. Rev. Genet.* 15 (2014) 107–120, <https://doi.org/10.1038/nrg3643>.
- [35] H.A. Herrmann, B.C. Dyson, L. Vass, G.N. Johnson, J.-M. Schwartz, Flux sampling is a powerful tool to study metabolism under changing environmental conditions, *npj Syst. Biol. Appl.* 5 (2019) 1–8, <https://doi.org/10.1038/s41540-019-0109-0>.
- [36] J.Y. Ryu, H.U. Kim, S.Y. Lee, Reconstruction of genome-scale human metabolic models using omics data, *Integr. Biol.* 7 (2015) 859–868, <https://doi.org/10.1039/c5ib00002e>.
- [37] A.S. Blazier, J.A. Papin, Integration of expression data in genome-scale metabolic network reconstructions, *Front. Physiol.* 3 (2012) 299, <https://doi.org/10.3389/fphys.2012.00299>.
- [38] J.S. Cho, C. Gu, T.H. Han, J.Y. Ryu, S.Y. Lee, Reconstruction of context-specific genome-scale metabolic models using multiomics data to study metabolic rewiring, *Curr. Opin. Struct. Biol.* 15 (2019) 1–11, <https://doi.org/10.1016/j.coisb.2019.02.009>.
- [39] A. Walakira, D. Rozman, T. Režen, M. Mraz, M. Moškon, Guided extraction of genome-scale metabolic models for the integration and analysis of omics data, *Comput. Struct. Biotechnol. J.* 8 (2021) 3521–3530, <https://doi.org/10.1016/j.csbj.2021.06.009>.
- [40] Q. Yang, Y. Wang, Y. Zhang, F. Li, W. Xia, Y. Zhou, Y. Qiu, H. Li, F. Zhu, NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data, *Nucleic Acids Res.* 48 (2020) W436–W448, <https://doi.org/10.1093/nar/gkaa258>.
- [41] B. Li, J. Tang, Q. Yang, S. Li, X. Cui, Y. Li, Y. Chen, W. Xue, X. Li, F. Zhu, NOREVA: normalization and evaluation of MS-based metabolomics data, *Nucleic Acids Res.* 45 (2017) W162–W170, <https://doi.org/10.1093/nar/gkx449>.
- [42] J. Fu, Y. Zhang, Y. Wang, H. Zhang, J. Liu, J. Tang, Q. Yang, H. Sun, W. Qiu, Y. Ma, Z. Li, M. Zhaorong, F. Zhu, Optimization of metabolomic data processing using NOREVA, *Nat. Protoc.* 17 (2022) 129–151, <https://doi.org/10.1038/s41596-021-00636-9>.
- [43] J. Tang, J. Fu, Y. Wang, B. Li, Y. Li, Q. Yang, X. Cui, J. Hong, X. Li, Y. Chen, W. Xue, F. Zhu, ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies, *Brief. Bioinform.* 21 (2020) 621–636, <https://doi.org/10.1093/bib/bby127>.
- [44] J. Tang, J. Fu, Y. Wang, Y. Luo, Q. Yang, B. Li, G. Tu, J. Hong, X. Cui, Y. Chen, L. Yao, W. Xue, F. Zhu, Simultaneous improvement in the precision, accuracy, and robustness of label-free proteome quantification by optimizing data manipulation chains, *Mol. Cell. Proteomics* 18 (2019) 1683–1699, <https://doi.org/10.1074/mcp.RA118.001169>.
- [45] Q. Yang, B. Li, J. Tang, X. Cui, Y. Wang, X. Li, J. Hu, Y. Chen, W. Xue, Y. Lou, Y. Qiu, F. Zhu, Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data, *Brief. Bioinform.* 21 (2020) 1058–1068, <https://doi.org/10.1093/bib/bbz049>.
- [46] D. Blanco-Melo, B.E. Nilsson-Payant, W.-C. Liu, S. Uhl, D. Hoagland, R. Møller, T. X. Jordan, K. Oishi, M. Panis, D. Sachs, et al., Imbalanced host response to SARS-CoV-2 drives development of COVID-19, *Cell* 181 (2020) 1036–1045, <https://doi.org/10.1016/j.cell.2020.04.026>.
- [47] S. Weingarten-Gabbay, S. Klaeger, S. Sarkizova, L.R. Pearlman, D.-Y. Chen, M. R. Bauer, H.B. Taylor, H.L. Conway, C.H. Tomkins-Tinch, Y. Finkel, et al., SARS-CoV-2 infected cells present HLA-I peptides from canonical and out-of-frame ORFs, *bioRxiv* (2020), <https://doi.org/10.1101/2020.10.02.324145>.
- [48] R. Leinonen, H. Sugawara, M. Shumway, International nucleotide sequence database collaboration, the sequence read archive, *Nucleic Acids Res.* 39 (2010) D19–D21, <https://doi.org/10.1093/nar/gkq1019>.
- [49] Y. Kodama, M. Shumway, R. Leinonen, The sequence read archive: explosive growth of sequencing data, *Nucleic Acids Res.* 40 (2012) D54–D56, <https://doi.org/10.1093/nar/gkr854>.
- [50] N.L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification, *Nat. Biotechnol.* 34 (2016) 525–527, <https://doi.org/10.1038/nbt.3519>.
- [51] N.C. Chung, B. Miasojedow, M. Startek, A. Gambin, Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data, *BMC Bioinform.* 20 (2019) 1–11, <https://doi.org/10.1186/s12859-019-3118-5>.
- [52] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (2008) 2579–2605.
- [53] L. van der Maaten, Accelerating t-SNE using tree-based algorithms, *J. Mach. Learn. Res.* 15 (2014) 3221–3245.
- [54] D.E. Kaufman, R.L. Smith, Direction choice for accelerated convergence in hit-and-run sampling, *Oper. Res.* 46 (1998) 84–95, <https://doi.org/10.1287/opre.46.1.84>.
- [55] L. Heirendt, S. Arreckx, T. Pfau, S.N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S.M. Keating, V. Vlasov, et al., Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0, *Nat. Protoc.* 14 (2019) 639–702, <https://doi.org/10.1038/s41596-018-0098-2>.
- [56] H. Wang, S. Marcišauskas, B.J. Sánchez, I. Domenzain, D. Hermansson, R. Agren, J. Nielsen, E.J. Kerkhoven, RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*, *PLoS Comput. Biol.* 14 (2018), e1006541, <https://doi.org/10.1371/journal.pcbi.1006541>.
- [57] J.S. Ayres, A metabolic handbook for the COVID-19 pandemic, *Nat. Metab.* 2 (2020) 572–585, <https://doi.org/10.1038/s42255-020-0237-2>.
- [58] S. Aggarwal, A. Acharjee, A. Mukherjee, M.S. Baker, S. Srivastava, Role of multiomics data to understand host–pathogen interactions in COVID-19 pathogenesis, *J. Proteome Res.* 20 (2021) 1107–1132, <https://doi.org/10.1021/acs.jproteome.0c00771>.
- [59] D. Shi, R. Yan, L. Lv, H. Jiang, Y. Lu, J. Sheng, J. Xie, W. Wu, J. Xia, K. Xu, et al., The serum metabolome of COVID-19 patients is distinctive and predictive, *Metabolism* 118 (2021) 154739, <https://doi.org/10.1016/j.metabol.2021.154739>.
- [60] B. Shen, X. Yi, Y. Sun, X. Bi, J. Du, C. Zhang, S. Quan, F. Zhang, R. Sun, L. Qian, et al., Proteomic and metabolomic characterization of COVID-19 patient sera, *Cell* 182 (2020) 59–72, <https://doi.org/10.1016/j.cell.2020.05.032>.
- [61] A. Crafa, R. Cannarella, R.A. Condorelli, L.M. Mongioi, F. Barbagallo, A. Aversa, S. La Vignera, A.E. Calogero, Influence of 25-hydroxy-cholecalciferol levels on SARS-CoV-2 infection and COVID-19 severity: a systematic review and meta-analysis, *EclinicalMedicine* 37 (2021), 100967, <https://doi.org/10.1016/j.eclinm.2021.100967>.
- [62] A. Kazemi, V. Mohammadi, S.K. Aghababae, M. Golzarand, C.C. Clark, S. Babajafari, Association of vitamin D status with SARS-CoV-2 infection or COVID-

- 19 severity: a systematic review and meta-analysis, *Adv. Nutr.* 12 (2021) 1636–1658, <https://doi.org/10.1093/advances/nmab012>.
- [63] F. Petrelli, A. Luciani, G. Perego, G. Dognini, P.L. Colombelli, A. Ghidini, Therapeutic and prognostic role of vitamin D for COVID-19 infection: a systematic review and meta-analysis of 43 observational studies, *J. Steroid Biochem. Mol. Biol.* 211 (2021) 105883, <https://doi.org/10.1016/j.jsbmb.2021.105883>.
- [64] A. Teshome, A. Adane, B. Girma, Z.A. Mekonnen, The impact of vitamin D level on COVID-19 infection: systematic review and meta-analysis, *Front. Public Health* 9 (2021) 169, <https://doi.org/10.3389/fpubh.2021.624559>.
- [65] N. Liu, J. Sun, X. Wang, T. Zhang, M. Zhao, H. Li, Low vitamin D status is associated with coronavirus disease 2019 outcomes: a systematic review and meta-analysis, *Int. J. Infect. Dis.* 104 (2021) 58–64, <https://doi.org/10.1016/j.ijid.2020.12.077>.
- [66] L. Szarpak, Z. Rafique, A. Gasecka, F. Chirico, W. Gawel, J. Hernik, H. Kaminska, K. J. Filipiak, M.J. Jaguszewski, L. Szarpak, A systematic review and meta-analysis of effect of vitamin D levels on the incidence of COVID-19, *Cardiol. J.* 28 (2021) 647–654, <https://doi.org/10.5603/CJ.A2021.0072>.
- [67] A. Bassatne, M. Basbous, M. Chakhtoura, O. El Zein, M. Rahme, G.E.-H. Fuleihan, The link between COVID-19 and vitamin D (VIVID): a systematic review and meta-analysis, *Metabolism* 119 (2021), 154753, <https://doi.org/10.1016/j.metabol.2021.154753>.
- [68] J.J. Malin, C.D. Spinner, U. Janssens, T. Welte, S. Weber-Carstens, G. Schälte, P. Gastmeier, F. Langer, M. Wepler, M. Westhoff, et al., Key summary of German national treatment guidance for hospitalized COVID-19 patients, *Infection* 50 (2022) 93–106, <https://doi.org/10.1007/s15010-021-01645-2>.
- [69] A. Pickard, B.C. Calverley, J. Chang, R. Garva, S. Gago, Y. Lu, K.E. Kadler, Discovery of re-purposed drugs that slow SARS-CoV-2 replication in human cells, *PLoS Pathog.* 17 (2021), e1009840, <https://doi.org/10.1371/journal.ppat.1009840>.
- [70] S. Qayyum, T. Mohammad, R.M. Slominski, M.I. Hassan, R.C. Tuckey, C. Raman, A. T. Slominski, Vitamin D and lumisterol novel metabolites can inhibit SARS-CoV-2 replication machinery enzymes, *Am. J. Physiol. Endocrinol. Metab.* 321 (2021) E246–E251, <https://doi.org/10.1152/ajpendo.00174.2021>.
- [71] S. Kouhpayeh, L. Shariati, M. Boshtam, I. Rahimmanesh, M. Mirian, Y. Esmaeili, M. Najafu, N. Khanahmad, M. Zeinalian, M. Trovato, et al., The molecular basis of COVID-19 pathogenesis, conventional and nanomedicine therapy, *Int. J. Mol. Sci.* 22 (2021) 5438, <https://doi.org/10.3390/ijms22115438>.
- [72] C. Stephensen, G. Lietz, Vitamin A in resistance to and recovery from infection: relevance to SARS-CoV-2, *Br. J. Nutr.* 126 (2021) 1663–1672, <https://doi.org/10.1017/S0007114521000246>.
- [73] T.-M. Tomasa-Irriguible, L. Bielsa-Berocal, L. Bordejé-Laguna, C. Tural-Llàcher, J. Barallat, J.-M. Manresa-Domínguez, P. Torán-Monserrat, Low levels of few micronutrients may impact COVID-19 disease progression: an observational study on the first wave, *Metabolites* 11 (2021) 565, <https://doi.org/10.3390/metabo11090565>.
- [74] P.-R. Tepaspe, R. Vollenberg, M. Fobker, I. Kabar, H. Schmidt, J.A. Meier, T. Nowacki, A. Hüsing-Kabar, Vitamin A plasma levels in COVID-19 patients: a prospective multicenter study and hypothesis, *Nutrients* 13 (2021) 2173, <https://doi.org/10.3390/nu13072173>.
- [75] T. Morita, K. Miyakawa, S.S. Jeremiah, Y. Yamaoka, M. Sada, T. Kuniyoshi, J. Yang, H. Kimura, A. Ryo, All-trans retinoic acid exhibits antiviral effect against SARS-CoV-2 by inhibiting 3CLpro activity, *Viruses* 13 (2021) 1669, <https://doi.org/10.3390/v13081669>.
- [76] S. Yuan, H. Chu, J.F.-W. Chan, Z.-W. Ye, L. Wen, B. Yan, P.-M. Lai, K.-M. Tee, J. Huang, D. Chen, et al., SREBP-dependent lipidomic reprogramming as a broad-spectrum antiviral target, *Nat. Commun.* 10 (2019) 1–15, <https://doi.org/10.1038/s41467-018-08015-x>.
- [77] B. Vlieg-Boerstra, N. de Jong, R. Meyer, C. Agostoni, V. De Cosmi, K. Grimshaw, G. P. Milani, A. Muraro, H. Oude Elberink, I. Pali-Schöll, et al., Nutrient supplementation for prevention of viral respiratory tract infections in healthy subjects: a systematic review and meta-analysis, *Allergy* (2021), <https://doi.org/10.1111/all.15136>.
- [78] K.A. Mayer, J. Stöckl, G.J. Zlabinger, G.A. Gualdoni, Hijacking the supplies: metabolism as a novel facet of virus-host interaction, *Front. Immunol.* 10 (2019) 1533, <https://doi.org/10.3389/fimmu.2019.01533>.
- [79] Z. Wu, Z. Zhang, X. Wang, J. Zhang, C. Ren, Y. Li, L. Gao, X. Liang, P. Wang, C. Ma, Palmitoylation of SARS-CoV-2 s protein is essential for viral infectivity, *Signal Transduct. Target Ther.* 6 (2021) 1–4, <https://doi.org/10.1038/s41392-021-00651-y>.
- [80] J. Chu, C. Xing, Y. Du, T. Duan, S. Liu, P. Zhang, C. Cheng, J. Henley, X. Liu, C. Qian, et al., Pharmacological inhibition of fatty acid synthesis blocks SARS-CoV-2 replication, *Nat. Metab.* 3 (2021) 1466–1475, <https://doi.org/10.1038/s42255-021-00479-4>.
- [81] Q. Mei, A.Y. Wang, A. Bryant, Y. Yang, M. Li, F. Wang, S. Du, C. Kurts, P. Wu, K. Ma, et al., Survival factors and metabolic pathogenesis in elderly patients ( $\geq 65$ ) with COVID-19: a multi-center study, *Front. Med.* (2021), 595503, <https://doi.org/10.3389/fmed.2020.595503>.
- [82] I. Pérez-Torres, V. Guarner-Lans, E. Soria-Castro, L. Manzano-Pech, A. Palacios-Chavarría, R.R. Valdez-Vázquez, J.G. Domínguez-Cherit, H. Herrera-Bello, H. Castillejos-Suastegui, L. Moreno-Castañeda, et al., Alteration in the lipid profile and the desaturases activity in patients with severe pneumonia by SARS-CoV-2, *Front. Physiol.* 12 (2021) 624, <https://doi.org/10.3389/fphys.2021.667024>.
- [83] R.K. Mahat, V. Rathore, N. Singh, N. Singh, S.K. Singh, R.K. Shah, C. Garg, Lipid profile as an indicator of COVID-19 severity: a systematic review and meta-analysis, *Clin. Nutr. ESPEN* 45 (2021) 91–101, <https://doi.org/10.1016/j.clnesp.2021.07.023>.
- [84] A. Zinellu, P. Paliogiannis, A.G. Fois, P. Solidoro, C. Carru, A.A. Mangoni, Cholesterol and triglyceride concentrations, COVID-19 severity, and mortality: a systematic review and meta-analysis with meta-regression, *Front. Public Health* 9 (2021), <https://doi.org/10.3389/fpubh.2021.705916>.
- [85] E. Kočar, T. Režen, D. Rozman, Cholesterol, lipoproteins, and COVID-19: basic concepts and clinical applications, *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids* 1866 (2021), 158849, <https://doi.org/10.1016/j.bbalip.2020.158849>.
- [86] F. Mesquita, L. Abrami, O. Sergeeva, P. Turelli, B. Kunz, C. Raclot, J. Paz Montoya, L.A. Abriat, M. Dal Pararo, D. Trono, et al., S-acylation controls SARS-CoV-2 membrane lipid organization and enhances infectivity, *Dev. Cell* 56 (2021) 2790–2807, <https://doi.org/10.1016/j.devcel.2021.09.016>, e8.
- [87] D.W. Sanders, C.C. Jumper, P.J. Ackerman, D. Bracha, A. Donlic, H. Kim, D. Kenney, I. Castello-Serrano, S. Suzuki, T. Tamura, et al., SARS-CoV-2 requires cholesterol for viral entry and pathological syncytia formation, *Elife* 10 (2021), e65962, <https://doi.org/10.7554/eLife.65962>.
- [88] H.-H. Hoffmann, F.J. Sánchez-Rivera, W.M. Schneider, J.M. Luna, Y.M. Soto-Feliciano, A.W. Ashbrook, J. Le Pen, A.A. Leal, I. Ricardo-Lax, E. Michailidis, et al., Functional interrogation of a SARS-CoV-2 host protein interactome identifies unique and shared coronavirus host factors, *Cell Host Microbe* 29 (2021) 267–280, <https://doi.org/10.1016/j.chom.2020.12.009>.
- [89] R. Wang, C.R. Simoneau, J. Kulsuptrakul, M. Bouhaddou, K.A. Travisano, J. M. Hayashi, J. Carlson-Stevermer, J.R. Zengel, C.M. Richards, P. Fozouni, et al., Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses, *Cell* 184 (2021) 106–119, <https://doi.org/10.1016/j.cell.2020.12.004>.
- [90] Z. Daniloski, T.X. Jordan, H.-H. Wessels, D.A. Hoagland, S. Kasela, M. Legut, S. Maniatis, E.P. Mimitou, L. Lu, E. Geller, et al., Identification of required host factors for SARS-CoV-2 infection in human cells, *Cell* 184 (2021) 92–105, <https://doi.org/10.1016/j.cell.2020.10.030>.
- [91] C. Wei, L. Wan, Q. Yan, X. Wang, J. Zhang, X. Yang, Y. Zhang, C. Fan, D. Li, Y. Deng, et al., HDL-scavenger receptor B type 1 facilitates SARS-CoV-2 entry, *Nat. Metab.* 2 (2020) 1391–1400, <https://doi.org/10.1038/s42255-020-00324-0>.
- [92] R. Zang, J.B. Case, E. Yutuc, X. Ma, S. Shen, M.F.G. Castro, Z. Liu, Q. Zeng, H. Zhao, J. Son, et al., Cholesterol 25-hydroxylase suppresses SARS-CoV-2 replication by blocking membrane fusion, *Proc. Natl. Acad. Sci. USA* 117 (2020) 32105–32113, <https://doi.org/10.1073/pnas.2012197117>.
- [93] S. Wang, W. Li, H. Hui, S.K. Tiwari, Q. Zhang, B.A. Croker, S. Rawlings, D. Smith, A.F. Carlin, T.M. Rana, Cholesterol 25-hydroxylase inhibits SARS-CoV-2 and other coronaviruses by depleting membrane cholesterol, *EMBO J.* 39 (2020), e106057, <https://doi.org/10.15252/emboj.2020106057>.