Identification of key genes and pathways in the hPSC-derived lungs infected by the SARS-CoV-2

Hanming Gu
Shanghai Jiao Tong University

Gongsheng Yuan (gsyuan14@fudan.edu.cn)
Fudan University

Research Article

Keywords: SARS-CoV-2, COVID-19, lung, DEGs, brain

DOI: https://doi.org/10.21203/rs.3.rs-114578/v1

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

Coronavirus disease 2019 (COVID-19) is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has led to numerous infections and deaths in the world. Our research is to explore the differentially expressed genes (DEGs) and signaling pathways in hPSC-derived lungs by using a bioinformatics method to clarify their potential pathogenesis. The gene expression profile of GSE155241 dataset was originally created by using an Illumina NovaSeq 6000 (Homo sapiens) platform. Functional categories and significant pathways were identified by the KEGG and GO analysis. The results suggested that brain disorders and mitochondrial dysfunctions are the main signaling pathways affected by the SARS-CoV-2 infection. Furthermore, key genes e.g. CDC20, NCBP1 and inhibitors e.g. MEK1-2-inhibitor, tivozanib may play critical roles in COVID-19. Therefore, our study provides insights into the treatment of COVID-19 and related disorders.

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infects epithelial cells of the upper respiratory tract and the lungs\(^1,2\). Coronaviruses belong to the family of viruses which contain positive single RNA genome\(^3\). The transmission to human might have happened through direct contact of bats but some institutes are claiming the existence of the intermediate hosts\(^4\). For SARS-CoV-2, there is genomic similarity to the pangolins and bats\(^5\). However, the fatality rate of SARS-CoV-2 is less than those of other known CoVs\(^6\).

Cytokine storm is an important stage in worsening viruses associated with lung infection\(^7\). A large amount of IL-10, G-CSF, MIP1A, MCP1, and TNF\(\alpha\) were reported to exist in the critical patients in ICU\(^8\). Drugs targeting the cytokines may become an ideal strategy for treating the Coronavirus disease-2019 (COVID-19)\(^9\). Different treatments can be used in handling the respiratory infection\(^10\). For example, the use of JAK inhibitors and FcR inhibitors could decrease the virus infection and inflammation\(^11\). Furthermore, the drugs that can directly target the specific and non-specific process of COVID-19 may also play a key role in the treatment\(^12\). Some of them have been tested in different diseases such as rheumatoid arthritis (RA) and cancers\(^13\). Numerous trials are currently under study in decreasing the cytokine storm of COVID-19\(^14\).

Here, we studied the difference between the virus infected lung tissues which were iPSC differentiated and those normal ones. We analyzed and discovered several differentially expressed genes (DEGs) and the relevant biological process and function by performing comprehensive bioinformatics analysis. We utilized the functional enrichment, pathway analysis, and protein-protein interaction (PPI) for finding target genes in infected lungs by comparing to non-infection ones. These key genes could be critical to guide future therapeutic interventions for COVID-19.
Methods

Data collection

Gene expression profile dataset GSE152418 was downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/). The data was produced by using an Illumina NovaSeq 6000 (Homo sapiens) (Developmental and Cognitive Neuroscience, Yerkes National Primate Research Center, Atlanta, GA30329-4208, US). The GSE152418 dataset contained data including 17 COVID-19 subjects and 17 healthy controls.

Data acquisition and preprocessing

The raw microarray data between SARS-CoV-2 positive samples and negative controls were subsequently conducted by R script as previously described\textsuperscript{15-17}. We set a classical t test to identify DEGs with $P<.01$ and fold change $\geq 1.5$ as being statistically significant.

Gene ontology (GO) and pathway enrichment analysis

Gene ontology (GO) analysis is a widely used approach to annotate genomic data and identify characteristic biological information. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database is commonly used for systematic analysis of gene functions and annotation of biological pathways. GO analysis and KEGG pathway enrichment analysis of DEGs in this study were analyzed by the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (http://david.ncifcrf.gov/). $P<.05$ and gene counts $>10$ were considered statistically significant.

Module analysis

The Molecular Complex Detection (MCODE) was used to analyze the densely connected regions in PPI networks. The significant modules were from the constructed PPI network using MCODE. the function and pathway enrichment analyses were performed by using DAVID, and $P<.05$ was used as the cutoff criterion.

Results

Identification of DEGs of SARS-CoV-2 infected lungs in comparison to controls

To get the insights on the difference between SARS-CoV-2 infected lungs and control, the modular transcriptional signature was compared. A total of 1020 genes were identified to be differentially
expressed in SARS-CoV-2 infected lungs with the threshold of \( P < 0.05 \). The top 10 up- and down-regulated genes are listed in Table 1.

**Enrichment analysis of SARS-CoV-2 infected lungs in comparison to controls**

To further analyze the potential mechanisms of DEGs of SARS-CoV-2 infected lungs, the GO categories enrichment analysis and KEGG pathway were utilized. The Gene ontology (GO) is a structured vocabulary for the classification of gene function and a powerful resource for drug development.

We identified top 5 biological processes: “Generation of precursor metabolites and energy”, “Oxidative phosphorylation”, “Ion transmembrane transport”, “ATP synthesis coupled proton transport”, and “Down electrochemical gradient” (Figure 1). We also identified the cellular components including “Mitochondrion”, “Mitochondrial part”, “Mitochondrial envelope”, “Mitochondrial membrane”, and “Organelle inner membrane” (Figure 1). We then identified top 5 molecular functions: “Structural constituent of ribosome”, “Manganese ion binding”, “Inorganic cation transmembrane transporter activity”, “Monovalent inorganic cation transmembrane transporter activity”, and “Hydrogen ion transmembrane transporter activity” (Figure 1).

KEGG pathway (http://www.genome.jp/kegg/) is a collection of pathway maps representing the molecular interaction networks such as metabolism, genetic information processing, environmental information processing, and drug development. Our study showed top 5 enriched KEGG pathways including “Alzheimer's disease”, “Ribosome”, “Huntington's disease”, “Parkinson's disease”, and “Oxidative phosphorylation” (Figure 1).

**PPI network and Module analysis**

The PPI network was for exploring the relationship of DGEs at the protein level. We set the criterion of combined score > 0.7 and created the PPI network by using the 384 nodes and 1384 interactions between SARS-CoV-2 infected lungs and controls. Among these nodes, the top 10 genes with highest scores are selected in Table 2.

The top two significant modules of SARS-CoV-2 infected lungs and controls were selected to show the functional annotation of genes (Figure 2 and 3). We then identified top five signaling pathways in module 1: Antigen processing: Ubiquitination & Proteasome degradation, Neddylation, Post-translational protein modification, Metabolism of proteins, Association of TriC/CCT with target, Inactivation of APC/C via direct, Conversion from APC/C:Cdc20 to APC/C:Cdh1 in late anaphase, APC/C:Cdc20 mediated degradation of Cyclin B, APC-Cdc20 mediated degradation of Nek2A, and APC/C:Cdc20 mediated degradation of Securin. We also identified five top signaling pathways in module 2: mRNA Splicing - Major Pathway, mRNA Splicing - Minor Pathway, Neddylation, mRNA decay by 5' to 3' exoribonuclease,
Processing of Intronless Pre-mRNAs, Formation of TC-NER Pre-Incision Complex, mRNA 3'-end processing, Cleavage of Growing Transcript in the Termination Region, RNA Polymerase II Transcription Termination, and Signaling by FGFR2 Illa TM by using Reactome Pathway Database (https://reactome.org/) (Supplemental Table S1 and Table S2).

**Potential inhibitors for COVID-19 in lungs**

We highlighted top ten inhibitors with the high scores identified by the L1000FWD analysis (Figure 4). Among them: MEK1-2-inhibitor blocks the mitogen-activated protein kinase kinase enzymes MEK1 and MEK2; tivozanib is a selective VEGFR inhibitor; daunorubicin interacts and inhibits the macromolecular biosynthesis; Vemurafenib is an inhibitor of the B-Raf enzyme; BRD-K83670234 is a potential Aurora kinase and PI3K inhibitor; CGP 57380 is a inhibitor of MAP kinase-interacting kinase 1; CGP-60474 is an inhibitor of cyclin-dependent kinase; Bisindolylmaleimide IX is a cell-permeable inhibitor of protein kinase C; WZ-3105 is the inhibitor of both NTRK and the SRC kinase family; AS-605240 selectively inhibits PI3Kγ.

**Discussion**

COVID-19, a huge threaten caused by SARS-CoV-2, has affected a large number of people and the death toll has reached hundreds of thousands worldwide\textsuperscript{15,17,18}. The trachea which connects the upper airway to the lung becomes a high virus contact zone. Viral loads are considered as the key point of the severity of COVID-19. People with more SARS-CoV-2 infections showed more loads and longer persistence in the respiratory tract than milder ones\textsuperscript{19}. ACE2 expression in the lung was approved to play a critical role in COVID-19 animal models\textsuperscript{20}. Our results report several findings to the growing understanding about the mechanism of SARS-CoV-2 infections.

Our KEGG analysis showed COVID-19 is the direct cause of brain disorders. Our study discovered that genes in many brain disorders such as Alzheimer's disease, Huntington's disease, and Parkinson's disease are significantly changed in the process of COVID-19. These disorders share a late onset in mid-adulthood or later time\textsuperscript{21}. Reports showed the early events might set the stage for later disease. For example, Alzheimer’s disease impairs cognition, memory, and dementia in late age, which is caused by the early issues\textsuperscript{22}. Thus, it is suggested that the infection of SARS-CoV-2 in early times such as childhood can be the trigger in the old ages. We therefore recommend that all ages should wear masks and maintain social distance, especially in environments that are currently considered to have no impact on children.

Our GO analysis suggested that COVID-19 affects the mitochondrial function. Based on our data, we found the genes were changed mainly in mitochondria including mitochondrial envelope and mitochondrial membrane. Mitochondria are the energy factories in cells and can be regulated by several
proteins and pathways, which is most related to the oxidative phosphorylation, apoptosis, cell death and survival\textsuperscript{23-27}. In addition to reactive oxygen species (ROS)-based regulation in immunity, mitochondria also take part in antiviral defense\textsuperscript{28}. Mitochondria play crucial roles in anti-viral issues and activate the innate immune system via DAMPs\textsuperscript{29}. Moreover, SARS-CoV-2 proteins regulate the mitochondrial system to block the immune response and may change the mitochondrial functions to promote the ROS production and signaling\textsuperscript{30}. Metabolism disorders such as obesity rank among the highest risk scores affecting COVID-19 diagnosis, ICU admission and mortality based on the meta-analysis studies\textsuperscript{31}. Obesity that can be regulated by several molecules such as circadian genes also enhances the severity of COVID-19 to young ages\textsuperscript{32-35}. Indeed, our GO analysis showed the SARS-CoV-2 infection can affect the generation of precursor metabolites and energy, ion transmembrane transport and ATP synthesis coupled proton transport.

In our study, several DEGs were recognized by the PPI network which could be considered as drug targets. The HTLV-1 is able to promote securin and cyclin B1 degradation through the accessory factor CDC20 of APC\textsuperscript{36}. NCBP3 is a pivotal under virus infection which can inhibit the virus growth\textsuperscript{37}. NF-κB is a key regulator during inflammation and involves in different inflammatory diseases\textsuperscript{38-40}. POLR2D is as an inflammatory molecule which can also be regulated by NF-κB pathways\textsuperscript{41}. DYNLL1 interacts with viral proteins by the dynein chains\textsuperscript{42}. FBXW5 facilitates the addition of Lys63-linked ubiquitin to enhance the inflammation\textsuperscript{43}. Protein quality control maintains the homeostasis of cells including Endoplasmatic reticulum-associated degradation (ERAD) and Ub-associated degradation\textsuperscript{44-46}. LRRC41 is a predominant nuclear dimeric protein which is involved in the protein degradation process such as ubiquitination\textsuperscript{47}. The lack of Fbxo21 impairs virus-induced activation of ASK1\textsuperscript{48}. FBXW9 regulates carcinogenesis via regulating the factors involved in cell cycle progression\textsuperscript{49}. FBXO44 Mediates BRCA1 Ubiquitination and protein metabolism\textsuperscript{50}. The activation of IFN-I signaling can be reversed by the FBXO6\textsuperscript{51}. Our analysis further discovered that antigen processing (Ubiquitination & Proteasome degradation) and mRNA Splicing are involved in the SARS-CoV-2 infection by analyzing the Reactome Pathway Database.

To sum up, we provided the basis for the biomarkers during SARS-CoV-2 infection in lungs. Brain disorders and mitochondrial dysfunctions were mostly related in COVID-19. This study provides further insights into the treatment of COVID-19.

**Declarations**

**Competing interests:**

The authors declare no competing interests.

**References**


Tables
<table>
<thead>
<tr>
<th>Entrez gene</th>
<th>Gene Symbol</th>
<th>Fold-change</th>
<th>Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10 down-regulated DEGs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27097</td>
<td>TAF5L</td>
<td>-1.818760446</td>
<td>Down</td>
</tr>
<tr>
<td>127391</td>
<td>TMCO2</td>
<td>-1.817160181</td>
<td>Down</td>
</tr>
<tr>
<td>57827</td>
<td>C6orf47</td>
<td>1.816969595</td>
<td>Down</td>
</tr>
<tr>
<td>22917</td>
<td>ZP1</td>
<td>-1.815479751</td>
<td>Down</td>
</tr>
<tr>
<td>146434</td>
<td>ZNF597</td>
<td>-1.814930811</td>
<td>Down</td>
</tr>
<tr>
<td>51088</td>
<td>KLHL5</td>
<td>-1.814353104</td>
<td>Down</td>
</tr>
<tr>
<td>8874</td>
<td>ARHGEF7</td>
<td>-1.813107418</td>
<td>Down</td>
</tr>
<tr>
<td>132720</td>
<td>C4orf32</td>
<td>-1.812761468</td>
<td>Down</td>
</tr>
<tr>
<td>262</td>
<td>AMD1</td>
<td>-1.812479327</td>
<td>Down</td>
</tr>
<tr>
<td>654</td>
<td>BMP6</td>
<td>-1.810848576</td>
<td>Down</td>
</tr>
<tr>
<td>Top 10 up-regulated DEGs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55577</td>
<td>NAGK</td>
<td>1.823564136</td>
<td>up</td>
</tr>
<tr>
<td>3074</td>
<td>HEXB</td>
<td>1.822745659</td>
<td>up</td>
</tr>
<tr>
<td>148398</td>
<td>SAMD11</td>
<td>1.819974513</td>
<td>up</td>
</tr>
<tr>
<td>64754</td>
<td>SMYD3</td>
<td>1.81993098</td>
<td>up</td>
</tr>
<tr>
<td>5920</td>
<td>RARRES3</td>
<td>1.819527811</td>
<td>up</td>
</tr>
<tr>
<td>23256</td>
<td>SCFD1</td>
<td>1.819426461</td>
<td>up</td>
</tr>
<tr>
<td>64518</td>
<td>TEKT3</td>
<td>1.81675265</td>
<td>up</td>
</tr>
<tr>
<td>51668</td>
<td>HSPB11</td>
<td>1.816714761</td>
<td>up</td>
</tr>
<tr>
<td>83590</td>
<td>TMUB1</td>
<td>1.816339851</td>
<td>up</td>
</tr>
<tr>
<td>2153</td>
<td>F5</td>
<td>1.815833075</td>
<td>up</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>Gene title</td>
<td>Degree</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>CDC20</td>
<td>cell division cycle 20</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>NCBP1</td>
<td>nuclear cap binding protein subunit 1</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>POLR2D</td>
<td>RNA polymerase II subunit D</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>DYNLL1</td>
<td>dynein light chain LC8-type 1</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>FBXW5</td>
<td>F-box and WD repeat domain containing 5</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>LRRC41</td>
<td>leucine rich repeat containing 41</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>FBXO21</td>
<td>F-box protein 21</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>FBXW9</td>
<td>F-box and WD repeat domain containing 9</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>FBXO44</td>
<td>F-box protein 44</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>FBXO6</td>
<td>F-box protein 6</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>