Serial co-expression analysis of host factors from SARS-CoV viruses highly converges with former high-throughput screenings and proposes key regulators

Antonio J. Pérez-Pulido, Gualberto Asencio-Cortés, Ana M. Brokate-Llanos, Gloria Brea-Calvo, Rosario Rodríguez-Griñolo, Andrés Garzón and Manuel J. Muñoz

Abstract

The current genomics era is bringing an unprecedented growth in the amount of gene expression data, only comparable to the exponential growth of sequences in databases during the last decades. This data allow the design of secondary analyses that take advantage of this information to create new knowledge. One of these feasible analyses is the evaluation of the expression level for a gene through a series of different conditions or cell types. Based on this idea, we have developed Automatic and Serial Analysis of CO-expression, which performs expression profiles for a given gene along hundreds of heterogeneous and normalized transcriptomics experiments and discover other genes that show either a similar or an inverse behavior. It might help to discover co-regulated genes, and common transcriptional regulators in any biological model. The present severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic is an opportunity to test this novel approach due to the wealth of data that are being generated, which could be used for validating results. Thus, we have identified 35 host factors in the literature putatively involved in the infectious cycle of SARS-CoV viruses and searched for genes tightly co-expressed with them. We have found 1899 co-expressed genes whose assigned functions are strongly related to viral cycles. Moreover, this set of genes heavily overlaps with those identified by former laboratory
high-throughput screenings (with P-value near 0). Our results reveal a series of common regulators, involved in immune and inflammatory responses that might be key virus targets to induce the coordinated expression of SARS-CoV-2 host factors.

**Key words:** SARS-CoV-2; SARS-CoV; coronavirus; reverse engineering; co-expressed genes; co-regulated genes

---

### Introduction

Genes are team players that rarely act in solitude but require the cooperation of others to carry out their physiological functions. Groups of genes are usually expressed together when necessary, under the conduction of regulatory proteins, conforming the so-called regulatory networks. In higher organisms, augmented complexity requires an increase in the number of biological functions. This is mainly achieved by expanding regulatory relationships rather than the number of participating genes [1, 2]. The combinatorial or regulatory activity of proteins on structural genes seem to account for the genetic plasticity required for development, organ functional diversity, responses to environmental changes and other emergent properties of multicellularity [3, 4]. One gene might play different roles in different regulatory contexts, but it will always be accompanied by other genes involved in that function. We hypothesize that, even under this complex scenario, this regulatory linkage among proteins sharing biological functions and pathways can be brought to light by carefully analyzing co-expression profiles from functional genomics and transcriptomic experiments.

The genomics era is generating a wealth of information about gene expression in many different biological processes and experimental approaches. Although designed to address specific questions, genomic expression data (as those from microarrays or RNA-Seq experiments) can provide a huge amount of information regarding regulatory relationships, usable to address different unrelated problems. One of the main obstacles encountered when analyzing data from different experimental sources is standardization. Lately, curated databases have been deployed to gather and normalize genomics experimental data. Expression Atlas is one of them [5]. As of July 2020, it holds normalized data from 1403 human transcriptomic experiments. These data contain valuable information on regulatory relationships that can be exploited to gain insights into any human biological system.

With all this in mind, we have devised a bioinformatics method based on an algorithm called Automatic and Serial Analysis of Co-expression (ASACO) that analyses information of multiple human transcriptomic experiments from Expression Atlas to predict novel regulators and functional partners of a given function. To challenge this algorithm, we have addressed the analysis of several known human proteins that are important to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection cycle and predicted new possible partners and regulators of these functions. The rationale behind this choice is the amount of experimental high-throughput genomic data, which is being produced because of the interest generated by the coronavirus disease 2019 (COVID-19) pandemics, providing an excellent opportunity to benchmark our in silico output. Moreover, we must also value the possibility of generating new knowledge on SARS-CoV-2 infection process. The virus infection cycle rests on the host’s physiological functions to take place, requiring the sequestration or co-option of multiple host factors. In the first place, it is clearly stated the general inhibition activity exerted by SARS viruses on host’s messenger RNA (mRNAs) translation [6, 7], but also other aspects of the host physiology are used or modified by the virus [8, 9]. The infection cycle heavily relies on tight interactions with the host endomembrane organelles, hijacking several host pathways [8, 10–15].

Understanding the basis of interactions of these pathways is crucial to fight infection. Translatomics and proteomics analysis of the response to viral infection have recently revealed cellular functions possibly involved in the infection process [16]. In the same way, interactomics using viral proteins as baits to find human proteins physically interacting with them, have also revealed possible new therapeutic targets [17]. Even more complex network analysis combining protein–protein interactions and transcriptomics pointed to new potential targets as well [18]. It is also noteworthy a study made on SARS-CoV knocking down kinases by sRNAi treatment. This study highlighted cellular kinases that are important to SARS-CoV virus infection and possibly crucial for SARS-CoV-2 as well [19]. Notably, not only the cellular but also the systemic response to the virus infection is important, and it seems to play a relevant role in SARS-CoV-2 infection as in many other viral cycles. Evidence is arising suggesting that SARS viruses are able to exploit the organism’s innate immune response on its own benefit, co-opting some components of the response, as stress granules and processing bodies [20–22], and recruiting as host factors those genes co-expressed with the innate immune response [23–25]. This evidence stresses the necessity of research to a broader systemic landscape to explain, for instance, SARS-CoV-2 influence on the inflammatory response [26–28].

To help identifying functional companions in regulatory networks for a given gene we have refined our algorithm to search for genes co-expressed with that initial gene along many unrelated transcriptomic experiments. We have applied this analysis to 35 human genes experimentally shown to play a role in SARS-CoV and/or SARS-CoV-2 infections. Here, we present the results of this test, compare with those from several experimental high-throughput genomic analysis and suggest some new functions possibly involved in SARS-CoV-2 biology.

### Materials and methods

#### Seed collecting

To collect human genes involved in the infection cycle of SARS-CoV-2 and SARS-CoV, we searched in PubMed database for articles until 29 March 2020 using the name of both viruses. Articles demonstrating host factors involved in specific steps of the infection were manually reviewed, and the viral activity and type of function was collected. Finally, both the gene name and the UniProtKB entry identifier were obtained for everyone. We called these genes as seeds.

#### Expression data collecting

We used the seed gene name in the Expression Atlas database to download human experiments where the seed was differentially expressed. This database provides 1352 standardized human transcriptomics experiments with a total of 3744 comparisons of different biological conditions coming from published results [5]. Then, we used a program written in Python language to get...
the expression matrices from all the considered experiments, including both microarray and RNA-sequencing data experiments. These matrices contain the logarithm in base 2 of the fold-change value (log2FC) for each gene within the experiments. We took the log2FC from all the genes with this value higher than 1 or lower than −1, and P-value lower than 0.05, in at least one of the collected experiments, and created a matrix of log2FC with the genes and the experiments. Finally, the Pearson correlation coefficient was independently calculated with the expression profile, for all the experiments, between the seed and each of the other genes.

ASACO algorithm

The ASACO algorithm involves a methodology to select genes that share similar behavior in terms of expression with the seed. The algorithm is based on fold change signs, as it is described as follows. The procedure begins with the matrix of fold change values extracted from the experiments of the Expression Atlas database in which the seed appears with a fold-change value. The ASACO algorithm is written in R language and available at https://github.com/UPOBioinfo/asaco

As metrics were defined, $P_i + N_i + Z_i = 1$ for $1 \leq i \leq n$. Then, a p-value $P(P_i)$ was computed for each $P_i$ from the sample distribution of $P_i$. Similarly, a p-value $P(N_i)$ was computed for each $N_i$ from the distribution of $N_i$.

The $P$-value of $P_1$, $P(P_1)$, is the probability that there are values greater than or equal to $P_1$ in the distribution of $P$, given the sample obtained from the experiment database. Analogously the $P$-value of $N_1$, $P(N_1)$, is the probability that there are values greater than or equal to $N_1$ in the distribution of $N$.

The p-values of $P_i$ and $N_i (1 \leq i \leq n)$ were computed by using the empirical cumulative distribution function, as it is defined in Equation 4 for $P_i$ (analogously for $N_i$).

$$P(P_i) = 1 - \frac{1}{n} \sum_{j = 1, j \neq i}^{n} (P_j \leq P_i)$$

Finally, selected genes were divided into two groups: a) those that are directly correlated in terms of the sign of their fold changes, and b) those that are inversely correlated.

The first group is defined as $D = \{G_k\}$ for each $k$ such that $P(P_k) \leq .01$ and $1 \leq k \leq |D|$, where $|D|$is the number of selected directly correlated genes. Analogously, the second group of selected genes is defined as $I = \{G_k\}$ for each $k$ such that $P(N_k) \leq .01$ and $1 \leq k \leq |I|$, where $|I|$is the number of selected inversely correlated genes.

In this way, selected directly correlated genes are those of which fold-change signs are mostly the same than the seed in the same experiments, since their value of $P$ is significantly high and, therefore, the probability of finding genes with a higher value of $P$ is very low ($\leq .01$). Analogously, selected inversely correlated genes are those of which fold change have mostly the opposite sign than the seed in the same experiments, since their value of $N$ is significantly high and, therefore, the probability of finding genes with a higher value of $N$ is very low.

The ASACO algorithm is written in R language and available at https://github.com/UPOBioinfo/asaco

Functional annotation and pathway analysis

Functional annotation for both seeds and co-expressed genes were obtained from Biomart [29]. The functional enrichment were made with KEGG Pathway [30], and Reactome [31], using the R libraries bioamR, clusterProfiler and ReactomePA, and a P-value cutoff of 0.05.

To find pathways with a significant averaged correlation with the seeds, all human genes were grouped by the Reactome pathway where they belong. A Wilcoxon test was calculated by each pathway and the P-value was adjusted by the FDR method. Finally, only the pathways with p-value equal or lower than 1e-05, and a median correlation higher than the third quartile of the distribution of all the analyzed genes were considered as significant pathways.

To discover transcription factors into the gene datasets, the gene ontology term ‘DNA-binding transcription factor activity’, together with those of ‘regulation of DNA-binding transcription factor activity’ were searched (GO:0003700, GO:0051090, GO:0051091, GO:0043433).
Table 1. Gene targets that are expected to be essential for the infectious cycle of SARS-CoV-2. Human genes used for the ASACO analysis. Their probable activity and the type of function observed in the tested virus are indicated.

| Gene name | UniProt Entry | Viral activity | Type of function | Virus | Reference |
|-----------|---------------|----------------|------------------|-------|-----------|
| ABL2      | P42684        | Entry           | Proviral         | SARS-CoV | (Coleman et al., 2016) |
| ACE2      | Q98YF1        | Entry           | Proviral         | SARS-CoV-2 | (Zhou et al., 2020) |
| BSG       | P35613        | Entry           | Proviral         | SARS-CoV | (Wang et al., 2020) |
| CTSL      | P07711        | Entry           | Proviral         | SARS-CoV-2 | (Cu et al., 2020) |
| FURIN     | P09958        | Entry           | Proviral         | SARS-CoV-2 | (Coutard et al., 2020) |
| NRP1      | O14786        | Entry           | Proviral         | SARS-CoV-2 | (Daly et al., 2020) |
| PKFVYE    | Q9Y217        | Entry           | Proviral         | SARS-CoV-2 | (Cu et al., 2020) |
| TMPRSS2   | O15393        | Entry           | Proviral         | SARS-CoV-2 | (Hoffmann et al., 2020) |
| TPCN2     | Q8NHX9        | Entry           | Proviral         | SARS-CoV-2 | (Cu et al., 2020) |
| CDK6      | Q00534        | Replication     | Antiviral        | SARS-CoV | (de Wilde et al., 2015) |
| DDX1      | Q92499        | Replication     | Antiviral        | SARS-CoV | (Wu et al., 2014) |
| DDX5      | P17844        | Replication     | Antiviral        | SARS-CoV | (Chen et al., 2009b) |
| EIF2AK2   | P19525        | Replication, Translation | Antiviral | SARS-CoV | (de Wilde et al., 2015) |
| EZK       | P15311        | Replication     | Antiviral        | SARS-CoV | (Millet et al., 2012) |
| GSK3A     | P49840        | Replication     | Proviral         | SARS-CoV | (Wu et al., 2009) |
| GSK3B     | P49841        | Replication     | Proviral         | SARS-CoV | (Wu et al., 2009) |
| HNRNPA1   | P09651        | Replication, Transcription | Proviral | SARS-CoV | (Luo et al., 2005) |
| IMPDH1    | P20839        | Replication     | Proviral         | SARS-CoV-2 | (Saijo et al., 2009) |
| IMPDH2    | P12268        | Replication     | Proviral         | SARS-CoV-2 | (Saijo et al., 2009) |
| PPIA      | P62937        | Replication     | Proviral         | SARS-CoV | (Fefferle et al., 2011) |
| PPPB      | P23284        | Replication     | Proviral         | SARS-CoV | (Fefferle et al., 2011) |
| PPPG      | Q13427        | Replication     | Proviral         | SARS-CoV | (Fefferle et al., 2011) |
| PPPIH     | O43447        | Replication     | Proviral         | SARS-CoV | (Fefferle et al., 2011) |
| TOP3B     | O95985        | Replication     | Proviral         | SARS-CoV-2 | (Prasanth et al., 2020) |
| ZCRB1     | Q8TTBF4       | Replication     | Proviral         | SARS-CoV | (Tan et al., 2012) |
| BST2      | Q10588        | Vesicle fusion  | Antiviral        | SARS-CoV | (Taylor et al., 2015) |
| COPB2     | P35606        | Vesicle fusion  | Proviral         | SARS-CoV | (de Wilde et al., 2015) |
| COPB1     | P35618        | Vesicle fusion  | Proviral         | SARS-CoV | (de Wilde et al., 2015) |
| GBF1      | Q92538        | Vesicle fusion  | Proviral         | SARS-CoV | (de Wilde et al., 2015) |
| IFITM1    | P13164        | Vesicle fusion  | Antiviral        | SARS-CoV | (Huang et al., 2011) |
| IFITM2    | Q01629        | Vesicle fusion  | Antiviral        | SARS-CoV | (Huang et al., 2011) |
| IFITM3    | Q01628        | Vesicle fusion  | Antiviral        | SARS-CoV | (Huang et al., 2011) |
| OSPB      | P22059        | Vesicle fusion  | Proviral         | SARS-CoV | (Amini-Bavil-Olyaee et al., 2013) |
| PRKCI     | P41743        | Vesicle fusion  | Proviral         | SARS-CoV | (de Wilde et al., 2015) |
| VAPA      | Q8P0L0        | Vesicle fusion  | Proviral         | SARS-CoV | (Amini-Bavil-Olyaee et al., 2013) |

Comparison with high-throughput experiments

Supplementary files with proteins from the published interactome of SARS-CoV-2 were downloaded [17], as well as proteins differentially expressed in a published translatome and proteome (P-value < 0.05 when comparing infection versus control at any time point post-infection) [16]. All gene identifiers were mapped to UniProt accession numbers. Then, co-expressed proteins from ASACO were independently compared to every dataset. To calculate the number of expected matches we take 21 489 as the number of total proteins in the human proteome, based on the Biomart gene type ‘protein_coding’. To calculate the p-value of the number of found matches, a hypergeometric test was used (dhyper function in R).

Genes regulated by interferon and genes related to stress granules

To check if a gene was induced or repressed by interferon, its expression was evaluated using the Interferome database v2.01 [32]. Only experiments where a gene had a fold change higher than 2 were considered. When a gene appeared differently expressed in more than one experiment, the average value was calculated.

Genes related to stress granules were extracted from the MSGP database (Mammalian Stress Granules Proteome), that store a total of 464 proteins [33]. Available gene names were mapped to UniProt accession numbers.

Results

Genes involved in well-known pathways from host factors show a positive correlation to these factors over hundreds of experiments

A set of human genes are known to be involved in the infectious cycle of SARS-CoV and SARS-CoV-2. We searched the literature and found 35 protein-coding genes that participate in different stages of their infection (Table 1). Then, we classified those host factors by the viral activity where they are involved (entry, replication or vesicle fusion), and created two subgroups according to their type of function (proviral or antiviral). Proteins encoded by these genes show common cellular functions (Figure 3). The three most populated groups of genes encode for: five replication proteins involved in metabolism of RNA, mainly mRNA splicing (DDX1, DDX5, HNRNPA1, PPPIH and ZCRB1), three proteins of vesicle fusion involved in endoplasmic reticulum to Golgi apparatus trafficking (COPB1, COPB2 and GBF1), and a large number of proteins involved in immune system, mainly interferon
Serial co-expression analysis of host factors from SARS-CoV viruses

Figure 1. Reactome pathways shared between three or more seeds. Seeds are highlighted by its type of activity (proviral or antiviral), and its function in the infection (entry, replication or vesicle fusion). Seeds with none shared pathways are not showed (PPIG, TOP3B and TMPRSS2).

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Genes with expression profiles similar to the seed ones present common pathways as well as others related to viral infections

To assess the agreement of functions from the best correlated genes to the ones of seeds, the expression profile of each seed was constructed using the available transcriptomic conditions. Genes with a similar expression profile (co-expressed genes) as well as genes with an inverse profile (inversely expressed genes) were obtained (Figure 4). The total number of found co-expressed genes was 2567, although several of them were common to different seeds. Thus, the number of different co-expressed genes was 1899, while the number of the inversely expressed ones was 1578 (Supplementary Figure S1, Table S1).

Co-expressed genes present common functions again related to those of the seeds (Figure 5). Pathways related to cytokine response such as interleukin or interferon signaling appear enriched to the expected antiviral genes IFITM1, IFITM2, IFITM3, EIF2AK2 and BST2, but also to the entry gene CTSL, and the replication genes PPIB and DDX1. In fact, the protease CTSL and cytokine signaling. This latter group includes five antiviral proteins (EIF2AK2, BST2, IFITM1, IFITM2 and IFITM3), together with the replication proteins IMPDH1, IMPDH2 and PPIA, and the vesicle protein VAPA. Furthermore, around the proteins of this group appear several proteins relevant for the virus entry into the host cell such as the proteases FURIN and cathepsin L1 (CTSL), as well as BSG and PIKFYVE. In fact, CTSL that improves the efficiency of the viral entry, is known to be involved in both adaptive and innate immune system [34].

All these proteins are part of autochthonous cell processes where they interact with others. Those proteins involved in the same biological processes frequently share expression profiles, which suggests they are co-regulated, often even showing common regulators. To study the expression relationships between the previous host factors and their co-expressed genes, we obtained transcriptomics experiments where they appear to be differentially expressed (Figure 2). Starting from these host factors, that we named as seeds from now on, 1381 different experiments were analyzed, 116 of them related to conditions involving viruses different from SARS-CoV-2. Then, the correlation value between the expression profile of these seeds versus all the other human genes through the different experiments was calculated.

It is expected that genes participating in the same biological process than the seed show a high correlation with it. Thus, we separated the correlation values for all the genes versus a given seed by the pathway they belong, and those pathways with a high average correlation value were analyzed. As a result, the most significant obtained pathways were usually those corresponding to the ones already annotated for the corresponding seed (Figure 3a), which would support the previous assumption that genes participating in the biological process where the seed is involved show a positive correlation with the seed. In addition, the most significant pathways found when analyzing all the seeds are once again discriminating those involved in replication (mainly metabolism of RNA, and cell cycle) from those related to vesicle fusion, including the main five antiviral genes (Figure 3b). Remarkably, seeds involved in cell entry appeared linked to these groups of genes. TPCN2 and BSG are linked to the group of replication seeds by the mitochondrial translation pathway, while NRPI and CTSL appear linked to the group of antiviral genes, once again by both interferon and cytokine signaling pathways (Figure 3b). These two groups would highlight the two important cell interactions with the virus: functions essential for its infectious cycle such as mRNA splicing, and the cytokine antiviral response.

and cytokine signaling. This latter group includes five antiviral proteins (EIF2AK2, BST2, IFITM1, IFITM2 and IFITM3), together with the replication proteins IMPDH1, IMPDH2 and PPIA, and the vesicle protein VAPA. Furthermore, around the proteins of this group appear several proteins relevant for the virus entry into the host cell such as the proteases FURIN and cathepsin L1 (CTSL), as well as BSG and PIKFYVE. In fact, CTSL that improves the efficiency of the viral entry, is known to be involved in both adaptive and innate immune system [34].

All these proteins are part of autochthonous cell processes where they interact with others. Those proteins involved in the same biological processes frequently share expression profiles, which suggests they are co-regulated, often even showing common regulators. To study the expression relationships between the previous host factors and their co-expressed genes, we obtained transcriptomics experiments where they appear to be differentially expressed (Figure 2). Starting from these host factors, that we named as seeds from now on, 1381 different experiments were analyzed, 116 of them related to conditions involving viruses different from SARS-CoV-2. Then, the correlation value between the expression profile of these seeds versus all the other human genes through the different experiments was calculated.

It is expected that genes participating in the same biological process than the seed show a high correlation with it. Thus, we separated the correlation values for all the genes versus a given seed by the pathway they belong, and those pathways with a high average correlation value were analyzed. As a result, the most significant obtained pathways were usually those corresponding to the ones already annotated for the corresponding seed (Figure 3a), which would support the previous assumption that genes participating in the biological process where the seed is involved show a positive correlation with the seed. In addition, the most significant pathways found when analyzing all the seeds are once again discriminating those involved in replication (mainly metabolism of RNA, and cell cycle) from those related to vesicle fusion, including the main five antiviral genes (Figure 3b). Remarkably, seeds involved in cell entry appeared linked to these groups of genes. TPCN2 and BSG are linked to the group of replication seeds by the mitochondrial translation pathway, while NRPI and CTSL appear linked to the group of antiviral genes, once again by both interferon and cytokine signaling pathways (Figure 3b). These two groups would highlight the two important cell interactions with the virus: functions essential for its infectious cycle such as mRNA splicing, and the cytokine antiviral response.

and cytokine signaling. This latter group includes five antiviral proteins (EIF2AK2, BST2, IFITM1, IFITM2 and IFITM3), together with the replication proteins IMPDH1, IMPDH2 and PPIA, and the vesicle protein VAPA. Furthermore, around the proteins of this group appear several proteins relevant for the virus entry into the host cell such as the proteases FURIN and cathepsin L1 (CTSL), as well as BSG and PIKFYVE. In fact, CTSL that improves the efficiency of the viral entry, is known to be involved in both adaptive and innate immune system [34].

All these proteins are part of autochthonous cell processes where they interact with others. Those proteins involved in the same biological processes frequently share expression profiles, which suggests they are co-regulated, often even showing common regulators. To study the expression relationships between the previous host factors and their co-expressed genes, we obtained transcriptomics experiments where they appear to be differentially expressed (Figure 2). Starting from these host factors, that we named as seeds from now on, 1381 different experiments were analyzed, 116 of them related to conditions involving viruses different from SARS-CoV-2. Then, the correlation value between the expression profile of these seeds versus all the other human genes through the different experiments was calculated.

It is expected that genes participating in the same biological process than the seed show a high correlation with it. Thus, we separated the correlation values for all the genes versus a given seed by the pathway they belong, and those pathways with a high average correlation value were analyzed. As a result, the most significant obtained pathways were usually those corresponding to the ones already annotated for the corresponding seed (Figure 3a), which would support the previous assumption that genes participating in the biological process where the seed is involved show a positive correlation with the seed. In addition, the most significant pathways found when analyzing all the seeds are once again discriminating those involved in replication (mainly metabolism of RNA, and cell cycle) from those related to vesicle fusion, including the main five antiviral genes (Figure 3b). Remarkably, seeds involved in cell entry appeared linked to these groups of genes. TPCN2 and BSG are linked to the group of replication seeds by the mitochondrial translation pathway, while NRPI and CTSL appear linked to the group of antiviral genes, once again by both interferon and cytokine signaling pathways (Figure 3b). These two groups would highlight the two important cell interactions with the virus: functions essential for its infectious cycle such as mRNA splicing, and the cytokine antiviral response.
presented several interleukins as co-expressed genes, such as Interleukin-1 beta (IL1B), and the X-C-C motif chemokines 2 and 3 (CXCL2, CXCL3). Other shared pathways are those related to the cell cycle. Five genes show enrichment in cell cycle checkpoints (DDX1, PPBP, PPBH, PRKCI and ZCRB1). The cell cycle disarrangement is an action that many viruses perform in the infected cells, where they induce a cell cycle arrest. For example, the coronavirus avian infectious bronchitis virus activates the cell ATR signaling, which contributes to S-phase arrest and is required for efficient virus replication and progeny production [35]. Genes of vesicle fusion such as COPB1, COPB2, OSBP, together with PPBH, are enriched in the expected processes of trafficking between the endoplasmic reticulum and the Golgi apparatus, glycosylation or the unfolded protein response. Other highlighted pathway is related to the nuclear export protein (NEP/NS2) of Influenza A virus that helps the transport of viral ribonucleoprotein complexes from the nucleus [36], and it is enriched from the correlators of DDX1, PPBP and PPBH. Other unexpected pathways were those relating DDX1, IMPDH2, PPBH and ZCRB1 to mitochondrial translation. Noteworthy, this latter pathway appeared previously related to PPBP, as well as the entry genes BSG and TPCN2, in the previous analysis of pathway average correlations (Figure 3b). Finally, the most remarkable common pathway in the inversely expressed genes was interferon-alpha/beta pathway, which appeared for both HNRNPA1 and PPBH genes. These genes are involved in mRNA splicing, that is one of the essential host functions for the virus, so it is expected that the cell response try to silence them, and because of this we expect that their negative correlators were related to cytokine response.

The best seeds’ correlators overlap with results of high-throughput experiments on SARS-CoV-2

To evaluate the list of co-expressed genes obtained, we compared them against genes identified in high-throughput laboratory experiments performed upon viral infection or in the presence of viral proteins. A recent work has completed the interactome of viral versus human proteins, and they found 332 human proteins interacting with viral proteins that are candidates to be involved in the viral infectious cycle [17]. In addition, other experimental group has published both the translatome and proteome in human cells infected by the virus [16], and they describe proteins that differentially change during the infection. In the latter case, we only considered for comparison proteins differentially expressed at any time post-infection. Only one seed used in the present work, IMPDH2, appears in the interactome dataset, as well as nine in the differential proteome. So, on the assumption that both proteomics and interactomics constitute an experimental approach of the cellular response to the viral infection, the co-expressed genes proposed by ASACO were compared to the discovered genes in these
Figure 3. Significant pathways found by ASACO. A) Distribution of correlation values of genes belonging to the most significant pathways (maximum 10) for six of the seeds. Pathways already annotated for the seed are highlighted in red color. The number of the genes that the pathway has annotated is shown in brackets. The solid line marks correlation 0, and the dashed line marks the median of the correlation for all the genes. B) Significant pathways shared between three or more seeds. Seeds are highlighted by its type of activity (proviral or antiviral), and its function in the infection (entry, replication or vesicle fusion). Seeds with none shared pathways are not showed (ABL2, ACE2, COPB1, DDX5, EZR, FURIN, GBP1, GSK3A, GSK3B, IMPDH1, OSBP, PIKFYVE, TMPRSS2 and VAPA). Pathways already annotated for any seed are highlighted in red color.
Figure 4. Expression profile of all seeds together with their positive and negative correlated genes. The black colored line represents the expression profile for the target gene through the different experiments (X axis), which are ordered from its higher to lower log2FC. Correlated genes are labeled in green (co-expressed genes), or red (inversely expressed genes), together with its deviation (Q1 an Q3 quartiles). Note that the number of ticks on the X axis is relative to the number of experiments for that seed.

experiments. The expected coincidence between the ASACO and the experimental results based on the size of the datasets was initially low. In contrast, around 20% of the positively correlated genes proposed by our approach appear in any of those experimental datasets, with 19 genes shared by the three approaches (14 positive and 5 negative correlators) (Figure 6a). This suggests that proposed co-expressed genes are similar to those obtained in infection experiments. However, 1539 of our proposed genes did not appear as results in these experimental analyses. To assess whether these new genes could perform virus related functions their associated pathways were analyzed. In fact, a good number of genes are involved in processes related to the infectious cycle of other viruses such as Influenza A, Epstein–Barr, HIV, Hepatitis B and C and Measles (Figure 6b). Another functions found were systems for DNA repair, which are also used by viruses such as the Epstein–Barr [37]. Furthermore, it is noteworthy the enrichment in RNA metabolism and cell cycle checkpoints (Figure 6c), which we have previously mentioned as important for the efficient viral replication [35]. Finally, there are functions well-known for SARS-CoV viruses related with the immune system response and involving the interferon/interleukin signaling [26].

Other correlation coefficients, different to the one presented here, are commonly used to discover co-expressed genes. One of the most used bioinformatics tools for this kind of analysis is WGCNA, that uses the Pearson coefficient, but also allows to use both the Spearman and the Biweight midcorrelation coefficient [38]. We calculated the correlations for the 35 seed using the three alternative coefficients, and all of these are able to find a similar number of positive correlators, with the latter one finding a bit more genes that match with the high-throughput experiments (Supplementary Figure S2; Table 2). However, negative correlators are seldom found by any coefficient except for ASACO.

These results suggest that co-expressed genes found by our approach can offer an accurate landscape of the cellular pathways and proteins affected by the virus when the infection is progressing, even though it is based on heterogeneous experiments from many different conditions that do not include coronavirus infections.
Transcription factors co-expressed with seeds are regulated by interferon and could induce the expression of genes involved in cell entry

The coincidence in the assigned pathways for seeds and their correlators suggests that SARS-CoV-2 host factors may belong to common regulatory networks, possibly sharing transcriptional regulators. Moreover, these regulators could be co-regulated with seeds as well. To test this hypothesis, transcription factors were identified from the correlators. So, 116 transcription factors, or putative upstream regulators, were found among the co-expressed genes, and 155 in the inversely expressed genes. Among them, 23 co-expressed regulators were common to two or more seeds (Figure 7a). They form an interrelated network with the main antiviral genes (EIF2AK2, BST2, IFITM1, IFITM2 and IFITM3), but some of them also appeared co-expressed with proviral genes involved in vesicle fusion such as COPB1, COPB2 and VAPA, replication, as DDX5, and entry, suggesting common regulatory features. Related to viral entry, the transcription factor NUPR1, a stress-response protein induced by the Hepatitis B virus [39], is co-expressed with the protease TMRPSS2. Moreover, the factor ZNF267, which is an antiviral zinc finger protein [40], is co-expressed with the kinase ABL2. Finally, TRIM14, which is a member of a family of E3 ubiquitin ligases linked to the mitochondria that plays an important role in innate defense against viruses facilitating the interferon response [41], is co-expressed with both the receptor ACE2 and the lysosomal channel TPCN2. Remarkably, all the regulators connecting the antiviral seeds with several proviral activities, including the cell entry, are genes induced by interferon. However, other transcription factors mainly connecting replication genes are contrarily repressed by interferon. For example, YEATS4 is co-expressed with DDX1, IMPDH2, and ZC3H1, that were co-expressed with genes involved in mitochondrial translation. Other remarkable co-expressed gene is CEBPZ, which belongs to a family of CCAAT/enhancer-binding proteins, some of them related to immune and inflammatory response [42].

CEBPZ is a gene co-expressed with the replication seeds DDX1 and PPIG, but also inversely expressed with the protease FURIN. Other correlators from this protein family are CEBPD, which is a co-expressed gene of IFITM3, but inversely expressed with COPB2, and CEBPB that is a co-expressed gene of the protease CTSL.

Conversely, regulators inversely expressed with seeds are mainly interferon repressed genes (Figure 7b). Among these genes, three zinc-finger proteins that act as antivirals against Herpes simplex virus 1 stand out in this dataset [43]: ZNF91 inversely expressed with CTSL, IFITM2, IFITM3 and VAPA, that is not induced or repressed by interferon, is a transcription factor specifically required to repress SINE-VNTR-Alu (SVA) retrotransposons [44]; the transcription factor ZNF550, which is repressed by interferon and appeared inversely correlated with the entry genes IFITM1 and BSG; and ZNF768, that appeared as inversely expressed to the genes involved in vesicle fusion COPB1 and OSBP. Two others inversely expressed genes were the transcriptional repressors YBX3 and CREBBP, which are repressed by interferon and link the entry genes ACE2 and TPCN2. Y-box-binding protein 3 (YBX3) restricts Influenza A virus by impairing viral ribonucleoprotein complexes [45], and also controls amino acid levels by regulating solute carrier amino acid transporter mRNA abundance, a pathway that is related with inversely expressed genes of HPRNP41 (Figure 5b).

To further analyze the interferon signaling related with the seeds, the response to interferon was independently evaluated for each seed, together with their co-expressed and inversely expressed genes. As expected, the main interferon induced genes, the antivirals IFITM1–3, BST2 and EIF2AK2, presented an important activation together with its co-expressed genes, and a repression of their inversely expressed genes (Figure 7c). Unexpectedly, proviral genes involved in entry, ACE2, ABL2 and CTSL, together with the replication gene PPB, present a similar interferon response, suggesting an undesired effect of interferon promoting SARS-CoV-2 infection. On the other hand, the seeds HNRNPA1, COPB2 and VAPA present a reversed profile.
Figure 6. Overlapping between co-expressed proteins found by ASACO and other proteins related with the SARS-CoV-2 infection from high-throughput experiments.

A) Co-expressed genes found by ASACO are shown inside the blue box. The interactome obtained by Gordon et al. [17] is displayed in the green box, the translatome and the proteome by Bojkova et al. [16] are shown in the pink and the yellow boxes, respectively. The total number of proteins in each dataset is shown inside the boxes as well as the number of overlapping proteins (n), the number of coincidences expected by chance (e), and the P-value calculated with the hypergeometric distribution (p). Results for positively (green color), and negatively (red color) correlated genes are separated. The table represents the number of proteins found by ASACO for each seed, together with those that overlap with any other dataset in brackets and separated by the correlation sign. Darker background color indicates that this seed is found in one of the high-throughput experiments. The total number of proteins is indicated below together with the total number of shared proteins and the number of positively correlated proteins exclusively found by ASACO. Outside the blue box, the number of proteins shared by the four datasets is shown.

B) KEGG pathway enrichment for genes exclusively found by ASACO.

C) Reactome pathway enrichment for genes exclusively found by ASACO.

Specifically, HNRNPA1 is a gene involved in mRNA splicing, which is moved toward the stress granules during viral infection. These structures include ribonucleoprotein complexes together with the cell translation machinery, and it is proposedly used by the virus to perform its replication [46]. These cytosolic particles seem to be targeted by the viral nucleocapsid protein (N). The N protein interact with 15 human proteins [17], and 3 of them were found as co-expressed genes of COPB1 and COPB2 (FAM98A), EIF2AK2 (MOV10), and HNRNPA1 and IMPDH2 (PABPC4). These three proteins have been associated to the stress granules, together with DDX1, EIF2AK2 and HNRNPA1, which reinforce the relation of the N viral protein with this stress structures [47–49]. In fact, the antiviral seed EIF2AK2 has been seen as the kinase activated by double-stranded RNA of viruses that activate the...
formation of the stress granules [50]. Currently, 464 proteins are known to form part of this liquid–liquid structures, and the seed co-expressed genes include 131 of them (41 were expected by chance; 4.6e-35) (Figure 7d), 21 from the interactome (7 expected; 7.7e-06), and 100 from the proteome (24 expected, 4.6e-35), which support the importance of these structures in the SARS-CoV-2, as well as the fact that the virus could use them for its replication and mRNA translation.

Discussion
Currently, databases sharing gene expression data are exponentially growing due to the universalization of transcriptomics techniques [51]. A secondary analysis of these data allows to reconstruct gene networks based on co-expressed genes using the so-called reverse engineering [52, 53]. We have developed a new in silico method called ASACO based on standardized gene expression data analysis. Starting from an initial seed gene, it finds the closest neighbors in terms of transcriptional regulation. Compared to other previous procedures, it has the advantage of evaluating thousands of experiments whose outcomes are normalized, allowing co-expression analysis for different heterogeneous genes over hundreds of experimental conditions [54, 55]. The ASACO algorithm search for co-expressed genes using a novel strategy that prioritizes genes with fold-change signs similar to the seed expression profile (see Material and Methods for more details). Usual metrics to measure

![Figure 7](https://example.com/figure7.png)

**Figure 7.** Transcription factor network from co-expressed genes, and relation with interferon and stress granules. A) Positively expressed transcription factors common to, at least, two seeds. Seeds are highlighted by its type of activity (proviral or antiviral), and its function in the infection (entry, replication, or vesicle fusion). Transcription factor induced by interferon are highlighted in red color and those repressed by interferon in blue color. B) The same as A) but for the negatively expressed transcription factors. C) Average interferon induced fold change of seeds, together with the value for both positively and negatively expressed genes. D) Number of genes related to stress granules in both positively and negatively expressed genes for each seed. The first column of the heatmap shows if the seed is related or not to stress granules.
correlation between expression profiles are Pearson and Spearman coefficients, or more robust indices such as the biweight midcorrelation used by the suite WGCNA [38]. But these are based on linear correlations, where it is not penalized when a gene has an inverse behavior compared to the seed or an outlier appears in either one or a few experiments. This is overtaken by ASACO, since it gives a higher weight to genes with the same behavior than the seed (overexpression or underexpression), and this strategy has shown a better result when the final aim is to discover negative regulators (Table 2). Contrary to positive correlators, negative ones can be useful to identify relevant therapeutic targets whose inhibition could activate antiviral genes or induce the expression of disease modifiers.

We have used this strategy on a selection of 35 cellular genes reportedly involved in SARS-CoV and/or SARS-CoV-2 infection, identifying their closest co-expressed genes. Even though the experiments employed to find them are not focused on coronaviruses infection, the functional enrichment of the co-expressed genes identify many of the already known pathways for these seeds (Figure 3). Furthermore, both seeds and correlators fairly match most of the cellular pathways relevant to the infection cycle (Figure 5). Moreover, these co-regulated genes show a high coincidence with the ones identified in recent high-throughput studies on cell responses to SARS-CoV-2 infection (Figure 6a). This match can be interpreted as a cross-validation of both, experimental and in silico approaches, providing relevance to our identified functions and supporting the idea that co-regulation, as we identify it, can be a hallmark of co-function. These results complement experimental data, since they show regulatory genes with subtle expression changes during the infection, which could constitute new candidates for therapeutic targets. Probably, not all the correlated genes will be involved in the viral infection, since the seeds can perform different unrelated activities in the cell. However, both the coincidence with infection studies and the functional relation with viral pathways would be useful in this case to propose the most valuable targets.

When viruses enter the host, one of the first systems that respond to the infection is the interferon signaling pathway that induces the expression of interferon stimulated genes (ISG). Five of the seeds used in this study, considered as antiviral genes, are ISGs. EIF2AK2 inhibits viral replication via the integrated stress response, and blocks the cellular and viral translation through the phosphorylation of EIF2α [56]. BST2 blocks the release of viruses by directly tethering nascent virions to the membranes of infected cells [57]. Finally, the transmembrane proteins IFITM1-3 inhibit the entry of enveloped virus by preventing vesicle fusion, even though they also could facilitate the infection of other viruses [58]. As expected, genes positively correlated with these antiviral genes consistently belong to interferon and interleukins response pathways (Figure 5a), and our co-expression analysis links them to regulators as STAT1, STAT3 or TRIM21 (Figure 7a). Surprisingly, several proviral seeds, remarkably those involved in virus entry (ACE2, TPCN2, ABL2 and TMPRSS2), are found to be linked to these antiviral genes by means of common co-expressed transcriptional regulators. The main viral receptor, ACE2, has already been found as induced by interferon, and this fact has been interpreted as ‘evidence that coronaviruses, as well as other viruses, have evolved to leverage features of the human IFN pathway’ [25]. This could be true for other entry genes. TRIM14 is one of the regulators found in common to ACE2, TPCN2 and four out of five antivirals (excluding IFITM2). This regulator is known to interact with MAVS at the outer mitochondria membrane and attenuates the antiviral response by the type I interferon response [59]. Since TRIM14 is co-expressed with ACE2, it could be responsible of, or related to, the receptor’s positive response to interferon. Interestingly, not only ACE2 seems to be induced by interferon, but also its positive correlators, as well as those of ABL2 (which could be induced by the ISG ZNF267 according to our results), and the cathepsin L protease, CTSL. All of these are genes involved in the viral entry, and the latter activates the membrane fusion function of the spike viral protein S of SARS-CoV [60], and could substitute to the TMPRSS2 protease in cell types different to lung cells [61]. CTSL co-expressed genes are not only interferon but also interleukin induced genes, and three cytokines are co-expressed with it (IL18, CXCL2 and CXCL3). In addition, it shares with IFITM2, IFITM3 and the vesicle gene VAPA links to the transcriptional repressor ZNF91. All of this points to both ZNF91 and TRIM14 as putative regulators responsible for the presumptive interferon and interleukin dependent induction of entry proviral genes and pinpoints them as useful pharmacological targets to interfere with the viral infection. In fact, the cellular response to SARS-CoV-2 has been shown to be lightly induced by interferon, but strongly by chemokines [62]. Furthermore, CTSL appeared also linked to the co-expressed transcription factor CBPB that regulates the expression of genes involved in immune and inflammatory responses [42]. This transcription factor can form heterodimers with CEBPD, which is also a co-expressed gene of IFITM3, but inversely expressed gene of the vesicle protein COBP2. It suggests that the interleukin response could repress proviral genes such as COBP2, but undesirably induce CTSL.

Contrarily, genes required for viral replication such as PPIG, DDX1, 2CR1, IMPDH2, GSK3A and GSK3B, are bound to interferon repressed transcriptional regulators such as RFX7 and YEAT54, which could negatively affect the infection in the presence of interferon.

One essential cellular pathway needed by the virus is the mRNA splicing. In this function, HNRNPA1 is a key gene [63]. Interferon triggers EIF2AK2-dependent HNRNPA1 protein translocation to stress granules resulting in mRNA translation inhibition. Until now, 464 different human proteins have been identified to participate in this liquid–liquid cytosolic structures [33]. Remarkably, proteins from stress granules show a broad convergence with our host factors co-expressed genes (131 out of 464), as well as those from the high-throughput experiments.

| Correlation | ASACO | Pearson | Spearman | Biweight | ASACO | Pearson | Spearman | Biweight |
|-------------|-------|---------|----------|----------|-------|---------|----------|----------|
| Positive    | 1899  | 1895    | 1989     | 1989     | 360   | 339     | 362      | 369      |
| Negative    | 1578  | 14      | 51       | 50       | 166   | 1       | 5        | 5        |

Table 2. Number of genes found when using different correlation metrics, and matching with genes found in the interactome, translate and proteome. The threshold P-value was customized in every metrics to obtain a comparable number of co-expressed genes. Biweight = Biweight midcorrelation.
Conclusions

We here presented ASACO, an algorithm with the capacity to generate key functional knowledge on specific genes based on their co-expressed genes. We have tested this algorithm with host factors involved in the SARS-CoV-2 infection, by means of the analysis of co-expression data extracted from public transcriptomics databases. Although further experiments using in vitro and in vivo approaches will be required to further confirm the results obtained here, our results have allowed the discovery of relevant gene networks and cell pathways, and pointed to a series of transcription regulators as potential targets useful in the fighting against SARS-CoV-2. The consistency of our results with those obtained by other experimental approaches represents a proof of concept of the utility of this algorithm, which could be used for the study in other pathologies where there is still a need for discovering new functional knowledge, such as molecularly uncharacterized rare diseases and other microbial infections.

Key Points

- ASACO identifies regulatory associations of genes using public transcriptomics data.
- ASACO highlights new cell functions likely involved in the infection of coronavirus.
- Comparison with high-throughput screenings validates candidates proposed by ASACO.
- Genes co-expressed with host’s genes used by SARS-CoV-2 are related to stress granules.

Supplementary Data

Supplementary data are available online at Briefings in Bioinformatics.

Acknowledgements

We thank to GaliciAME for the project that allowed the initial development of ASACO, to Gustavo Aguilar for the development of GUPO to download the transcriptomics experiments, to C3UPO for the HPC support, and to Dr. Javier Sánchez-Céspedes for the critical reading of our manuscript.

We want to dedicate this article to the memory of our colleague and friend Jose Luis Gómez-Skarmeta.

Data availability

Scripts are available in the GitHub repository, https://github.com/upobioinfo/asaco

Funding

The author received no specific funding for this work.

Conflict of Interest

The authors declare that they have no conflict of interests.

References

1. Rogers JM, Bulyk ML. Diversification of transcription-factor–DNA interactions and the evolution of gene regulatory networks. Wiley Interdiscip Rev Syst Biol Med 2018;10:e1423.
2. Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell 2016;167:1170–7.
3. Lobo I. Biological complexity and integrative levels of organization. Nat Edu 2008;1:141.
4. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet 2012;13:613–26.
5. Papaetheodorou I, Moreno P, Manning J, et al. Expression atlas update: from tissues to single cells. Nucleic Acids Res 2020;48:D77–83.
6. Kamitani W, Huang C, Narayanan K, et al. A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. Nat Struct Mol Biol 2009;16:1134–40.
7. Thoms M, Buschauer R, Ameismeier M, et al. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. Science 2020;369:1249–55.
8. Fung TS, Liu DX. Post-translational modifications of coronavirus proteins: roles and function. Future Virol 2018;13:405–30.
9. Sanche S, Lin YT, Xu C, et al. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 2020;26:1470–7.
10. Belouzard S, Millet JK, Licitra BN, et al. Mechanisms of coronavirus cell entry mediated by the spike protein. Viruses 2012;4:1011–33.
11. Bouland S, Stanifer M, Lozach PY. Dynamics of virus-receptor interactions in virus binding, signaling, and endocytosis. Viruses 2015;7:2794–815.
12. Heald-Sargent T, Gallagher T. Ready, set, fuse! The coronavirus spike protein and Acquisition of Fusion Competence. Viruses 2012;4:557–80.
13. Fukushi M, Yoshinaka Y, Matsuoka Y, et al. Monitoring of S protein maturation in the endoplasmic reticulum by Calnexin is important for the infectivity of severe acute respiratory syndrome coronavirus. J Virol 2012;86:11745–53.
14. Fung TS, Liu DX. Human coronavirus: host-pathogen interaction. Annu Rev Microbiol 2019;73:529–57.
15. Masters PS. The molecular biology of coronaviruses. Adv Virus Res 2006;65:193–292.
27. Huang C, Wang Y, Li X, et al. A Kinome-wide small interfering RNA screen identifies Provilial and antiviral host factors in severe acute respiratory syndrome coronavirus replication, including double-stranded RNA-activated protein kinase and early secretory pathway proteins. J Virol 2015;89:8318–33.

28. de Wilde AH, Wannee KF, Scholte FM, et al. ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across the protein components of the mammalian stress granules. J Virol 2009;83:681–7.

29. Durinck S, Spellman PT, Birney E, et al. MSGP: the first database of the protein components of the mammalian stress granules. Database 2019;2019:baz031.

30. Ziegler CGK, Allon SJ, Nyquist SK, et al. Nuclear traffic of different DNA repair systems participate in Epstein-Barr virus genome maturation. J Virol 2011;85:6127–35.

31. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.

32. Rusinova I, Forster S, Yu S, et al. Cellular functions are highly expressed in nasal epithelial cells together with innate immune genes. Nat Med 2020;26:681–7.

33. Pinto BG, Oliveira AE, Singh Y, et al. ACE2 expression is increased in the lungs of patients with comorbidities associated with severe COVID-19. J Infect. Dis. 2020;222:556–63.

34. Zavaňnik-Bergant V, Schweiger A, Bevec T, et al. Inhibitory p41 isoform of invariant chain and its potential target enzymes cathepsins L and H in distinct populations of macrophages in human lymph nodes. Immunology 2004;112:378–85.

35. Xu LH, Huang M, Fang SG, et al. Coronavirus infection induces DNA replication stress partly through interaction of its nonstructural protein 13 with the p125 subunit of DNA polymerase δ. J Biol Chem 2011;286:39546–59.

36. Boulo S, Akarsu H, Ruigrok RWH, et al. Nuclear traffic of influenza virus proteins and ribonucleoprotein complexes. Virus Res 2007;124:12–21.

37. Sugimoto A, Kanda T, Yamashita Y, et al. Spatiotemporally different DNA repair systems participate in Epstein-Barr virus genome maturation. J Virol 2011;85:6127–35.

38. Melchjorsen J, Matikainen S, Paludan SR. Activation and evasion of innate antiviral immunity by herpes simplex virus. Viruses 2009;1:737–59.

39. Jacobs FMJ, Greenberg D, Nguyen N, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature 2014;516:242–5.

40. Qin Z, Qu X, Lei L, et al. Y-box-binding protein 3 (YBX3) restricts influenza virus by interacting with viral ribonucleoprotein complex and impairing its function. J Gen Virol 2020;101:385–98.

41. Perdikari TM, Murthy AC, Ryan VH, et al. SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins. bioRxiv 2020;2020.06.09.141101.

42. Ozeki K, Sugiyama M, Akter KA, et al. FAM98A is localized to stress granules and associates with multiple stress granule-localized proteins. Mol Cell Biochem 2019;451:107–15.
53. Basso K, Margolin AA, Stolovitzky G, et al. Reverse engineering of regulatory networks in human B cells. Nat Genet 2005;37:382–90.
54. Gibbs DL, Grajinski L, Baric RS, et al. Multi-omic network signatures of disease. Front Genet 2013;4:309.
55. Cardozo LE, Russo PST, Gomes-Correa B, et al. WebCENITool: co-expression modular analysis made easy. Front Genet 2019;10:146.
56. Kang JH, Kwon SN, Park SH, et al. PKR protein kinase is activated by hepatitis C virus and inhibits viral replication through translational control. Virus Res 2009;142:51–6.
57. Dafa-Berger A, Kuzmina A, Fassler M, et al. Modulation of hepatitis C virus release by the interferon-induced protein BST-2/tetherin. Virology 2012;428:98–111.
58. Lim YY, Ng Y, Tam J, et al. Human coronaviruses: a review of virus-host interactions. Diseases 2016;4:26.
59. Zhou Z, Jia X, Xue Q, et al. The antiviral mechanism of BST-2/tetherin. J Virol 2018;92:8887–90.
60. Bosch BJ, Bartelink W, Rottier PJM. Cell biology of SARS-CoV-2: receptor binding and entry in the human airway. J Cell Biol 2020;209:1013–10145.e9.
61. Liu H, Gai S, Wang X, et al. Single-cell analysis of SARS-CoV-2 receptor ACE2 and spike protein priming expression of proteases in the human heart. Cardiovasc Res 2020;116:1733–41.
62. Blanco-Melo D, Nilsson-Payant B, Liu WC, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell 2020;181:1013–10145.e9.
63. Luo H, Chen Q, Chen J, et al. The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. FEBS Lett 2005;579:2623–8.
64. Savastano A, Opakua AI d, Rankovic M, et al. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. Nat Commun 2020;11:579.
65. Chen JY, Chen WN, Poon K MV, et al. Interaction between SARS-CoV helicase and a multifunctional cellular protein (Ddx5) revealed by yeast and mammalian cell two-hybrid systems. Arch Virol 2009;154:507–12.
66. Lindquist ME, Lifland AW, Utlej Tj, et al. Respiratory syncytial virus induces host RNA stress granules to facilitate viral replication. J Virol 2010;84:12274–84.
67. Amini-Bavil-Olyaee S, Choi S, Lee YJ, et al. The antiviral effector IFITM3 disrupts intracellular cholesterol homeostasis to block viral entry. Cell Microbe 2013;13:452–464. doi:10.1016/j.jcom.2013.03.006
68. Coleman CM, Sisk JM, Mingo RM, et al. Abelson Kinase Inhibitors Are Potent Inhibitors of Severe Acute Respiratory Syndrome Coronavirus and Middle East Respiratory Syndrome Coronavirus Fusion. J Virol 2016;90:8924–8933. doi:10.1128/jvi.01429-16
69. Coutard B, Valle C, de Lamballerie X, et al. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. Antiviral Res 2020;176:104742. doi:10.1016/j.antiviral.2020.104742