



OPEN

Plasma metabolomics and gene regulatory networks analysis reveal the role of nonstructural SARS-CoV-2 viral proteins in metabolic dysregulation in COVID-19 patients

V. A. Ivanisenko^{1,2,3}✉, E. V. Gaisler¹, N. V. Basov¹, A. D. Rogachev^{1,4}, S. V. Cheresiz¹, T. V. Ivanisenko^{2,3}, P. S. Demenkov^{2,3}, E. L. Mishchenko², O. P. Khripko⁵, Yu. I. Khripko⁵, S. M. Voevoda^{5,6}, T. N. Karpenko⁷, A. J. Velichko⁷, M. I. Voevoda⁵, N. A. Kolchanov^{1,2} & A. G. Pokrovsky¹

Metabolomic analysis of blood plasma samples from COVID-19 patients is a promising approach allowing for the evaluation of disease progression. We performed the metabolomic analysis of plasma samples of 30 COVID-19 patients and the 19 controls using the high-performance liquid chromatography (HPLC) coupled with tandem mass spectrometric detection (LC-MS/MS). In our analysis, we identified 103 metabolites enriched in KEGG metabolic pathways such as amino acid metabolism and the biosynthesis of aminoacyl-tRNAs, which differed significantly between the COVID-19 patients and the controls. Using ANDSystem software, we performed the reconstruction of gene networks describing the potential genetic regulation of metabolic pathways perturbed in COVID-19 patients by SARS-CoV-2 proteins. The nonstructural proteins of SARS-CoV-2 (orf8 and nsp5) and structural protein E were involved in the greater number of regulatory pathways. The reconstructed gene networks suggest the hypotheses on the molecular mechanisms of virus-host interactions in COVID-19 pathology and provide a basis for the further experimental and computer studies of the regulation of metabolic pathways by SARS-CoV-2 proteins. Our metabolomic analysis suggests the need for nonstructural protein-based vaccines and the control strategy to reduce the disease progression of COVID-19.

Coronaviruses (CoVs) represent a group of enveloped viruses with exceptionally long (26–32 kb) single strand positive sense (+) RNA genome¹. Human coronaviruses primarily affect the respiratory tract, causing fever, cough, and, in severe cases, the shortness of breath and systemic inflammation, possibly, resulting in sepsis, cardiac insufficiency and polyorgan dysfunction in high-risk patients. The pediatric problem of common cold coronaviruses (ccCoVs) discussed in the publication² deserves a special interest. The virus causing COVID-19 (2019-nCoV or SARS-CoV-2) is more contagious, than previously identified human beta-coronaviruses (Severe Acute Respiratory Syndrome CoV/SARS-CoV and Middle East Respiratory Syndrome CoV/MERS-CoV). The

¹Novosibirsk State University, Pirogova Str., 2, 630090 Novosibirsk, Russia. ²Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences, Acad. Lavrentiev Ave., 10, 630090 Novosibirsk, Russia. ³Kurchatov Genomics Center of Institute of Cytology and Genetics of Siberian Branch of Russian Academy of Sciences, Acad. Lavrentiev Ave., 10, 630090 Novosibirsk, Russia. ⁴N. N. Vorozhtsov, Novosibirsk Institute of Organic Chemistry, Acad. Lavrentiev Ave., 9, 630090 Novosibirsk, Russia. ⁵Federal Research Center for Fundamental and Translational Medicine, Timakov Str., 2, 630117 Novosibirsk, Russia. ⁶Research Institute of Internal and Preventive Medicine—Branch of the Institute of Cytology and Genetics, Acad. Lavrentiev Ave., 10, 630090 Novosibirsk, Russia. ⁷State Budgetary Healthcare Institution of Novosibirsk Region 'City Clinical Hospital No. 11', Tankistov str., 23, 630120, Novosibirsk, Russia. ✉email: salix@bionet.nsc.ru

origin of new SARS-CoV-2 variants due to the virus genome mutability is associated with the acquisition of stronger contagiousity and the ability to escape the immune control, which accounts to the faster spread of the infection³.

The 5' regions of coronaviral genome include the cap, 5' untranslated region (5' UTR) and a long replicase gene comprising about 2/3 of the genome and coding for 16 non-structural proteins (nsp). The 3' genomic regions code for several structural proteins (spike (S), surface (E), membrane (M), and nucleocapsid (N) proteins) and a number of accessory proteins, which is variable in different coronaviruses. They also comprise the 3' -UTR and poly(A) tract⁴.

The replicase gene consists of two overlapping ORFs (ORF1a and ORF1b) coding for the components of viral replication and transcription complex (RTC). Replicase gene expression results in the translation of pp1a polyprotein from ORF1a, while the -1 ribosomal frameshift before the ORF1a translation termination codon causes a switch to ORF1b translation and production of pp1ab polyprotein representing a longer variant of pp1a. Cleavage by two viral cysteine proteases (papain-like PLpro and 3C-like 3Clpro) produces the intermediate precursor proteins and the highly conservative mature nonstructural protein, which can associate together to form the RTC⁴. The RTC is comprised of several enzymes, such as the nsp3 and nsp5 proteases, nsp7/nsp8 primase, nsp12 RNA-dependent RNA polymerase, nsp13 helicase/triphosphatase, nsp 14 exoribonuclease, nsp 15 endonuclease and nsp10/nsp16 N7- and 2'O-methyltransferase⁵. The 3' SARS-CoV genomic region codes for structural proteins and eight accessory proteins designated as ORF3a, 3b, 6, 7a, 7b, 8a, 8b and 9b. The SARS-CoV-2 3' genomic region codes for the same set of accessory proteins, with the exception of ORF3b⁶. Functions have been attributed to several SARS-CoV accessory proteins: ORF3a and ORF8a trigger the apoptosis, ORF7a activates NF- κ B cascade, ORF3b regulates the expression of several cytokines and chemokines, ORF6 down-regulates IFN production, while ORF8b induces the cellular DNA synthesis⁷.

The molecular mechanisms of viral infections and the related pathological processes are actively studied using metabolomic^{8,9}, proteomic¹⁰, and transcriptomic analyses¹¹.

Metabolomics measures the profiles of metabolites, such as amino acids, organic acids, bioamines, acylcarbinols, glycerophospholipids, sphingolipids, sugars and other compounds. Plasma metabolomics analyses of COVID-19 patients are represented in recent publications^{12–16}. The metabolomic profiles of critical COVID-19 patients of intensive care facilities were characterized by the dominating change of kynurenine and arginine content (and arginine to kynurenine ratio), sarcosine and lysophosphatidylcholines. The creatinine and arginine to kynurenine ratio were the perfectly accurate predictors of mortality⁹. The systematic metabolomic analysis of sera from COVID-19 patients revealed a significant decrease in > 100 metabolites, such as aminoacids and their derivatives, compared to the control subjects⁸.

An investigation into COVID-19-associated processes and pathways is important for the identification of diagnostic and prognostic biomarkers and improvement of COVID-19 treatments. The data on the interactions between SARS-CoV-2 and human proteins laid the background for the study of molecular mechanisms of virus-host relationship¹⁷. Identification of the involvement of particular viral proteins in signaling pathways, via which these proteins affect the host biological functions, is a promising approach to the search of pharmacological targets of the new generation antivirals. However, while interpreting the experimental data, the role of viral proteins is often disregarded.

An establishment of the links between the different omics data and the viral proteins requires the use of gene network reconstruction methods. Previously, we developed the ANDSystem software and information system for gene network reconstruction based on the information extracted from factual databases or obtained by text-mining of research publications^{18–20}. In particular, the reconstruction of pre-eclampsia associate²¹, identification of a novel tuberculosis susceptibility candidate²², reconstruction and analysis of the HCV interactome²³, the search for novel candidate genes important for asthma and hypertension comorbidity²⁴, the analysis of programmed cell death upon SARS-CoV-2 infection²⁵, etc., has been performed with the use of ANDSystem.

In this study, we performed metabolomic analysis of blood plasma from COVID-19 patients and the control group using the high-performance liquid chromatography coupled with tandem mass spectrometric detection (LC-MS/MS). The proposed approach to metabolomic analysis has been previously employed by ourselves for the study of metabolomic profiles of cerebrospinal fluid and blood plasma samples of high-grade glioma patients²⁶.

We identified 103 metabolites, which significantly differ between the compared groups. The highest level of significance was demonstrated by mevalonolactone (a cyclic form of mevalonic acid), which plays a key role in cholesterol biosynthesis. The content of this metabolite in the plasma of COVID-19 patients was higher than that in the control samples. The role of cholesterol biosynthesis in the mechanisms of COVID-19 pathologies including the cytokine storm drags the attention of many researchers²⁷.

The amino acid metabolism and the biosynthesis of aminoacyl-tRNAs represented KEGG metabolic pathways²⁸ enriched with metabolites, which significantly differ between the COVID-19 patients and the controls. Employment of ANDSystem enabled us to reconstruct the gene networks describing the potential molecular mechanisms, by which the viral proteins affect metabolic pathways perturbed in COVID-19 patients. Some viral proteins had multiple ways via which they can regulate different metabolic processes. Three viral proteins were involved in the greater number of regulatory pathways, namely, ORF8, E and nsp5. Noteworthy, the reconstructed gene networks only provide the hypothesis on the molecular mechanisms of virus-host relationship in COVID-19 pathology, which require further corroboration by experimental studies and computer simulation.

Materials and methods

Study subjects. Study subjects were enrolled at the State budgetary healthcare institution of Novosibirsk region 'City Clinical Hospital No. 11' and included patients with COVID-19 diagnosis confirmed by PCR and healthy controls. General patient cohort characteristics are shown in Table 1.

Group	Age		Gender	
	Confidence limits, lower	Confidence limits, upper	Male	Female
Control	66.25	75.92	9	10
COVID-19	62.92	73.08	16	14

Table 1. Patient characteristics in two cohorts showing matching in terms of age and gender.

The samples of blood plasma from the biobank subjects obtained in 2019 (before the COVID-19 epidemics in Russia) and stratified for sex and age were used as the controls. No exclusion criteria have been applied to the subjects of this cohort.

Compliance with ethical standards. The study was reviewed and all experimental protocols were approved by the Ethics Committee of Novosibirsk State University Zelman's Institute of Medicine and Psychology (Meeting Minutes of 02.11.2020). All procedures involving human participants were found to be compliant with the ethical standards of the institutional research committee and the 1964 Helsinki Declaration and its subsequent amendments or similar ethical standards. An informed consent form was completed and signed by every study subject.

Plasma collection. Plasma samples were collected from study subjects by venipuncture and blood collection into a vacutainer containing potassium EDTA as stabilizer. Blood cells were pelleted down by centrifugation, and plasma samples were aliquoted and kept frozen at $-80\text{ }^{\circ}\text{C}$ until further use.

Sample preparation. All samples were processed at the same time according to the protocol described by²⁹. Briefly, 100 μL of plasma sample were precipitated with 400 μL of cooled methanol and incubated overnight at $-80\text{ }^{\circ}\text{C}$ for protein precipitation. Then, samples were centrifuged at $+4\text{ }^{\circ}\text{C}$ and 16,000 G for 15 min. Supernatant was transferred into a new polypropylene tube and dried in a SpeedVac concentrator centrifuge (Thermo Fisher Scientific/Savant, Waltham, MA). Samples were reconstituted in 100 μL of water/acetonitrile (95:5) and subjected to a modified targeted metabolomics analysis with relative quantification. Each sample was analyzed in three replicates.

LC–MS/MS analysis. Samples were analyzed using a Shimadzu LC-20AD Prominence chromatograph (Shimadzu Corporation, Japan) equipped with SIL-20AC autosampler (Shimadzu Corporation, Japan) thermostated at $10\text{ }^{\circ}\text{C}$. Sample (10 μL) was injected onto a Prontosil 120–5-C18 AQ (2.1 \times 75 mm) (Econova LLC, Russia). The mobile phases consisted of HPLC buffer A (water containing 0.1% formic acid) and HPLC eluent B (100% acetonitrile), the flow rate during analysis was 0.25 mL/min. The HPLC elution gradient was as follows: from 0 to 3 min, the mobile phase B was decreased from 97% B to 85%; from 3 to 4 min, the percentage of solvent B was decreased from 85 to 30%; from 4 to 10 min, the mobile phase B was decreased to 2% and was kept at 2% for additional 4.5 min. At minute 14.5, solvent B was increased back to 97% and the column was equilibrated for additional 2.5 min at the flow rate of 0.5 mL/min.

Metabolites ($n = 289$) were analyzed in MRM mode. Data acquisition was performed on API 6500 QTRAP mass-spectrometer (AB SCIEX, USA) equipped with an electrospray ionization source operating in the positive/negative switch mode. The main mass spectrometric parameters were as follows. The IS (ion spray) voltages were set at 5500 V and -4500 V for positive and negative modes, respectively. The ion source temperature was set at $475\text{ }^{\circ}\text{C}$, CAD gas was set as “medium”, Gas1, Gas2 and curtain gas were 35, 35 and 30 psi, respectively. Declustering potential was at 93 V, entrance potential at 10 V, and collision cell exit potential at 20 V for positive and negative ion modes. In addition, the polarity switching (settling) time was set at 5 ms, and dwell time was 3 ms for each MRM transition. The precursor ion and fragment ion transitions, the metabolite names, dwell times, and the appropriate collision energies for both positive and negative ion modes were adapted from²⁹ with several metabolite transitions added by our group. The device was controlled and information collected using Analyst 1.6.3 software (AB SCIEX).

Statistical analysis. To get the significance of the difference between metabolite level means in COVID-19 patients and the controls, Welch's t-test implemented in the SciPy package v1.8.0 was used³⁰. The multiple hypothesis testing was performed using the Benjamini-Yekutieli procedure from the statsmodels Python package v0.13.2³¹ (<https://www.statsmodels.org/stable/index.html>).

ANDSystem. ANDSystem¹⁹ includes the global gene network defining the interactions between the molecular genetic objects constructed via automatized mining of research literature and data extraction from factological databases. Overall, the ANDSystem considers 13 types of objects (proteins, genes, metabolites, etc.) and 24 types of interactions (physical interactions, expression regulation, activity regulation, stability regulation, etc.). ANDVisio program module provides the graphical interface offering the user's access to the database, which enables one to search for the pathways in the global gene network using the templates. The templates represent the linear chain of objects and their interactions. Additionally, the objects can be defined by the concrete names/

Template name	Template description*
P ₁	Vp - PPI -> Kp
P ₂	Vp - PPI -> Hp - PPI -> Kp
P ₃	Vp - PPI -> Hp - Act/Stab/Pr/PPM/Tr -> Kp
P ₄	Vp - PPI -> Hp - Exp reg -> Kg - Exp -> Kp
P ₅	Vp - PPI -> Hp - Expreg-> Hg - Exp-> Hp - Expreg-> Kg - Exp-> Kp
P ₆	Vp - PPI -> Hp - Exp reg -> Hg - Exp -> Hp - Act/Stab/Pr/PPM/Tr -> Kp
P ₇	Vp - PPI -> Hp - Expreg-> Hg - Exp-> Hp - PPI -> Kp

Table 2. Templates for virus-host interaction pathways. Interactions: PPI protein–protein interactions, Act/Stab/Pr/PPM/Tr regulation of activity or stability, or proteolysis, or post translational modifications, or transport (release). Exp reg regulation of gene expression. Exp gene expression (protein production). *Objects: Vp SARS-CoV2 proteins, Kp KEGG metabolic pathway proteins, Kg KEGG metabolic pathway genes, Hp any host proteins involved in the interactions, Hg any host genes involved in the interactions.

identifiers or by an object type, only. In the first case, the pathway search will consider only the identified objects, otherwise, all the objects of a particular type will be considered.

In an attempt to identify the potential molecular mechanisms, via which the viral proteins could perturb metabolic pathways, we reconstructed the pathways describing different types of interactions between molecular objects using ANDSystem, which employs the rules described by special predesigned templates.

The program searches for those pathways in the global network, which meet the template description. In this work, we analyzed 7 such templates (Table 2). The first position in every template is occupied by SARS-CoV-2 proteins, the last one, the KEGG metabolic pathway proteins. Viral proteins can participate in protein–protein interactions with human counterparts, only. The length of the template chain was 2 to 5 objects. Only proteins and genes were considered as the objects here. The interactions included the protein–protein interactions, the regulatory interactions, and the links between the gene expression and the gene product, or similar ones.

Statistical significance of the virus-host signaling pathways. The statistical significance of the link between the virus-host signaling pathways and KEGG metabolic pathways²⁸ (p value) was estimated for each pathway template (Table 2) using hypergeometric distribution. For the calculations, all human proteins from KEGG database ($N = 1393$) were taken. The signaling pathways that link the viral proteins to the KEGG proteins from the list (N) were searched for with the use of templates. The use of a single template may have resulted in the identification of a set of signaling pathways. Based on the analysis of these sets, for each template P_i , the number of proteins n_i from the list (N), which represent the last object in, at least, a single signaling pathway from the set of pathways of a particular template, is calculated. Similarly, for each metabolic pathway j including K_j proteins, the number of proteins, k_{ij} , which are the targets of virus-host signaling pathways of template P_i , is calculated. The probability of observing k_{ij} number due to accidental causes is estimated with the standard hypergeometric distribution using hypergeom function of SciPy 1.8.0 package (<https://scipy.org>).

Results

The metabolite content in plasma samples of COVID-19 patient and control cohorts is shown in the Supplementary Data. The filtration of metabolites was first approached. Only those metabolites that showed the non-zero values in at least two of three repeats were used for the further analysis. Thus, only 140 metabolites out of 289 metabolites that can be detected by the employed method were selected. The comparison of mean metabolite content in plasma samples of COVID-19 patients and the controls showed the statistically significant differences (Benjamini-Yekutieli test, $p < 0.05$) for 103 out of 140 metabolites studied (Supplementary Table S1).

KEGG metabolic pathway overrepresentation analysis (Table 3) performed with MetaboAnalyst webservice³² enabled us to identify the pathways enriched with metabolites from the list of 103 significant ones (Supplementary Table S1). The numbers of significantly overrepresented pathways ($q < 0.05$) are 3 and 6, respectively, when Holm or FDR corrections for multiple comparisons are accounted for. All KEGG processes from this list were linked to the amino acid metabolism or the aminoacyl-tRNA biosynthesis.

We estimated the statistical significance of associations between the virus-host signaling pathways and KEGG metabolic processes (Table 4). Three overrepresented metabolic pathways, for which HOLM p value is < 0.05 (Table 3), are analyzed here.

As one can see from (Table 4), five out of seven templates are significant for arginine biosynthesis. The regulation of a particular KEGG pathway by viral proteins can occur via different signaling pathways and can involve different types of interactions. Conversely, only template P_5 is significant for glycine, serine and threonine metabolism pathway. The signaling pathways described by templates P_2 and P_7 can significantly contribute to the regulation of aminoacyl-tRNA biosynthesis. A typical feature of these pathways is that the last link in the chain of interactions is represented by protein–protein interaction involving KEGG enzyme. Such pathways affect the activity and stability of proteins via protein–protein interactions, but not the expression of KEGG pathway enzymes.

A detailed analysis of signaling pathways associated with each of KEGG metabolic processes is provided below.

Metabolite set	Total	Hits	Expect	<i>p</i> value	<i>q</i> value, Holm P	<i>q</i> value, FDR
Aminoacyl-tRNA biosynthesis	48	14	2.93	3.33E-7	2.8E-5	2.8E-5
Glycine, serine and threonine metabolism	33	10	2.02	1.29E-5	0.00107	5.41E-4
Arginine biosynthesis	14	6	0.855	8.95E-5	0.00734	0.00251
Valine, leucine and isoleucine biosynthesis	8	4	0.489	7.58E-4	0.0614	0.0159
Arginine and proline metabolism	38	8	2.32	0.00153	0.123	0.0258
Histidinemetabolism	16	5	0.978	0.00195	0.154	0.0273
Pantothenate and CoA biosynthesis	19	5	1.16	0.00449	0.35	0.0538
Alanine, aspartate and glutamate metabolism	28	6	1.71	0.00558	0.429	0.0586
Phenylalanine, tyrosine and tryptophan biosynthesis	4	2	0.244	0.0204	1.0	0.191
Purine metabolism	65	8	3.97	0.0403	1.0	0.29
Pentose phosphate pathway	22	4	1.34	0.0409	1.0	0.29
Glyoxylate and dicarboxylate metabolism	32	5	1.96	0.0414	1.0	0.29

Table 3. Analysis of KEGG metabolic pathway overrepresentation. Significant values ($q < 0.05$) are in bold.

Virus-host pathway template (P_i)	Aminoacyl-tRNA biosynthesis ($K_1 = 40$ proteins)		Glycine, serine and threonine metabolism ($K_2 = 34$ proteins)		Arginine biosynthesis ($K_3 = 18$ proteins)	
	k_{i1}	<i>P</i> value	k_{i2}	<i>P</i> value	k_{i3}	<i>P</i> value
P_1	3	0.092	0	–	0	–
P_2	19	0.00011	11	0.26	10	0.0075
P_3	3	0.23	0	–	2	0.99
P_4	0	–	1	0.69	3	0.009
P_5	5	0.981	10	0.049	10	0.004
P_6	4	0.712	3	0.78	5	0.024
P_7	19	0.0089	10	0.54	11	0.021

Table 4. Statistical significance of associations between virus-host signaling pathways and KEGG metabolic processes. Significant values ($p < 0.05$) are in bold. K_j is the number of proteins in KEGG metabolic pathway j , k_{ij} is the number of proteins, which represent the targets of virus-host pathway template P_i (Table 2).

Aminoacyl-tRNA biosynthesis. The first column in the table of KEGG metabolic pathway overrepresentation analysis guided by the identified set of metabolites, which significantly differ between COVID-19 and control groups, is occupied by aminoacyl-tRNA biosynthesis (Table 3). Overall, 48 metabolites are involved in the process, 14 of which are on the list of significantly differing metabolites (Supplementary Table S2).

Of them, ten metabolites are significantly increased in plasma samples of COVID-19 patients (L-Aspartic acid, L-Serine, L-Glutamic acid, etc.), while L-Asparagine, L-Tyrosine, L-Methionine and L-Leucine/L-Isoleucine contents are reduced there.

The reconstruction of signaling pathways potentially involved in the regulation of aminoacyl-tRNA biosynthesis by viral proteins was performed for the proteins localized to the mitochondria or the cytoplasm, separately (Supplementary Table S3). Of seven types of pathways, those described by P_2 and P_7 templates (Table 4) are significant. As previously described, these templates share a common feature. In them, the effect on the last target in the pathway is exerted by protein–protein interactions. P_2 pathways are shorter, since they include only a single intermediate. These pathways involve protein–protein interactions only.

Mitochondrial network comprising all signaling pathways identified with the use of P_2 template includes 28 proteins and 31 interactions (Fig. 1A). The network contains 7 viral proteins and 9 enzymes of aminoacyl-tRNA biosynthesis. The cytoplasmic P_2 pathways include 32 proteins, nine of which are viral proteins, while ten are the aminoacyl-tRNA biosynthesis enzymes (Fig. 1B). Merge of cytoplasmic and mitochondrial P_2 networks results in 53 protein members (11 viral proteins and 19 aminoacyl-tRNA biosynthesis enzymes) and increases the number of interactions to 57.

The merged P_7 network of mitochondrial and cytoplasmic signaling pathways is shown in Fig. 2. P_7 pathways include 3 intermediates (Table 2). The first intermediate is a protein interacting with viral protein. The expression of a gene, which represents the second intermediate, is regulated by intermediate 1. The third intermediate is the protein product of intermediate gene 2. At the end of pathway, intermediate 3 is involved in protein–protein interactions with a target protein of the considered KEGG metabolic process. As seen in Fig. 2, the P_7 pathway network is greater than the P_2 network. It includes 111 proteins (15 viral proteins and 19 aminoacyl-tRNA biosynthesis enzymes), as well as 23 genes. The network also involves 110 interactions, with 60 of them being

cytoplasm, the RNA exosome complex is known to be involved in general mRNA turnover due to its specific degradation of inherently unstable mRNAs containing AU-rich elements in 3' untranslated regions and in RNA surveillance pathways, due to prevention of aberrant mRNA translation³⁴.

According to Fig. 1A, the mitochondrial Leucine-tRNA ligase (SYLM) can be affected by viral proteins E and M. M protein can act onto SYLM via two pathways. One is mediated by MPPB (Mitochondrial-processing peptidase subunit beta), the other one by FAKD5 (FAST kinase domain-containing protein 5). MPPB cleaves the mitochondrial sequence off the newly imported precursor proteins³⁵. BRD2 (Bromodomain-containing protein 2) turns out to be an intermediate between E and SYLM. The data on MPPB, FAKD5 and BRD2 interactions with SYLM is contained in BioGrid (Id 2,857,559, Id 28,550,757 and Id 2,538,529, respectively).

Of note, all the discussed signaling pathways are only hypothetical and based on the integration of knowledge obtained from different experiments, thus, they require further corroboration.

Glycine, serine and threonine metabolism. Glycine, serine and threonine metabolism is the second top process of overrepresented KEGG processes (Table 3). Of 33 metabolites participating in this process, ten are significantly different between plasma samples of COVID-19 patients and the controls (Supplementary Table S4). Nine of the latter are increased in content, while glyceric acid is decreased ($\log_{FC} = -0.67$).

Reconstruction of signaling pathways describing the potential regulation of glycine, serine and threonine metabolism is performed with the use of P₅ template, which proves significant for this metabolic process (Table 4). P₅ template describes the pathway of potential metabolic gene expression regulation by viral protein and two human intermediate genes/proteins. Of 40 genes of KEGG process of glycine, serine and threonine metabolism, ten are the potential targets of viral proteins (Supplementary Table S5). In Fig. 4, nine viral proteins participate in gene expression regulation of this metabolic process. Particularly, orf8 protein can influence the expression regulation of ALDH7A1 coding for AL7A1 protein. This enzyme (EC 1.2.1.8) participates in betaine biosynthesis (Fig. 3). The potential signaling pathway initiated by orf8 is represented by the following chain of interactions in gene network. Orf8 interacts with SMOC1¹⁷, SMOC1 can activate BMP2 gene expression³⁶, while BMP2 inhibits ALDH7A1 expression³⁷. As another example of signaling pathway from gene network shown in Fig. 4, we can consider the expression regulation of PHGDH gene coding for SERA (D-3-phosphoglycerate dehydrogenase) by viral protein nsp5. This enzyme (EC 1.1.1.95) catalyzes the reversible oxidation of 3-phospho-D-glycerate to 3-phosphonoxy pyruvate, the first step of the phosphorylated L-serine biosynthesis pathway. According to the recent publication¹⁷, nsp5 interacts with histone deacetylase 2 (HDAC2). HDAC2 can, in turn, inhibit the expression of tumor suppressor p53^{38–40}. The last link in nsp5-SERA signaling pathway is represented by p53 and PHGDH interaction. The former is known to suppress PHGDH (SERA) expression and inhibit serine biosynthesis⁴¹.

Arginine biosynthesis. Arginine biosynthesis is the third top process among overrepresented KEGG processes (Table 3).

Of 23 metabolites participating in this process, six are significantly different between the plasma samples of COVID-19 patients and the controls (Supplementary Table S6). Five of them are increased in content, while ornithine is decreased ($\log_{FC} = -1.98$).

The potential regulation of arginine biosynthesis by viral proteins differs significantly from the two KEGG metabolic processes considered above (Supplementary Table S7). Of 22 genes involved in arginine biosynthesis, fourteen represent the potential targets of viral proteins (Fig. 5). Five types of signaling pathways potentially involved in the regulation prove statistically significant including pathways of P₂, P₄, P₅, P₆ and P₇ types (Table 4). The viral proteins can potentially regulate the expression of enzymes of this metabolic process or the expression of human proteins, which interact with these enzymes, or they can control the activity or stability of these enzymes via the mentioned types of pathways.

Gene network of expression regulation is shown in Fig. 6. It includes 74 genes and 132 proteins, nine of which are arginine biosynthesis enzymes, while 15 represent the viral proteins. The network also includes 194 links between proteins and genes, which describe expression regulation and 49 links defining the protein–protein interactions between viral and human counterparts.

Figure 7 shows an example of signaling pathways included in gene network in Fig. 6, which can be involved in arginase 2 expression regulation. ARG2 (EC:3.5.3.1) participates in L-ornithine and urea synthesis from L-arginine and may play a role in the regulation of extra-urea cycle of arginine metabolism (Fig. 5). As seen in Fig. 7, viral proteins E, nsp5, orf8, and orf3a can regulate ARG2 expression. In particular, nsp5 can form complexes with histone deacetylase 2 (HDAC2) protein¹⁷. In turn, HDAC2 can suppress arginase 2 expression⁴². nsp5 binding to HDAC2 can potentially affect HDAC2 function. However, further experimental studies and computer simulations are needed to clarify the functional role of these protein–protein interactions. In particular, interesting results can be obtained by the computer-assisted structural modeling of protein complexes.

An analysis of pathways related to the regulation of activity/stability of arginine biosynthesis enzymes reveals 6 viral proteins (E, nsp5, nsp14, orf3a, orf8 и orf9c) potentially involved in these pathways (Fig. 8). Six enzymes are associated with this type of regulation. NOS3 protein takes the central place in the network in this figure. Its activity can be potentially regulated by 4 viral proteins. Additionally, some of the latter can be involved in multiple ways, for example, orf8 and E participate in 3 and 2 pathways, respectively. Thus, one of the pathways involving orf8 is realized via its interaction with disintegrin and metalloproteinase domain-containing protein 9 (ADAM9), while another includes the interaction with protein-lysine 6-oxidase (LYOX). ADAM9 is known to increase vascular endothelial growth factor A (VEGFA) expression in lung cancer metastasis⁴³. In the gene network, this interaction between ADAM9 and VEGFA is presented as a potential one. According to one publication⁴⁴, LYOX

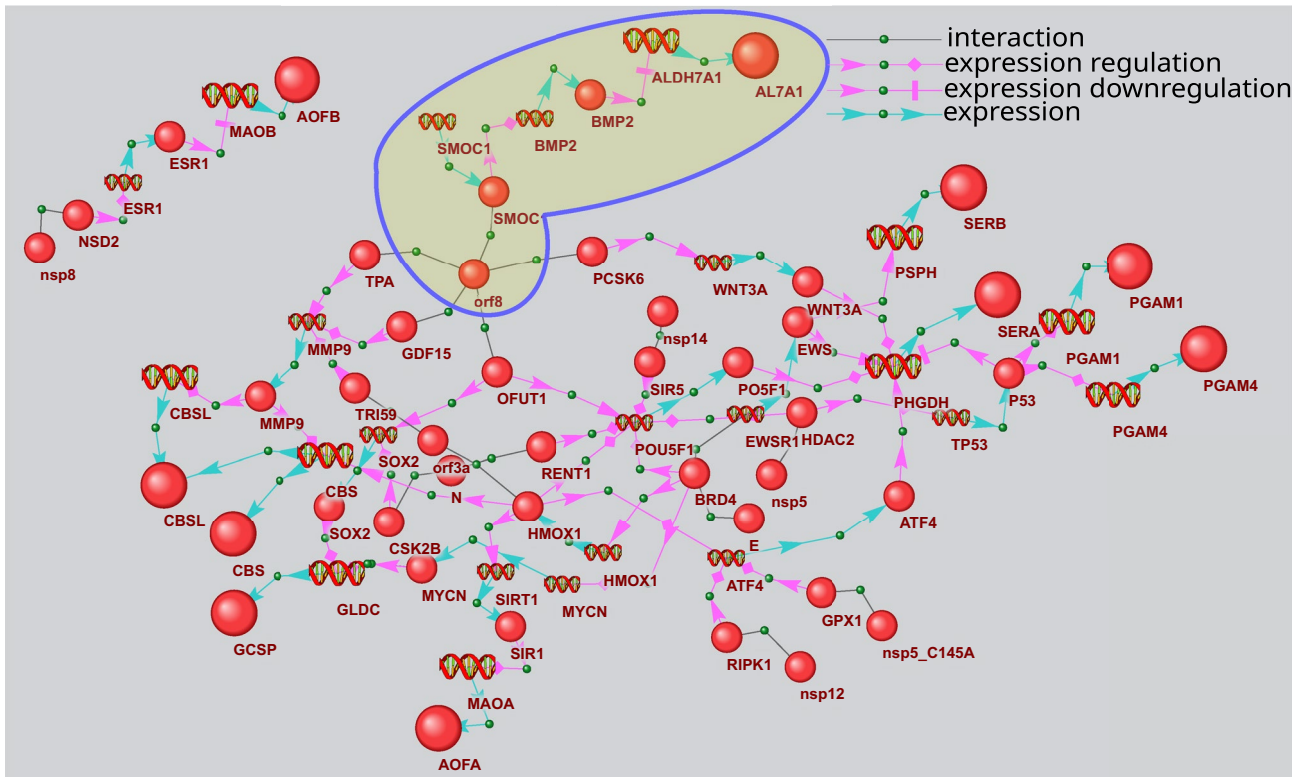


Figure 4. Gene network describing P₅ pathways of gene expression regulation of glycine, serine and threonine metabolism. The bigger balls show the proteins of glycine, serine and threonine metabolism, while the smaller ones designate other proteins. Spirals designate the genes. The pathways discussed in the text are outlined.

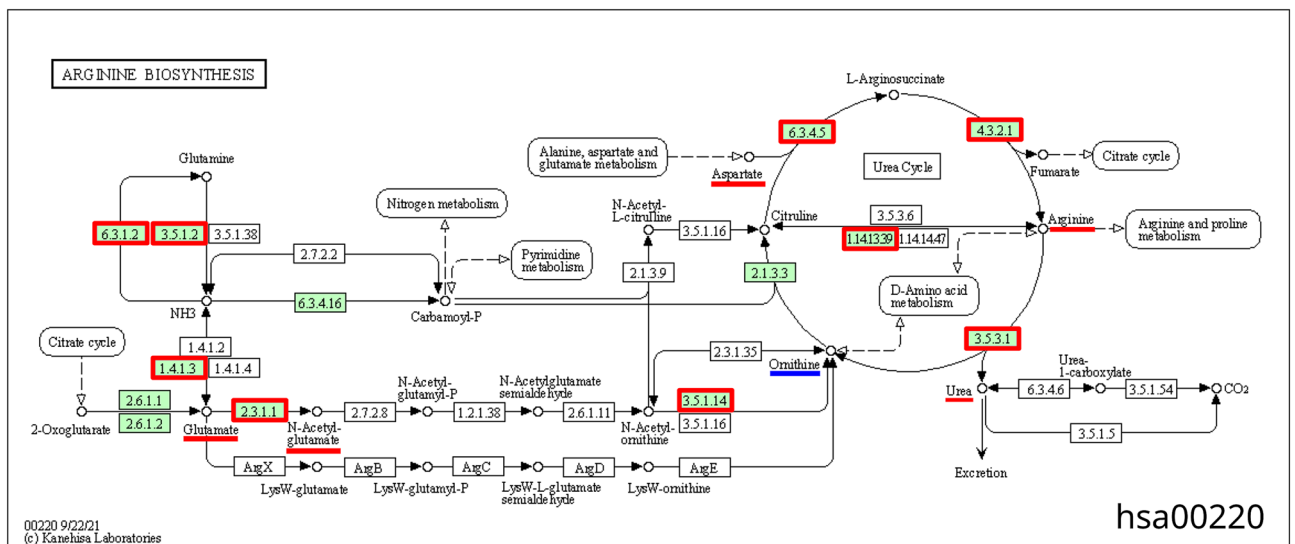


Figure 5. The scheme of arginine biosynthesis extracted from KEGG database²⁸ (Id hsa00220). Enzymes that can be the potential targets of viral proteins are shown by red boxes, the metabolites increased in the plasma of COVID-19 patients are underscored by red line, while those decreased are underscored by blue line.

Endothelin receptor type B (EDNRB) gene expression⁴⁶, while, according to HPRD database⁴⁷, EDNRB interacts with NOS3 (HPRD Id 01,224, HPRD Id 01,224).

Discussion and conclusions

Here, we report the results of metabolomic analysis of plasma samples of COVID-19 patients and the controls, which revealed 103 metabolites significantly differing between the two groups. According to the overrepresentation analysis of KEGG metabolic processes, the changes of metabolite content observed in the patients can be

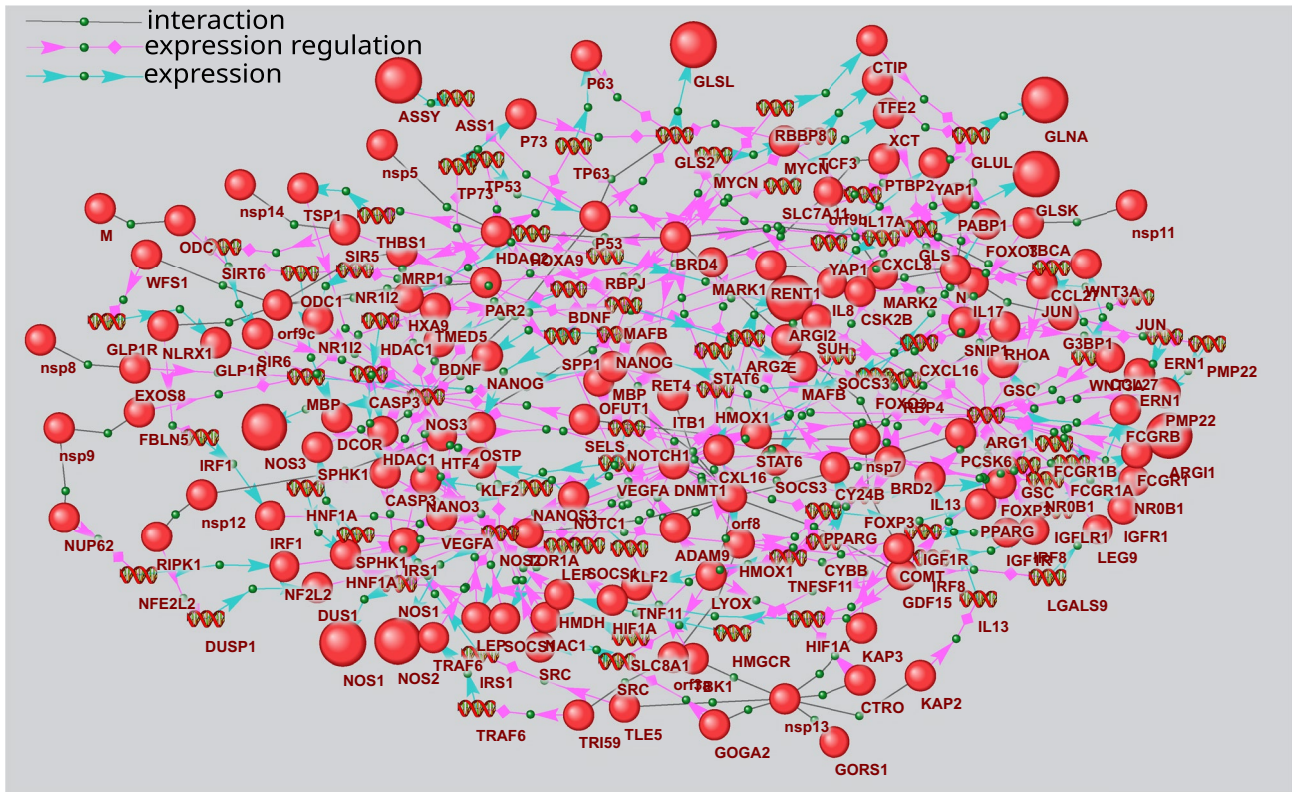


Figure 6. Gene network describing P₄ and P₅ pathways of arginine biosynthesis gene expression regulation. The bigger balls show the arginine biosynthesis proteins, while the smaller ones designate other proteins. Spirals designate the genes.

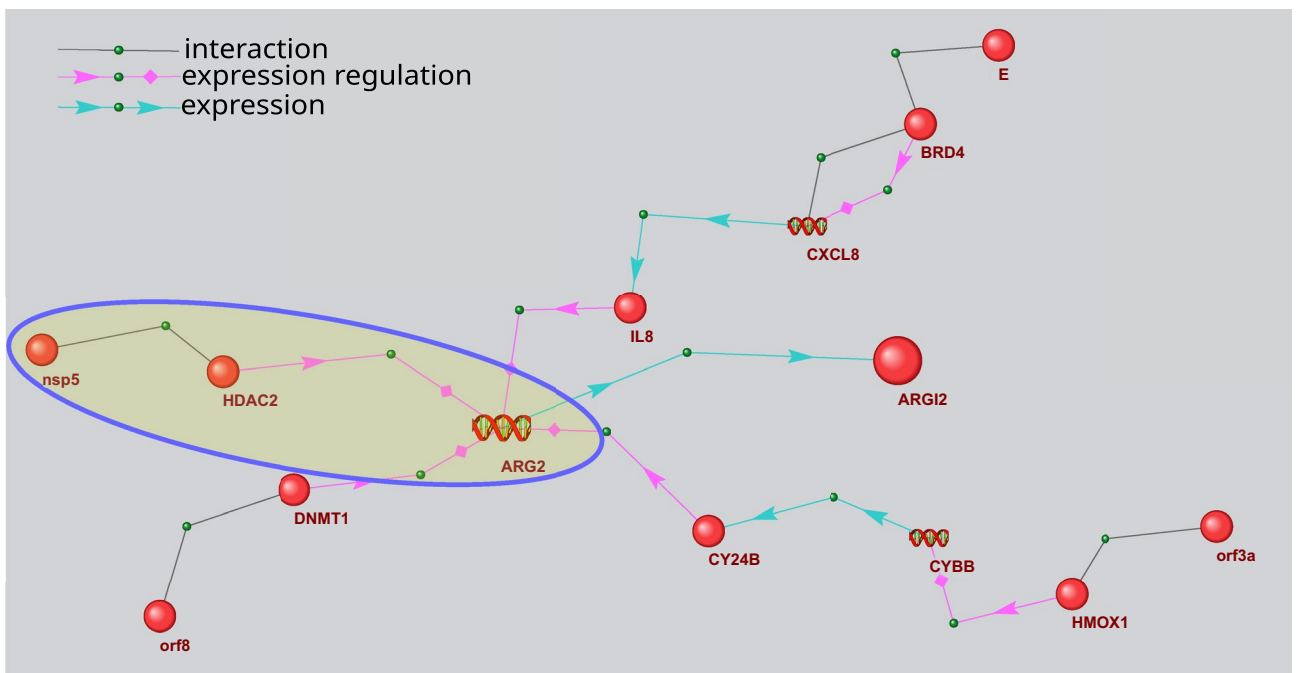


Figure 7. Pathways of arginase 2 (ARG2) expression regulation by viral proteins. The bigger ball designates ARG2, while the smaller ones designate other proteins. Spirals designate the genes. The pathways discussed in the text are outlined.

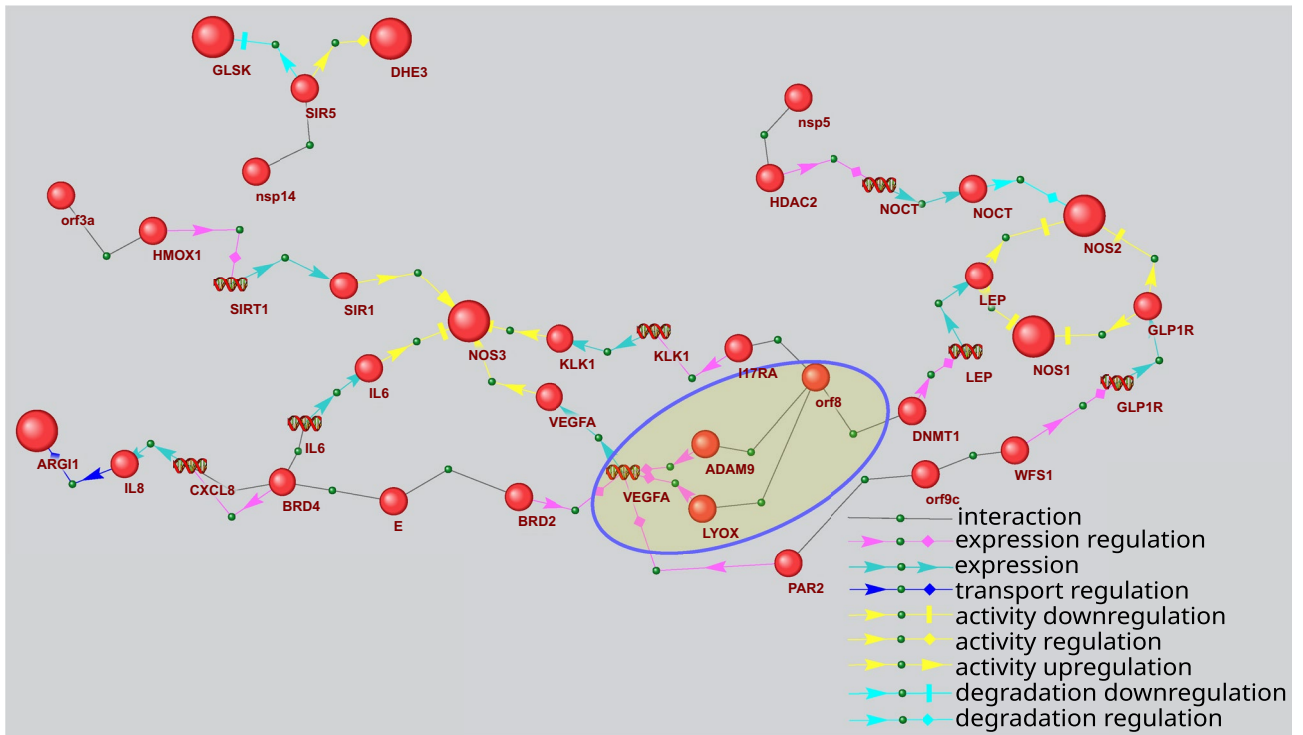


Figure 8. Gene network describing P_3 and P_6 pathways regulating activity/stability of arginine biosynthesis enzymes by viral proteins. The bigger balls show the proteins of arginine biosynthesis, while the smaller ones designate other proteins. Spirals designate the genes. The pathways discussed in the text are outlined.

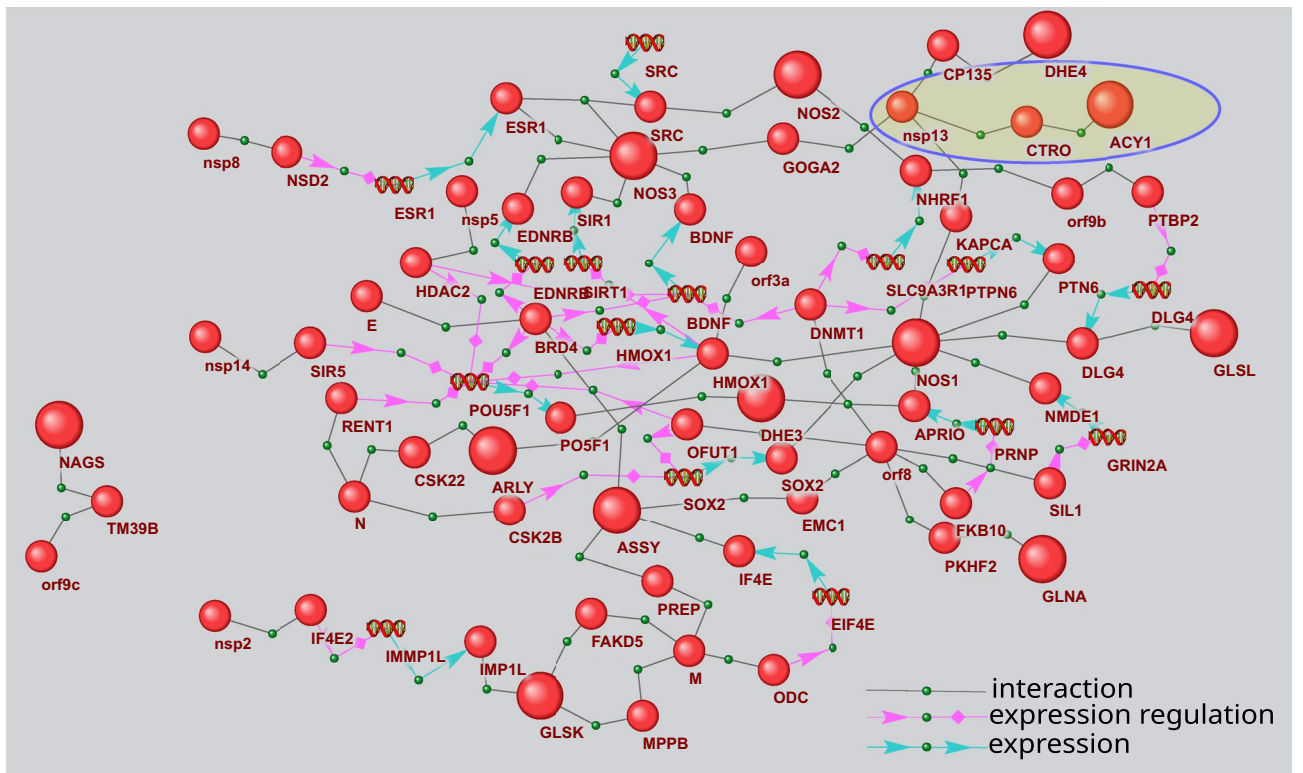


Figure 9. Gene network describing P_2 and P_7 pathways by which the viral proteins can influence the arginine biosynthesis proteins. The bigger balls show the proteins of arginine biosynthesis, while the smaller ones designate other proteins. Spirals designate the genes. The pathways discussed in the text are outlined.

closely related to amino acid metabolism including arginine biosynthesis, glycine, serine and threonine metabolism, and aminoacyl-tRNA biosynthesis. The obtained results well correspond to the reported metabolomic analysis data of COVID-19 patient plasma samples^{48–50}.

The metabolite most significantly differing between plasma samples of COVID-19 patients and the controls is mevalonolactone (Supplementary Table S1, Supplementary Note), a cyclic form of mevalonic acid, which plays a key role in cholesterol biosynthesis. Its content in plasma samples of COVID-19 patients was increased compared to that of the controls. The role of cholesterol biosynthesis in the mechanisms of COVID-19 pathology including the cytokine storm attracts keen attention of the researchers²⁷. Additionally, the increased cholesterol content could be associated with the higher rate of infection due to the proposed capacity of the virus to use lipid rafts for cellular entry⁵¹.

Noteworthy, the previously published results of metabolomic studies were, typically, limited to the analysis of metabolic profiles and metabolic processes only. An interpretation of these data in terms of molecular mechanisms and roles of viral proteins in the revealed disturbed function of metabolic processes is often overlooked.

We used ANDSys software, which enables one to reconstruct the gene networks from the data obtained by literature mining^{18–20}, to analyze the potential role of SARS-CoV-2 proteins in the disturbed function of metabolic processes.

The potential contribution of viral proteins in the regulation of four metabolic processes discussed above is summarized in Supplementary Table S8. As seen from this table, most of the viral proteins can potentially be involved in the regulation of the reported metabolic pathways. Surprisingly, spike protein is not listed among the potential contributors, since the regulatory pathways, via which it could affect metabolic processes, were not revealed. Viral proteins E, N, nsp5, nsp8 and orf8 were the potential top contributors to the regulation of metabolic processes. Each of them was related to all four metabolic pathways reported above.

The results of analysis of aminoacyl-tRNA biosynthesis regulation pathways were rather interesting. First, we discovered that cytoplasmic and mitochondrial pathways of aminoacyl-tRNA biosynthesis regulation by viral proteins are different. Similar enzymes of the mitochondrial and cytoplasmic aminoacyl-tRNA biosynthesis are typically affected by different viral proteins. Also, the regulatory pathways via which the virus exerts its effects are different. Moreover, the reconstructed gene networks show that viral proteins do not affect the expression of genes coding for aminoacyl-tRNA biosynthesis enzymes, but rather influence the protein–protein interactions of these enzymes with other human proteins. Such interactions are known to affect protein activity and stability. This suggestion on the mechanisms of modulation of aminoacyl-tRNA biosynthesis by viral proteins well agrees with the results of publication by Huang and collaborators¹¹, which focuses on the differential gene expression in SARS-CoV-2 infected pluripotent stem cell-derived human lung alveolar type 2 cells. This work demonstrates that LogFC of gene expression in the infected cells were not significantly higher or lower, than in non-infected cells (ranging from -0.51 to 0.11).

Unlike the aminoacyl-tRNA biosynthesis, the two others overrepresented KEGG metabolic processes (glycine, serine and threonine metabolism and arginine biosynthesis) can be affected by viral proteins at the level of expression regulation of the genes coding for the enzymes involved in these pathways.

The regulation of glycine, serine and threonine metabolism can occur via the control over gene expression only. When analyzing the regulatory pathways of different types, we discovered that these were exactly the expression regulation pathways that proved statistically significant.

Of 40 genes of glycine, serine and threonine metabolism, ten represent the potential targets of viral proteins (Supplementary Table S8), and 9 viral proteins can be involved in the expression regulation of these genes.

The feature of the potential regulation of arginine biosynthesis by the viral proteins is that this process can be controlled via gene expression regulation, or the regulation of protein activity and stability, or the protein–protein interactions. Nine genes of this pathway show the expression regulation by viral proteins, while 7 and 13 proteins are regulated at the levels of activity/stability and protein–protein interactions, respectively.

Importantly, the reconstructed gene networks represent an initial attempt to the identification of molecular mechanisms of virus–host interactions in COVID-19 patients. The obtained results suggest the promising directions of research into the interactions between viral and host proteins. Our analysis shows that such interactions can initiate the transmission of regulatory signals along the chains of links between genes and proteins, which can enhance or inhibit gene expression or the activity of key enzymes.

According to our analysis of the reconstructed gene networks, the protein–protein interactions between viral and human proteins including orf8/GDF15, N/MOV10, nsp5/HDAC2 and others will have the most significant effect on the metabolic pathway functions.

The significance of the reconstruction of molecular mechanisms of pathogen–host interactions, including the approaches involving metabolomic analysis, is dictated by the serious problem of emerging drug resistance of viruses due to the high mutation rates of their genomes. The pharmacological influence over the functions of host genes, which are exploited by the virus for the benefit of its life cycle, may result in the development of new generation therapies⁵². One can expect that drug resistance to such therapies will be considerably lower, than to those drugs, which target the viral enzymes and genomes.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 30 May 2022; Accepted: 11 November 2022

Published online: 20 November 2022

References

- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170 (2020).
- McIntosh, K. Proving etiologic relationships to disease. *Pediatr. Infect. Disease J.* **41**, e102–e103 (2021).
- Willett, B. J. *et al.* SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nat. Microbiol.* **7**, 1161–1179 (2022).
- Naqvi, A. A. T. *et al.* Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim. Biophys. Acta BBA Mol. Basis Disease* **1866**, 165878 (2020).
- Rohaim, M. A., El Naggar, R. F., Clayton, E. & Munir, M. Structural and functional insights into non-structural proteins of coronaviruses. *Microb. Pathog.* **150**, 104641 (2021).
- Singh, K. K., Chaubey, G., Chen, J. Y. & Suravajhala, P. Decoding SARS-CoV-2 hijacking of host mitochondria in COVID-19 pathogenesis. *Am. J. Physiol. Cell Physiol.* **319**, C258–C267 (2020).
- McBride, R. & Fielding, B. The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. *Viruses* **4**, 2902–2923 (2012).
- Shen, B. *et al.* Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* **182**, 59–72.e15 (2020).
- Fraser, D. D. *et al.* Metabolomics profiling of critically ill coronavirus disease 2019 patients: Identification of diagnostic and prognostic biomarkers. *Crit. Care Explor.* **2**, e0272 (2020).
- Hekman, R. M. *et al.* Actionable cytopathogenic host responses of human alveolar type 2 cells to SARS-CoV-2. *Mol. Cell* **81**, 212 (2021).
- Huang, J. *et al.* SARS-CoV-2 infection of pluripotent stem cell-derived human lung alveolar type 2 cells elicits a rapid epithelial-intrinsic inflammatory response. *Cell Stem Cell* **27**, 962–973.e7 (2020).
- Lee, J. W. *et al.* Integrated analysis of plasma and single immune cells uncovers metabolic changes in individuals with COVID-19. *Nat. Biotechnol.* **40**, 110–120 (2021).
- Danlos, F.-X. *et al.* Metabolomic analyses of COVID-19 patients unravel stage-dependent and prognostic biomarkers. *Cell Death Disease* **12**, 1–11 (2021).
- Sindelar, M. *et al.* Longitudinal metabolomics of human plasma reveals prognostic markers of COVID-19 disease severity. *Cell Rep. Med.* **2**, 100369 (2021).
- Valdés, A. *et al.* Metabolomics study of COVID-19 patients in four different clinical stages. *Sci. Rep.* **12**, 1–11 (2022).
- Yang, Z. *et al.* Plasma metabolome and cytokine profile reveal glycyproline modulating antibody fading in convalescent COVID-19 patients. *Proc. National Acad. Sci.* **119**, 2117089119 (2022).
- Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
- Ivanisenko, V. A. *et al.* ANDSystem: An associative network Discovery system for automated literature mining in the field of biology. *BMC Syst. Biol.* **9**, S2 (2015).
- Ivanisenko, V. A., Demenkov, P. S., Ivanisenko, T. V., Mishchenko, E. L. & Saik, O. V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinform* **20**, 5–15 (2019).
- Ivanisenko, T. V. *et al.* ANDDigest: A new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinform.* **21**, 1–12 (2020).
- Glotov, A. S. *et al.* Molecular association of pathogenetic contributors to pre-eclampsia (pre-eclampsia associome). *BMC Syst. Biol.* **9**, 1–12 (2015).
- Bragina, E.Yu., Tiys, E. S., Rudko, A. A., Ivanisenko, V. A. & Freidin, M. B. Novel tuberculosis susceptibility candidate genes revealed by the reconstruction and analysis of associative networks. *Infect. Genet. Evol.* **46**, 118–123 (2016).
- Saik, O. V., Ivanisenko, T. V., Demenkov, P. S. & Ivanisenko, V. A. Interactome of the hepatitis C virus: Literature mining with ANDSystem. *Virus Res.* **218**, 40–48 (2016).
- Saik, O. V. *et al.* Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. *BMC Med. Genomics* **11**, 61–76 (2018).
- Ivanisenko, N. V., Seyrek, K., Kolchanov, N. A., Ivanisenko, V. A. & Lavrik, I. N. The role of death domain proteins in host response upon SARS-CoV-2 infection: Modulation of programmed cell death and translational applications. *Cell Death Discov.* **6**, 1–10 (2020).
- Rogachev, A. D. *et al.* Correlation of metabolic profiles of plasma and cerebrospinal fluid of high-grade glioma patients. *Metabolites* **11**, 133 (2021).
- Lee, W. *et al.* COVID-19-activated SREBP2 disturbs cholesterol biosynthesis and leads to cytokine storm. *Signal Transduct. Target. Ther.* **5**(1), 1–11 (2020).
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
- Yuan, M., Breitkopf, S. B., Yang, X. & Asara, J. M. A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nat. Protoc.* **7**, 872–881 (2012).
- Virtanen, P. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Meth.* **17**, 261–272 (2020).
- Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference 2010* **57**, 61 (2010)
- Chong, J., Wishart, D. S. & Xia, J. Using metaboanalyst 4.0 for comprehensive and Integrative metabolomics data analysis. *Current Protoc. Bioinform.* **68**, e86 (2019).
- Tempio, T. & Anelli, T. The pivotal role of ERp44 in patrolling protein secretion. *J. Cell Sci.* **133**(21), 240336 (2020).
- Mukherjee, D. The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J.* **21**, 165–174 (2002).
- Greene, M. W. Circadian rhythms and tumor growth. *Cancer Lett.* **318**, 115–123 (2012).
- Wang, Y. & Wu, X. SMOC1 silencing suppresses the angiotensinII induced myocardial fibrosis of mouse myocardial fibroblasts via affecting the BMP2/Smad pathway. *Oncol. Lett.* <https://doi.org/10.3892/ol.2018.8989> (2018).
- van den Hoogen, C. *et al.* The aldehyde dehydrogenase enzyme 7A1 is functionally involved in prostate cancer bone metastasis. *Clin. Exp. Metas.* **28**, 615–625 (2011).
- Fritsche, P. *et al.* HDAC2 mediates therapeutic resistance of pancreatic cancer cells via the BH3-only protein NOXA. *Gut* **58**, 1399–1409 (2009).
- Witt, O., Deubzer, H. E., Milde, T. & Oehme, I. HDAC family: What are the cancer relevant targets?. *Cancer Lett.* **277**, 8–21 (2009).
- Li, H. *et al.* Clinical significance of HDAC1, -2 and -3 expression levels in esophageal squamous cell carcinoma. *Exp. Ther. Med.* **20**, 315–324 (2020).
- Simabuco, F. M. *et al.* p53 and metabolism: From mechanism to therapeutics. *Oncotarget* **9**, 23780–23823 (2018).
- Chen, X. *et al.* Histone deacetylases (HDACs) and atherosclerosis: A mechanistic and pharmacological review. *Front. Cell Dev. Biol.* **8**, 581015 (2020).
- Lin, C.-Y. *et al.* ADAM9 promotes lung cancer metastases to brain by a plasminogen activator-based pathway. *Can. Res.* **74**, 5229–5243 (2014).

44. Baker, A.-M. *et al.* Lysyl oxidase plays a critical role in endothelial cell stimulation to drive tumor angiogenesis. *Can. Res.* **73**, 583–594 (2012).
45. Musicki, B., Palese, M. A., Crone, J. K. & Burnett, A. L. Phosphorylated endothelial nitric oxide synthase mediates vascular endothelial growth factor-induced penile erection. *Biol. Reprod.* **70**, 282–289 (2004).
46. Wilkinson, M. *et al.* The molecular effects of a high fat diet on endometrial tumour biology. *Life* **10**, 188 (2020).
47. Prasad, T. S. K., Kandasamy, K. & Pandey, A. Human protein reference database and human proteinpedia as discovery tools for systems biology. *Meth. Mol. Biol.* **577**, 67–79. https://doi.org/10.1007/978-1-60761-232-2_6 (2009).
48. Masoodi, M. *et al.* Disturbed lipid and amino acid metabolisms in COVID-19 patients. *J. Mol. Med.* **100**, 555–568 (2022).
49. Pang, Z., Zhou, G., Chong, J. & Xia, J. Comprehensive meta-analysis of COVID-19 global metabolomics datasets. *Metabolites* **11**, 44 (2021).
50. Wu, D. *et al.* Plasma metabolomic and lipidomic alterations associated with COVID-19. *Natl. Sci. Rev.* **7**, 1157–1168 (2020).
51. Seyran, M. *et al.* The structural basis of accelerated host cell entry by SARS-CoV-2. *FEBS J.* **288**, 5010–5020 (2020).
52. Seyran, M. Artificial intelligence and clinical data suggest the T cell-mediated SARS-CoV-2 nonstructural protein intranasal vaccines for global COVID-19 immunity. *Vaccine* **40**, 4296–4300 (2022).

Acknowledgements

The bioinformatics analysis was supported by the budget project “Systems biology and bioinformatics: reconstruction, analysis and modeling of the structural and functional organization and evolution of human, animal, plant and microorganism gene networks” (No. FWNR-2022-0020). Plasma sample collection and processing were supported by RFBR grant 20-04-60314. Metabolomic analyses were supported by the state contract FSUS-2020-0035.

Author contributions

All authors participated in the study. I.V.A. and S.V.C. drafted the manuscript, and the other authors provided corrections and comments regarding the text. The final version of the paper was read and approved by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-24170-0>.

Correspondence and requests for materials should be addressed to V.A.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022