Bioinformatics analysis of expression profiling by high throughput sequencing for identification of potential key genes among SARS-CoV-2/COVID 19

Basavaraj Mallikarjunayya Vastrad
Department of Pharmaceutics, SET’S College Of Pharmacy, SR Nagar, Dharwad

Chanabasayya Mallikarjunayya Vastrad (✉ channu.vastrad@gmail.com)
Chanabasava Nilaya  https://orcid.org/0000-0003-3615-4450

Research Article

Keywords: SARS-CoV-2 infection, bioinformatics analysis, differentially expressed genes, novel biomarkers, hub genes

DOI: https://doi.org/10.21203/rs.3.rs-122015/v2

License: ☝️  This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) is pandemic recently emerged and is rapidly spreading in humans. However, the precise molecular mechanisms of the advancement and progression of SARS-CoV-2 infection remain unclear. The current investigation attempted to identify and functionally analyze the differentially expressed genes (DEGs) between SARS-CoV-2 infection and mock by using comprehensive bioinformatics analyses. The GSE148729 expression profiling by high throughput sequencing was downloaded from the Gene Expression Omnibus (GEO) and analyzed using the limma package in R software to identify DEGs. Pathway and gene ontology (GO) enrichment analysis of the up and down regulated genes were performed in ToppGene. The HIPPIE database was used to evaluate the interactions of up and down regulated genes and to construct a protein-protein interaction (PPI) network using Cytoscape software. Hub genes were selected using the Network Analyzer plugin. Subsequently, extensive target prediction and network analyses methods were used to assess, target gene - miRNA regulatory network and target gene - TF regulatory network. Receiver operating characteristic (ROC) analysis was utilized for validation. A total of 928 DEGs (461 up regulated genes and 467 down regulated genes) were identified between SARS-CoV-2 infection and mock samples. The Pathway enrichment analysis results showed that these up and down regulated genes were significantly enriched in cytokine-cytokine receptor interaction, and ascorbate and aldarate metabolism. Several significant GO terms, including the response to biotic stimulus and oxoacid metabolic process, were identified as being closely associated with these up and down regulated genes. The top hub genes and target genes were screened and included JUN, FBXO6, PCLAF, CFTR, TXNIP, PMAIP1, BRI3BP, FAHD1, PROX1, CXCL11, SERHL2 and CFI. ROC curve analysis showed that messenger RNA levels of these ten genes (DDX58, IFITM2, IRF1, PML, SAMHD1, ACSS1, CYP2U1, DDC, PNMT and UGT2A3) exhibited better diagnostic efficiency for SARS-CoV-2 infection and mock. The current investigation identified a series of key genes and pathways that may be involved in the progression of SARS-CoV-2 infection, providing a new understanding of the underlying molecular mechanisms of SARS-CoV-2 infection.

Introduction

Severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) is key causes of respiratory disease in pandemic form [1] and named as COVID - 19. In 2020, more than 406,207 people die of SARS-CoV-2 complications in the entire world. Restriction of the spread of SARS-CoV-2 virus has been difficult, but transmissions of this virus develop to require close contact through large-particle aerosols [2]. SARS-CoV-2 infections in humans cause a severe pneumonia that is disastrous in about 11% of infected individuals [3].

Intensive investigation has been achieved on the virulence and evolution of the SARS-CoV-2 virus [4]. However, very little is known about the influence of specific genes or biomarkers in humans that add to the susceptibility to SARS-CoV-2 infections. Consequently, it needs more effort to clarify the molecular mechanism underlying SARS-CoV-2 infections advancement and progression, holding promise for finding potential drug targets and diagnostic biomarkers of SARS-CoV-2 infections.
Expression profiling by high throughput sequencing analysis based on microarray technology is a widely used approach, which can simultaneously detect expression change of thousands of genes on the mRNA level. By gene expression profiling analysis with microarray technology, some investigations have found many differently expressed genes which played a key role in SARS-CoV-2 infections initiation and development and could be assessed as possible molecular targets and diagnostic markers. ACE2 and TMPRSS2 could function as a key mediator of SARS-CoV-2 infections by means of gene expression analysis [5]. In the present investigation, we identified SARS-CoV-2 infection linked with differential expressed genes (DEGs) between SARS-CoV-2 infected samples and mock samples, and successively performed pathway and GO term enrichment analysis, PPI network analysis, modules analysis, target gene - miRNA regulatory network and target gene - TF regulatory network analysis to discover the hub and target genes and pathways closely related to SARS-CoV-2 infections. Finally, hub genes were validated by receiver operating characteristic (ROC) curve analysis.

**Material And Methods**

**Microarray data**

The expression profiling by high throughput sequencing dataset GSE148729 was downloaded from the National Center for Biotechnology Information Gene Expression Omnibus Dataset (https://www.ncbi.nlm.nih.gov/gds) [6] and was based on the GPL18573 Illumina NextSeq 500 (Homo sapiens) array. The GSE148729 dataset, which is expression profiling by high throughput sequencing dataset, contains 22 samples, including 14 SARS-CoV-2 infections samples and 8 mock samples.

**Data pre-processing and identification of DEG**

Limma package in R bioconductor was used for the pre-processing for GSE148729 dataset, which included background correcting, normalizing and calculating expression [7]. The DEGs in SARS-CoV-2 infections samples compared with mock samples were determined using limma package. DEGs were screened with a adj. P value corrected P<0.05 and |log fold-change (FC)|>1. The adj.P.Val for up and down regulated genes corrected P<0.05 and | log FC|>1 were used as the screening thresholds.

**Pathway enrichment analysis of DEGs**

(https://toppgene.cchmc.org/enrichment.jsp) [17] to perform the pathway enrichment analysis. A P-value <0.05 was considered to be statistically significant.

**Gene ontology (GO) enrichment analysis**

Gene ontology (GO) (http://www.geneontology.org) enrichment analysis defines approach used to characterize gene function and relationships [18]. It classifies functions such as biological process (BP), cellular component (CC) and molecular function (MF). ToppGene (ToppFun) (https://toppgene.cchmc.org/enrichment.jsp) [17] was used to classify significant DEGs by their BP, CC and MF through GO and the significant transcripts (Benjamini–Hochberg false discovery rate <0.05) were identified using the Functional Annotation clustering tool.

**Protein-protein interaction network construction and module analysis**

The online Human Integrated Protein-Protein Interaction rEference (HIPPIE) (http://cbdm.uni-mainz.de/hippie/) [19] database was used to identify potential interaction among the up and down regulated genes and associated with various PPI data bases such as IntAct (https://www.ebi.ac.uk/intact/) [20], BioGRID (https://thebiogrid.org/) [21], HPRD (http://www.hprd.org/) [22], MINT (https://mint.bio.uniroma2.it/) [23], BIND (http://download.baderlab.org/BINDTranslation/) [24], MIPS (http://mips.helmholtz-muenchen.de/proj/ppi/) [25] and DIP (http://dip.doe-mbi.ucla.edu/dip/Main.cgi) [26]. Cytoscape 3.8.0 (http://www.cytoscape.org/) [27] was used to visualize the results from the PPI network. The Network Analyzer Cytoscape plug-in was used to explore hub genes were generated using highest node degree [28], betweenness centrality [29], stress centrality [30], closeness centrality [31] and lowest clustering coefficient [Wang et al 2012] methods and hub genes were further analyzed for pathway and GO enrichment analysis. Module analysis on the PPI network results was performed using the PEWCC1 (http://apps.cytoscape.org/apps/PEWCC1) [32] clustering algorithm that comes with Cytoscape. Module analysis was applied to analyze more connected gene groups. In addition, the module analysis results were further analyzed for pathway and GO enrichment analysis.

**Construction of target genes - miRNA regulatory network**

miRNet database (https://www.mirnet.ca/) [33] is a bioinformatics platform for predicting target genes (up and down regulated genes) and miRNA pairs. In the current investigation, the targets genes were predicted using ten databases: TarBase (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index) [34], miRTarBase (http://mirтарбасе.mbc.nctu.edu.tw/php/download.php) [35], miRecords (http://miRecords.umn.edu/miRecords) [36], miR2Disease (http://www.mir2disease.org/) [37], HMDD (http://www.cuilab.cn/hmdd) [38], PhenomiR (http://mips.helmholtz-muenchen.de/phenomir/) [39], SM2miR (http://bioinfo.hrmbu.edu.cn/SM2miR/) [40], PharmacomiR (http://www.pharmacomi-r.org/) [41], EpimiR (http://bioinfo.hrmbu.edu.cn/EpimiR/) [42] and starBase (http://starbase.sysu.edu.cn/) [43]. The screening criterion was that the miRNA target exists in the ten miRNA databases concurrently. The target genes - miRNA regulatory network was depicted and visualized using Cytoscape software.
Construction of target genes - TF regulatory network

NetworkAnalyst database (https://www.networkanalyst.ca/) [44] is a bioinformatics platform for predicting target genes (up and down regulated genes) and TF pairs. In the current investigation, the targets genes were predicted using JASPAR (http://jaspar.genereg.net/) [45]. The screening criterion was that the TF target exists in the TF database concurrently. The target genes - TF regulatory network was depicted and visualized using Cytoscape software.

ROC analysis

Receiver operating characteristic (ROC) curve analysis was implemented to calculate the sensitivity and specificity of the up and down regulated genes for SARS-CoV-2 infections diagnosis using the R “pROC” package [46]. An area under the curve (AUC) value was determined and used to label the ROC effect.

Results

Data pre-processing and identification of DEG

Expression profiling by high throughput sequencing dataset GSE148729 was downloaded from GEO database. After quality detection of microarray raw data, microarray based on GPL18573 platform, including 14 SARS-CoV-2 infections samples and 8 mock samples. Limma in R package was used for preprocessing and gene differential expression analysis of microarray data (Fig. 1A and Fig. 1B). Total of 928 DEGs (fold change >1, Fig. 2) consisting 461 significantly up regulated genes and 467 significantly down regulated genes, for the subsequent bioinformatics analysis. The hierarchical clustering heat map of up and down regulated genes shows the differences between SARS-CoV-2 infections and mock (Fig. 3 and Fig. 4).

Pathway enrichment analysis of DEGs

To analyze the biological classification of up and down regulated genes, functional and pathway enrichment analyses were performed using ToppGene. Pathway analysis revealed that the up regulated genes were mainly enriched in superpathway of steroid hormone biosynthesis, aspirin triggered resolvin E biosynthesis, cytokine-cytokine receptor interaction, TNF signaling pathway, calcineurin-regulated NFAT-dependent transcription in lymphocytes, ATF-2 transcription factor network, cytokine signaling in immune system, interferon signaling, steroid hormone metabolism, androgen and estrogen metabolism, genes encoding secreted soluble factors, cytokines and inflammatory response, toll receptor signaling pathway, interleukin signaling pathway, nuclear factor kappa B signaling, c-Jun N-terminal kinases MAPK signaling, 17-beta hydroxysteroid dehydrogenase III deficiency and adrenal hyperplasia type 3 or congenital adrenal hyperplasia due to 21-hydroxylase deficiency are listed in Table 2, while down regulated genes were mainly enriched in catecholamine biosynthesis, noradrenaline and adrenaline degradation, ascorbate and aldolase metabolism, steroid hormone biosynthesis, signaling mediated by p38-gamma and p38-delta, aurora A signaling, biological oxidations, the citric acid (TCA) cycle and
respiratory electron transport, bile acid biosynthesis, valine-leucine and isoleucine degradation, angiotensin-converting enzyme 2 regulates heart function, apoptotic DNA fragmentation and tissue homeostasis, nicotine degradation, serine glycine biosynthesis, renin-angiotensin system signaling, histidine metabolic, fatty acid elongation in mitochondria and gluconeogenesis are listed in Table 3.

**Gene ontology (GO) enrichment analysis**

GO term enrichment analysis results varied from GO classification and expression change of up and down regulated genes are listed in Table 4 and Table 5. As to BP, the up regulated genes significantly enriched in response to biotic stimulus and defense response, and the down regulated genes significantly enriched in oxoacid metabolic process, and monocarboxylic acid metabolic process. For CC, the up regulated genes significantly enriched in cell surface and synapse, and the down regulated genes significantly enriched in mitochondrion and polymeric cytoskeletal fiber. About MF, the up regulated genes significantly enriched in cytokine activity and signaling receptor binding and the down regulated genes significantly enriched in cofactor binding and oxidoreductase activity.

**Protein-protein interaction network construction and module analysis**

PPI network of up regulated genes, consisting of 5698 nodes and 11384 edges, was constructed by Cytoscape software, based on HIPPIE database (Fig. 5). The up and down regulated genes with highest node degree, betweenness centrality, stress centrality, closeness centrality and lowest clustering coefficient were selected as the hub genes of SARS-CoV-2 infections and are listed in Table 6. The scatter plots of node degree, betweenness centrality, stress centrality, closeness centrality and clustering coefficient for up regulated genes are shown in Fig. 6. The up regulated hub genes (JUN, FBXO6, CYLD, IFI16, ADRB2, PML, ZC3HAV1, SYNPR, CH25H, LSMEM1, CRTAM and RAET1L) in this PPI network were significantly enriched in TNF signaling pathway, RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways, NOD-like receptor signaling pathway, cAMP signaling pathway, influenza A, response to biotic stimulus, IL-17 signaling pathway, cytokine-cytokine receptor interaction, positive regulation of RNA metabolic process, transition metal ion binding and response to endogenous stimulus. Similarly, PPI network of down regulated genes, consisting of 5712 nodes and 11002 edges (Fig. 7). The scatter plots of node degree, betweenness centrality, stress centrality, closeness centrality and clustering coefficient for down regulated genes are shown in Fig. 8. The down regulated hub genes (PCLAF, CFTR, TUBA1A, AURKA, ARRB1, KPNA2, PRDX2, PGK1, GALM, ACYP2, HSD17B2, CEACAM7 and SEMA3E) in this PPI network were significantly enriched in microtubule cytoskeleton, peptidase activity, metabolism of proteins, aurora A signaling, morphine addiction, regulation of cytoplasmic and nuclear SMAD2/3 signaling, oxidation-reduction process, metabolic pathways, protein homodimerization activity, identical protein binding, steroid hormone biosynthesis and genes encoding proteins affiliated structurally or functionally to extracellular matrix proteins.

PEWCC1 was used to identified the total 1503 modules in the PPI network of up regulated genes and the four significant modules were selected (Fig. 9). Pathway and GO enrichment analysis, module 25 (nodes 92 and edges 173), module 33 (nodes 75 and edges 153), module 37 (nodes 63 and edges 109) and
module 70 (nodes 20 and edges 41) revealed that the four significant modules were mainly enriched in cytokine-cytokine receptor interaction, influenza A, Jak-STAT signaling pathway, inflammation mediated by chemokine and cytokine signaling pathway, response to cytokine, response to biotic stimulus, response to virus and response to endogenous stimulus. Similarly, PEWCC1 was used to identified the total 1405 modules in the PPI network of down regulated genes and the four significant modules were selected (Fig. 10). Pathway and GO enrichment analysis, module 14 (nodes 111 and edges 135), module 15 (nodes 107 and edges 121), module 27 (nodes 81 and edges 97) and module 45 (nodes 54 and edges 75) revealed that the four significant modules were mainly enriched in hedgehog signaling events mediated by Gli proteins, metabolism of proteins, metabolism of lipids and lipoproteins, the citric acid (TCA) cycle and respiratory electron transport, mitochondrion, growth, intrinsic component of plasma membrane and cellular lipid metabolic process.

**Construction of target genes - miRNA regulatory network**

To further understand the regulatory network between miRNA and target genes, various miRNA databases such as TarBase, miRTarBase, miRecords, miR2Disease, HMDD, PhenomiR, SM2miR, PharmacomiR, EpimiR and starBase through miRNet were constructed by Cytoscape. As shown in Fig. 11, the target genes - miRNA regulatory network for up regulated genes with 5667 edges and 2144 nodes was obtained for target genes. Different up regulated target genes regulated by miRNA are shown in Table 7, which top five miRNAs such as hsa-mir-4511, hsa-mir-3924, hsa-mir-4478, hsa-mir-3650 and hsa-mir-4252 had been predicted to regulate target genes such as TXNIP, PMAIP1, APOL6, CHAC1 and KLF2. Pathway and GO enrichment analysis revealed that target genes in this network were mainly involved in NOD-like receptor signaling pathway, Direct p53 effectors, signaling receptor binding and response to cytokine. Similarly, as shown in Fig. 12, the target genes - miRNA regulatory network for down regulated genes with 6679 edges and 2236 nodes was obtained for target genes. Different down regulated target genes regulated by miRNA are shown in Table 7, which top five miRNAs such ashsa-mir-4312, hsa-mir-4527, hsa-mir-4673, hsa-mir-4496 and hsa-mir-3153 had been predicted to regulate target genes such as BRI3BP, FAHD1, CPM, HOXA13 and TMBIM6. Pathway and GO enrichment analysis revealed that target genes in this network were mainly involved in mitochondrion, metabolic pathways, metabolism of proteins and ion transport.

**Construction of target genes - TF regulatory network**

To further understand the regulatory network between TFs and target genes, JASPAR through NetworkAnalyst were constructed by Cytoscape. As shown in Fig. 13, the target genes - TF regulatory network for up regulated genes with 2844 edges and 439 nodes was obtained for target genes. Different up regulated target genes regulated by TF are shown in Table 8, which top five TFs such as FOXC1, GATA2, YY1, FOXL1 and NFKB1 had been predicted to regulate target genes such as PROX1, CXCL11, C4A, JAK2 and IL15RA. Pathway and GO enrichment analysis revealed that target genes in this network were mainly involved in viral genome replication, cytokine-cytokine receptor interaction, innate immune system, influenza A and Jak-STAT signaling pathway. Similarly, as shown in Fig. 14, the target genes - TF regulatory network for down regulated genes with 3191 edges and 505 nodes was obtained for target
genes. Different down regulated target genes regulated by TF are shown in Table 8, which top five TFs such as FOXC1, GATA2, YY1, NFIC and FOXL1 had been predicted to regulate target genes such as SERHL2, CFI, TCN1, PGP and PGK1. Pathway and GO enrichment analysis revealed that target genes in this network were mainly involved in mitochondrion, peptidase activity, drug metabolic process, metabolic pathways and carbon metabolism.

**ROC analysis**

As these five up and five down regulated genes are prominently expressed in SARS-CoV-2 infection, we performed a ROC curve analysis to evaluate their sensitivity and specificity for the diagnosis of SARS-CoV-2 infection. As shown in Fig. 15, DDX58, IFITM2, IRF1, PML, SAMHD1, ACSS1, CYP2U1, DDC, PNMT and UGT2A3 achieved an AUC value of 0.777, 0.857, 0.929, 0.804, 0.759, 0.714, 0.728, 0.754, 0.714 and 0.763, demonstrating that these genes have high sensitivity and specificity for SARS-CoV-2 infection diagnosis. The results suggested that DDX58, IFITM2, IRF1, PML, SAMHD1, ACSS1, CYP2U1, DDC, PNMT and UGT2A3 can be used as biomarkers for the diagnosis of SARS-CoV-2 infection.

**Discussion**

SARS-CoV-2 infection transmits worldwide and its mortality rate has increased in recent years. This research represents the first complete study of a genes linked with SARS-CoV-2 virus infection, using bioinformatics analysis. However, the molecular mechanisms of SARS-CoV-2 infection remain poorly understood. Thus, probable biomarkers for diagnosis and treatment with high competence are crucially demanded. Microarray technology has been proved to be a needful approach to diagnose novel biomarkers in SARS-CoV-2 infection.

In the current investigation, expression profiling by high throughput sequencing dataset was analyzed to obtain DEGs between SARS-CoV-2 infections samples and mock samples. A total of 928 DEGs were identified, including 461 up regulated genes and 467 down regulated genes. Genes such as SOCS3 [47] and IL1A [48] were liable for progression of influenza virus infection, but these genes may be linked with progression of SARS-CoV-2 infection. Genes such as KLF6 [49], DUSP1 [50] and OLFM4 [51] were associated with development of respiratory syncytial virus infection, but these genes may be responsible for advancement of SARS-CoV-2 infection. IL6 was involved in progression of SARS-CoV-2 infection [52]. Genes such as MAGT1 [53], CD24 [54] and UGT1A7 [55] were involved in progression of various viral infections, but these genes may be associated with development of SARS-CoV-2 infection. AQP1 was linked with development of porcine reproductive and respiratory syndrome virus infection [56], but this gene may be responsible for progression of SARS-CoV-2 infection.

In the pathway enrichment analysis, many up and down regulated genes may play a very key role in progression of SARS-CoV-2 infection. Enriched genes such as CXCL9 [57], CX3CL1 [58], IFNB1 [59], IL7 [60], TNFSF10 [61], IFNL3 [62], EGR1 [63], USP18 [64], DDX58 [65], UBA7 [66], IFI35 [67], IFIT2 [68], IFIT1 [69], TRIM5 [70], RSAD2 [71], XAF1 [72], MX2 [73], GBP1 [74], TRIM21 [75], FGF19 [76], PELI1 [77], OAS1 [78], DUSP10 [79], TLR6 [80], TICAM1 [81], DDC (dopa decarboxylase) [82], UGT1A1 [83], STMN1 [84],
CYP27A1 [85] and ACE (angiotensin I converting enzyme) [86] were linked with advancement of various viral infections, but these genes may responsible for progression of SARS-CoV-2 infection. Enriched genes such as CCL2 [87], FLT3 [88], CXCL8 [89], IL11 [90], CXCL10 [91], TNF (tumor necrosis factor) [92], EGR2 [93], TRIM22 [94], PML (promyelocyticleukemia) [95], ICAM1 [96], IFITM1 [97], IFITM2 [98], ISG15 [99], SOCS1 [100], IRF1 [101], IRF7 [102], ISG20 [103], STAT2 [104], OAS2 [105], HERC5 [106] and TNFAIP3 [107] were associated with advancement of influenza virus infection, but these genes may be liable for development of SARS-CoV-2 infection. Enriched genes such as CCL20 [108], TSLP (thymic stromal lymphopoietin) [109], IL7R [110], IFIT3 [111] and OASL (2'-5'-oligoadenylate synthetase like) [112] were involved in the development of respiratory syncytial virus infection, but these genes may be associated with development of SARS-CoV-2 infection. SAMHD1 was involved in progression of porcine reproductive and respiratory syndrome virus infection [113], but this gene may be liable for progression of SARS-CoV-2 infection. Enriched genes such as JAK2 [114] and ACE2 [115] were responsible for progression of SARS-CoV-2 infection. As given in the pathway enrichment analysis for up and down regulated genes, HSD17B3, CYP21A2, TNFRSF4, CXCL11, FLT3LG, CSF2, TNFSF13B, IL12A, IL15RA, INHBA (inhibin subunit beta A), IL17C, IL22RA1, BMP2, CXCL2, CXCL3, LTB (lymphotoxin beta), IFNL2, IFNL1, FOS (Fos proto-oncogene, AP-1 transcription factor subunit), GATA3, PTGS2, JUN (Jun proto-oncogene, AP-1 transcription factor subunit), EGR3, EGR4, UBE2L6, HLA-F, SPTBN5, RASGRP3, TRIM38, KLB (klotho beta), MAP3K8, MT2A, PSMB9, FGF18, DUSP8, RELB (RELB proto-oncogene, NF-kb subunit), LGALS9, RASGRP1, FST (follistatin), WIFI1, NFKBIA (NFKB inhibitor alpha), PNMT (phenylethanolamine N-methyltransferase), ALDH2, UGT1A10, UGT1A8, UGT1A6, UGT1A5, UGT1A9, UGT1A4, UGT1A3, UGT2A3, SNTA1, EPHX1, ADH6, FDX1, ACSS1, GSTO2, GGT1, NR1H4, GLYATL2, GSTA4, CYP2U1, CYP4F3, HADHB (hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta), AGT (angiotensinogen), MME (membrane metalloendopeptidase), PPT1 and PRDX3 are novel biomarkers for SARS-CoV-2 infection.

Similarly, GO enrichment analysis, many up and down regulated genes may play a very key role in progression of SARS-CoV-2 infection. Enriched genes such as IFNL4 [116], ZFP36 [117], CD274 [118], MIR21 [119], CYP1A1 [120], CMPK2 [121], C4A [122], SERPINE1 [123], CCRL2 [124], CD69 [125], FAH (fumarylacetoacetate hydrolase) [126], ERFE (erythropherrone) [127], SERINC5 [128], ADI1 [129], AP2M1 [130], PRIMPOL (primase and DNA directed polymerase) [131], BIRC5 [132] and TF (transferrin) [133] were responsible for advancement of various viral infections, but these genes may be involved in progression of SARS-CoV-2 infection. Enriched genes such as APOBEC3G [134], NLRC5 [135], PTX3 [136], FAP (fibroblast activation protein alpha) [137] and CAT (catalase) [138] were linked with development of influenza viral infection, but these genes may be associated with progression of SARS-CoV-2 infection. Enriched genes such as IFI16 [139], TAP1 [140] and PIN1 [141] were involved in development of porcine reproductive and respiratory syndrome virus infection, but these genes may be associated with progression of SARS-CoV-2 infection. NLRP3 was key for advancement of SARS-CoV-2 infection [142]. PDK1 was liable for development of SARS corona virus infection [143], but this gene may be linked with advancement of SARS-CoV-2 infection. As given in the GO enrichment analysis for up and down regulated genes, PDE2A, PDE4B, NT5C3A, GGLYP2, DTX3L, PLAC8, NOCT (nocturnin), PMAIP1,
ZC3H12A, DHX58, TENT5A, IFI44, BATF2, TXNIP (thioredoxin interacting protein), SH2D1B, MIR181B1, RAET1L, STX11, ZC3HAV1, CYLD (CYLD lysine 63 deubiquitinase), TAGAP (T cell activation RhoGTPase activating protein), BCL3, PTGER4, RTP4, GFI1, IFI1H1, CLEC4A, GLI2, C4B, GPER1, CRTAM (cytotoxic and regulatory T cell molecule), PCK1, acetylcholinesterase (Cartwright blood group), TRPC4, DSCAML1, HEG1, PLAUR (plasminogen activator, urokinase receptor), SLITRK6, ACKR4, LAYN (layilin), GRIA2, MMP25, EPFA4, SECTM1, VSTM1, RIMKLA (ribosomal modification protein rimK like family member A), ACO1, ETFB (electron transfer flavoprotein subunit beta), ETFDH (electron transfer flavoprotein dehydrogenase), MDH1, HNMT (histamine N-methyltransferase), PGK1, FABP6, ACSM3, FBPI, HOGA1, SCPE2, ECI2, ALDH6A1, PCY01X1, MPC2, HACD3, FOLR1, SHMT1, ACSF2, SLC3A1, CRYD (crystallin mu), CSPG5, PPA2, AKR1C3, DHRP1, SPR (sepiapterinreductase), BDH2, GATM (glycine amidinotransferase), BCKDHB (branched chain keto dehydrogenase E1 subunit beta), PSMD10, ITIH2, PHGDH (phosphoglycerate dehydrogenase), BTD (bistidinase), SLC25A21, OAT (ornithine aminotransferase), OAZ1, HACD2, ECH1, BC02, BNIPL (BCL2 interacting protein like), PAH (phenylalanine hydroxylase), LTC4S, DSEL (dermatansulfateepimerase like), ATPSCKMT (ATP synthase c subunit lysine N-methyltransferase), COX16, ACYP2, TRMT2B, TSTD1, ASB9, UCP1, AK4, CEBPSOS (CEBPZ opposite strand), SERHL2, HSPE1, SLC25A27, COA3, AGR2, MACROD1, COX11, NUBPL (nucleotide binding protein like), CRLS1, BRI3BP, DNLZ (DNL-type zinc finger), CYBA (cytochrome b-245 alpha chain), KLL6, NDUFS7, PSAP (proaposin), NDUFA8, ACOT13, SRI (sorcin), TOMM20, NDUFA2, SGPP1, TMEM160, HEBP1, AKR1B10, DPYSL2, ATP5IF1, NRG1 (neurogranin), DTYMK (deoxythymidylate kinase), NTHL1, PRDX2, TMBIM6, GSR (glutathione-disulfidereductase), TFRC (transferrin receptor), ANXA10, SAMD13, FAHD1, ENDOG (endonuclease G), COX14, FMO5, CRYZ (crystallin zeta), ACBD7, KCNAB1, CREG2 and TCN1 are novel biomarkers for SARS-CoV-2 infection.

PPI network was constructed and modules extracted to explore the molecular mechanism SARS-CoV-2 infection. In this investigation, the expression of hub genes in SARS-CoV-2 infection. Hub gene CH25H was liable for progression of porcine reproductive and respiratory syndrome virus infection [144], but this gene may be involved in advancement of SARS-CoV-2 infection. Hub genes such as TRAF1 [145], IFIT5 [146], CFTR (CF transmembrane conductance regulator) [147] and SMARCA2 [148] were key for progression of influenza viral infection, but these genes may be linked with progression of SARS-CoV-2 infection. Hub gene SAP30 was responsible for advancement of Rift Valley fever virus infection [149], but this gene may be key for advancement of SARS-CoV-2 infection. Hub gene APOE (apolipoprotein E) was responsible for development of SARS-CoV-2 infection [150]. As given in the PPI network and module analysis for up and down regulated genes, FBXO6, ADRB2, SYNPR (synaptoporin), LSMEM1, ARRDC3, ACTA1, LMO2, ATF3, KLF4, PCLAF (PCNA clamp associated factor), TUBA1A, AURKA (aurora kinase A), ARRB1, KPNA2, GALK (galactosemutarotase), HSD17B2, CEACAM7, SEMA3E, EPN3, ROM1, SORL1 and PDCD4 are novel biomarkers for SARS-CoV-2 infection.

Target gene - miRNA regulatory network and target gene - TF regulatory network analysis (up and down regulated genes) can be regarded as key to the understanding of SARS-CoV-2 infection and might also lead to new therapeutic approaches. Hub genes such as PROX1 [151] and HOXA13 [152] were responsible for development of hepatitis B virus infection, but these genes may be essential for advancement of for
SARS-CoV-2 infection. As given in the target gene - miRNA regulatory network and target gene - TF regulatory network for up and down regulated genes, APOL6, CHAC1, KLF2, CPM (carboxypeptidase M), CFI (complement factor I) and PGP (phosphoglycolate phosphatase) are novel biomarkers for SARS-CoV-2 infection.

**Conclusion**

The current investigation provided a comprehensive bioinformatics analysis of up and down regulated genes in SARS-CoV-2 infection. Analysis of these genes provided information regarding the molecular mechanisms of SARS-CoV-2 infection and key biomarkers or targets for the diagnosis and treatment of SARS-CoV-2 infection. However, further molecular biological experiments are required to confirm the action of the up and down regulated genes and pathways in different types of SARS-CoV-2 infection.

**Declarations**

**Acknowledgements**

We also thank Emanuel Wyler, Max Delbrück Center for Molecular Medicine, Berlin Institute for Medical Systems Biology, Robert Roessle Str 10, Berlin, Germany., very much, the author who deposited their expression profiling by high throughput sequencing dataset, GSE148729, into the public GEO database.

**Conflict of interest**

The authors declare that they have no conflict of interest.

**Ethical approval**

This article does not contain any studies with human participants or animals performed by any of the authors.

**Informed consent**

No informed consent because this study does not contain human or animals participants.

**Author Contributions**

B. V - Writing original draft, and review and editing

C. V - Software and investigation

**Availability of data and materials**

The datasets supporting the conclusions of this article are available in the GEO (Gene Expression Omnibus) (https://www.ncbi.nlm.nih.gov/geo/) repository. [(GSE148729)]

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors

Basavaraj Vastrad ORCID ID: 0000-0003-2202-7637
Chanabasayya Vastrad ORCID ID: 0000-0003-3615-4450

References


50. Robitaille, A.C.; Caron, E.; Zucchi, N.; Mukawera, E.; Adam, D.; Mariani, M.K.; Gélinas, A.; Fortin, A.; Brochiero, E.; Grandvaux, N. DUSP1 regulates apoptosis and cell migration, but not the JIP1-protected
cytokine response, during Respiratory Syncytial Virus and Sendai Virus infection. *Rep.* 2017,7,17388. doi:10.1038/s41598-017-17689-0


111. Ternette, N.; Wright, C.; Kramer, H.B.; Altun, M.; Kessler, B.M. Label-free quantitative proteomics reveals regulation of interferon-induced protein with tetratricopeptide repeats 3 (IFIT3) and 5'3'-exoribonuclease 2 (XRN2) during respiratory syncytial virus infection. _J._ 2011, 8, 442. doi:10.1186/1743-422X-8-442


120. Anderson, G.R.; Hasan, A.; Yin, H.; Qadri, I.; Quattrocchi, L.C. Regulation of the CYP1A1 gene by 2,3,7,8-tetrachlorodibenzo-p-dioxin but not by beta-naphthoflavone or 3-methylcholanthrene is altered in


Tables

Tables 1-8 are available in the Supplemental Files section

Figures
Figure 4

Heat map of downregulated differentially expressed genes. Legend on the top left indicate log fold change of genes. (A1 – A8 = Mock samples; B1 – B14 = SARS-CoV-2 infection samples)

Figure 5

Protein–protein interaction network of upregulated genes. Green nodes denotes upregulated genes.
Figure 7

Protein–protein interaction network of down regulated genes. Red nodes denotes down regulated genes

Figure 8

Scatter plot for down regulated genes. (A- Node degree; B- Betweenness centrality; C- Stress centrality; D- Closeness centrality; E- Clustering coefficient)
Figure 9

Modules in PPI network. The green nodes denote the up regulated genes.

Figure 10
Modules in PPI network. The red nodes denote the down regulated genes.

**Figure 13**

The network of up regulated genes and their related TFs. The green circles nodes are the up regulated genes, and purple triangle nodes are the TFs.

**Figure 14**

The network of down regulated genes and their related TFs. The green circles nodes are the down regulated genes, and blue triangle nodes are the TFs.