An approach combining bioinformatics and machine learning to identify eight autophagy-related biomarkers and construct molecular mechanisms underlying COVID-19 and major depressive disorders


1Department of Traditional Chinese Medicine Classics, Shandong University of Traditional Chinese Medicine Affiliated Hospital, Jinan, Shandong, China
2College of Acupuncture and Massage, 3College of Traditional Chinese Medicine, 4First Clinical School of Medicine, Shandong University of Traditional Chinese Medicine, Jinan, China
5Department of Traditional Chinese Medicine, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China
6Department of Clinical Pharmacy, Shaoxing People’s Hospital, Shaoxing, China
7Department of Pediatric Surgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

Cheng Yu, Fengjun Zhang, Lili Zhang, Min Peng, Lin Zhang, and Shuai Wang contributed equally to this work

Abstract. - OBJECTIVE: A lack of objective biomarkers is preventing the screening and diagnosis of COVID-19 combined with major depression disorder (COVID-19-MDD). The purpose of this study was to identify diagnostic biomarkers and gene regulatory mechanisms associated with autophagy; a crucial process significantly involved in the pathogenesis of COVID-19-MDD.

MATERIALS AND METHODS: In this study, differentially expressed genes (DEGs) were screened using GSE98793 from the GEO2R analysis (GEO) database, and intersected with the COVID-19-related gene (CRGs) and autophagy-related genes (ARGs) to obtain common genes involved in. Then, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of these common genes were performed. Subsequently, the transcription factor (TF)—gene regulatory network and comorbidity network were constructed. In addition, 10 drug candidates were screened using the DSigDB database. To identify diagnostic markers, we used LASSO regression.

RESULTS: In total, 13 common genes were screened, which were primarily enriched in lysosomes, endoplasmic reticulum membranes, and other endomembrane systems also associated with autophagy. Additionally, these genes were involved in neurological cell signaling and have a functional role in pathways related to vascular endothelial growth factor, tyrosine kinase, autophagy, inflammation, immunity, and carcinogenesis. Tumors and psychiatric disorders were the most highly linked diseases to COVID-19. Finally, ten drug candidates and eight diagnostic markers (STX17, NRG1, RRAGD, XPO1, HERC1, HSP90AB1, EPHB2, and S1PR3) were screened.

CONCLUSIONS: This is the first study to screen eight diagnostic markers and construct a gene regulatory network for COVID-19-MDD from the perspective of autophagy. The findings of our study provide novel insights into the diagnosis and treatment of COVID-19-MDD.

Key Words: COVID-19, Major depressive disorder, Autophagy, Diagnostic biomarkers, Gene regulatory network.
Introduction

A serious psychiatric disorder that is major depressive disorder (MDD) is among the leading cause of suicide. A study by the World Health Organization (WHO) predicts that depression will be the top-burden disease in the world by 2030. It has been reported that the MDD prevalence among COVID-19 patients ranges between 31%-45% being significantly higher than in the general population. The COVID-19 epidemic has led to a dramatic increase in the number of patients with depression. Consequently, an early screening of patients with COVID-19 combined with depression and thereby proper intervention may help reduce MDD incidence. Currently, depression screening and diagnosis rely primarily on subjective assessments by psychiatrists which lead to higher rates of misdiagnosis. As a result, there is an urgent need for developing objective as well as reliable markers that may aid in screening and diagnosis of depression. Presently, studies on the diagnostic markers of MDD combined with COVID-19 (COVID-19-MDD) primarily focus on three aspects: inflammatory, kynurenine pathway, and growth factors, but their clinical application is restricted due to limited sensitivity and specificity. Therefore, it is crucial to identify potent diagnostic biomarkers for the diagnosis of COVID-19-MDD.

Autophagy, a basic cellular metabolic process known for regulating immune responses by mediating immune cell activity and cytokine release, plays a crucial role in viral infections as well as neurodegenerative diseases. Immune dysregulation functions such as monocyte activation, decreased number and/or activity of T cells as well as the release of pro-inflammatory factors in large amounts are reported as the main key pathological mechanisms for the pathogenesis of MDD. Additionally, immune cells and cytokines activate autophagy. An effect of autophagy on antidepressants or compounds that exert antidepressant-like effects has been reported. There is also some evidence that autophagy-related genes (ARGs) play a role in the diagnosis of MDD. Therefore, autophagy may be an important mechanism in the pathogenesis of MDD with significant implications for its diagnosis and treatment. Additionally, autophagy acts as an essential mechanism in the body’s fight against viruses where coronaviruses may replicate by usurping and exploiting these autophagic mechanisms. It has also been reported that autophagy inducers can antagonize coronavirus replication. Autophagy may, therefore, be involved in COVID-19 pathogenesis. In this study, we attempted to utilize autophagy as an entry point for identifying the pathogenesis and diagnostic markers of COVID-19-MDD thereby generating newer ideas for its good diagnosis and treatment. High-throughput screening is becoming increasingly important in the field of biomedical research. Microarray data analysis is one of the most prominent techniques among high-throughput methods used for large-scale analysis of gene expression. Therefore, analysis of genetic data using high-throughput techniques can help to screen potential diagnostic biomarkers for COVID-19-MDD. In this study, differentially expressed genes (DEGs) for MDD were screened by downloading the GSE98793 dataset from the GEO database and intersected with ARGs and COVID-19-related genes (CRGs) to obtain common genes. Subsequently, gene enrichment analyses of the common genes were performed to clarify the gene regulation mechanism of COVID-19-MDD, construct a transcription factor (TF)-gene regulatory network, and screen for target drugs. Finally, a machine learning algorithm was used to screen for diagnostic biomarkers of COVID-19-MDD.

The flow chart for this study is presented in Figure 1.

Materials and Methods

Data Acquisition

We have utilized the NCBI–GEO database from which the GSE98793 dataset was downloaded as the MDD dataset. This dataset contributed by Kelly et al contains whole blood sample measurements from 128 patients with MDD (64 of them were diagnosed with a generalized anxiety disorder) while the remaining 64 were healthy controls. We have used GPL570/HG-U133_Plus_2 Affymetrix Human Genome U133 Plus 2.0 Array as a platform for the analysis. A search in GeneCard (https://www.genecards.org/) for “COVID-19” yielded a total of 4585 related genes. Here, a total of 803 ARGs were obtained using the Human Autophagy Database (HADB, http://autophagy.lu/) and the Human Autophagy Regulator Database (HAMDB, http://hamdb.scbdd.com).
Identification of DEGs and Common Genes

A DEGs search was conducted on the GSE98793 dataset. In order to analyze the GSE98793 dataset, we used the GEO2RWeb tool (https://www.ncbi.nlm.nih.gov/geo/Geo2r/) and its LIMMA package. In order to screen for DEGs within GSE98793, we used the criterion of $p$-value < 0.05 and $|\log_{2}\text{Fold-Change}| \geq 0.2$. Subsequently, for the visualization of the DEGs, the heatmap package (R package version 4.1.1) and ggrepol2 package (R 4.1.1) were used to draw gradient volcano and heat maps. The common genes obtained from the intersection of DEGs with genes of the other two datasets are presented in a Venn diagram.

Enrichment Analyses of Common Genes

In order to analyze these common genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on these common genes\(^\text{20}\). A GO enrichment analysis of common genes has clarified their cellular component (CC), molecular function (MF), and biological process (BP)\(^\text{21}\). In addition, a KEGG enrichment analysis has been applied for detecting the specific functioning and metabolic pathways of common genes\(^\text{22}\). A viable network platform that is Enrichr (https://amp.pharm.mssm.edu/Enrichr/) is also being utilized for the enrichment analysis. As a final step, we have entered the common genes into the Enrichr web platform to yield GO and KEGG enrichment analysis results\(^\text{23}\).

Construction of the TF-Gene Regulatory Network

DNA-binding proteins, also known as trans-acting factors (TFs) can specifically interact with genes while affecting gene transcription either by posing activating or inhibiting effects. The construction of the TF–gene regulatory network facilitates the identification of pathways through which TFs affect gene expression\(^\text{24}\). Gene expression in numerous species can be analyzed using NetworkAnalyst3.0, a web-based platform. In the NetworkAnalyst 3.0 platform, the TF-gene regulatory network was constructed on the basis of the JASPAR database\(^\text{25}\).

COVID-19-MDD-Related Comorbidities

DisGeNET (http://www.disgenet.org/), a database integrating multiple disease-related genes and variants, covers the expression profiles and normal and abnormal traits of human diseases.
As a result of integrating existing relevant gene or mutation databases and obtaining the relevant information from the literature through machine learning methods, the current version constructs a unified database of disease-related genes and mutation loci covering more than 24,000 diseases and traits, 17,000 genes, and 117,000 genomic variants. The DisGeNET database in NetworkAnalyst 3.0 was used to examine relationships between common genes and diseases for identifying COVID-19-MDD-related comorbidities.

Identification of Drug Candidates

A key component of this study was drug-based molecular characterization to provide new ideas for clinical drug use. We have utilized the Drug Signature Database (DSigDB) on the Enrichr web platform which is a database that basically addresses the repurposing of targeted drugs. The common genes obtained were entered into the Enrichr platform (https://amp.pharm.mssm.edu/enrichr/) for screening common gene-related drug candidates in the DSigDB database.

Identification of Diagnostic Biomarkers Using a Machine Learning Algorithm

To address multicollinearity in regression analysis, the “glmnet” software package was used to screen for autophagy-related diagnostic biomarkers of COVID-19-MDD using the least absolute shrinkage and selection operator (LASSO) regression.

Results

Identification of DEGs in MDD and Their Common Genes with CRGs and ARGs

From an analysis of the GSE98793 dataset using the GEO2RWeb tool and its LIMMA package, we have identified 496 DEGs including 250 up-regulated and 246 down-regulated DEGs. A Heatmap and a gradient volcano plot were shown in Figure 2A–2B demonstrating DEGs. Based on the intersection of MDD DEGs with CRGs and ARGs, around 13 common genes are yielded including STX17, VEGFA, NRG1, RRAGD, NPC1, XPO1, HERC1, HSP90AB1, CAMP, ATM, EPHB2, S1PR3, and TRIM13. The Venn diagram as shown in Figure 3 illustrated the visualization of the common genes among the three datasets.

GO and KEGG Enrichment Analyses

We have applied Enrichr to perform GO and KEGG enrichment analyses of common genes (Table I). A GO enrichment analysis revealed that the common genes are primarily enriched in autophagy-related endomembrane systems such as lysosomes and endoplasmic reticulum membranes being involved in the signaling of nervous system cells while primarily functioning among vascular endothelial growth factors as well as tyrosine kinase (Figure 4A-4C). Further, KEGG enrichment analysis revealed that these common genes are primarily enriched in autophagy-, inflammation-, immune- and carcinogenesis-related pathways as shown in Figure 4D.

Figure 2. Identification of DEGs. A, The heatmap represented the expression of the DEGs of GSE98793. B, Asymptotic volcano plot showing the gene expression of GSE98793. The top 10 genes that met the threshold value (p-value < 0.05 and log2FoldChange (absolute value) of 0.2) were indicated on the plot, only FAM3B was found to be an up-regulated gene and all others were down-regulated genes. The two vertical lines represented gene expression ploidy of -0.2 and 0.2, respectively. The horizontal line indicated a p-value of 0.05. The color of the dots indicated the magnitude of the p-value.
Table I. Gene enrichment analysis (GO and KEGG pathway enrichment analyses).

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>p-value</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>**Biological</td>
<td>Positive regulation of cellular process (GO:0048522)</td>
<td>1.29E-06</td>
<td>HSP90AB1; TRIM13; NRG1; ATM; SIPR3; VEGFA</td>
</tr>
<tr>
<td>Process</td>
<td>Commissural neuron axon guidance (GO:0071679)</td>
<td>1.09E-05</td>
<td>EPHB2; VEGFA</td>
</tr>
<tr>
<td></td>
<td>Positive regulation of intracellular signal transduction (GO:1902533)</td>
<td>0.000016</td>
<td>HSP90AB1; RRAGD; TRIM13; NRG1; VEGFA</td>
</tr>
<tr>
<td></td>
<td>Central nervous system neuron axonogenesis (GO:0021955)</td>
<td>3.03E-05</td>
<td>HSP90AB1; EPHB2</td>
</tr>
<tr>
<td></td>
<td>Regulation of establishment of endothelial barrier (GO:1903140)</td>
<td>5.27E-05</td>
<td>SIPR3; VEGFA</td>
</tr>
<tr>
<td></td>
<td>Positive regulation of protein-containing complex assembly (GO:0031334)</td>
<td>5.48E-05</td>
<td>NRG1; ATM; VEGFA</td>
</tr>
<tr>
<td></td>
<td>Regulation of cell adhesion (GO:0030155)</td>
<td>7.83E-05</td>
<td>ATM; EPHB2; VEGFA</td>
</tr>
<tr>
<td></td>
<td>Response to gamma radiation (GO:001032)</td>
<td>0.000107</td>
<td>TRIM13; ATM</td>
</tr>
<tr>
<td></td>
<td>Cardiac muscle cell differentiation (GO:0055007)</td>
<td>0.000107</td>
<td>NRG1; VEGFA</td>
</tr>
<tr>
<td></td>
<td>Mammary gland development (GO:0030879)</td>
<td>0.000107</td>
<td>NRG1; VEGFA</td>
</tr>
<tr>
<td>**Cellular</td>
<td>Secretory granule lumen</td>
<td>0.000993</td>
<td>HSP90AB1; CAMP; VEGFA</td>
</tr>
<tr>
<td>Component</td>
<td>Lysosome</td>
<td>0.003226</td>
<td>STX17; NPC1; RRAGD</td>
</tr>
<tr>
<td></td>
<td>Smooth endoplasm reticulum membrane</td>
<td>0.003246</td>
<td>STX17</td>
</tr>
<tr>
<td></td>
<td>Integral component of lysosomal membrane</td>
<td>0.006482</td>
<td>NPC1</td>
</tr>
<tr>
<td></td>
<td>Integral component of vacuolar membrane</td>
<td>0.007774</td>
<td>NPC1</td>
</tr>
<tr>
<td></td>
<td>Lytic vacuole</td>
<td>0.008598</td>
<td>NPC1; RRAGD</td>
</tr>
<tr>
<td></td>
<td>Axonal growth cone</td>
<td>0.009064</td>
<td>HSP90AB1</td>
</tr>
<tr>
<td></td>
<td>Smooth endoplasm reticulum</td>
<td>0.012283</td>
<td>STX17</td>
</tr>
<tr>
<td></td>
<td>Lytic vacuole membrane</td>
<td>0.012569</td>
<td>STX17; NPC1</td>
</tr>
<tr>
<td></td>
<td>Autophagosome membrane</td>
<td>0.01421</td>
<td>STX17</td>
</tr>
<tr>
<td>**Molecular</td>
<td>Growth factor activity</td>
<td>0.001414</td>
<td>NRG1; VEGFA</td>
</tr>
<tr>
<td>Function</td>
<td>ErbB-3 class receptor binding</td>
<td>0.003246</td>
<td>NRG1</td>
</tr>
<tr>
<td></td>
<td>TPR domain binding</td>
<td>0.003246</td>
<td>HSP90AB1</td>
</tr>
<tr>
<td></td>
<td>Transmembrane receptor protein tyrosine kinase activator activity</td>
<td>0.004542</td>
<td>NRG1</td>
</tr>
<tr>
<td></td>
<td>Vascular endothelial growth factor receptor 2 binding</td>
<td>0.005189</td>
<td>VEGFA</td>
</tr>
<tr>
<td></td>
<td>Cytokine activity</td>
<td>0.00545</td>
<td>NRG1; VEGFA</td>
</tr>
<tr>
<td></td>
<td>1-phosphatidylinositol-3-kinase activity</td>
<td>0.006482</td>
<td>ATM</td>
</tr>
<tr>
<td></td>
<td>Histone methyltransferase binding</td>
<td>0.007129</td>
<td>HSP90AB1</td>
</tr>
<tr>
<td></td>
<td>Nuclear export signal receptor activity</td>
<td>0.007129</td>
<td>XPO1</td>
</tr>
<tr>
<td></td>
<td>Vascular endothelial growth factor receptor binding</td>
<td>0.007774</td>
<td>VEGFA</td>
</tr>
<tr>
<td><strong>KEGG</strong></td>
<td>Autophagy</td>
<td>0.003458</td>
<td>STX17; RRAGD</td>
</tr>
<tr>
<td></td>
<td>Fluid shear stress and atherosclerosis</td>
<td>0.003557</td>
<td>HSP90AB1; VEGFA</td>
</tr>
<tr>
<td></td>
<td>NOD-like receptor signaling pathway</td>
<td>0.00595</td>
<td>HSP90AB1; CAMP</td>
</tr>
<tr>
<td></td>
<td>Human T-cell leukemia virus 1 infection</td>
<td>0.008598</td>
<td>XPO1; ATM</td>
</tr>
<tr>
<td></td>
<td>Chemical carcinogenesis</td>
<td>0.01017</td>
<td>HSP90AB1</td>
</tr>
<tr>
<td></td>
<td>Shigelllosis</td>
<td>0.010748</td>
<td>RRAGD; ATM</td>
</tr>
<tr>
<td></td>
<td>MicroRNAs in cancer</td>
<td>0.016687</td>
<td>ATM; VEGFA</td>
</tr>
<tr>
<td></td>
<td>Human papillomavirus infection</td>
<td>0.018883</td>
<td>ATM; VEGFA</td>
</tr>
<tr>
<td></td>
<td>SNARE interactions in vesicular transport</td>
<td>0.021245</td>
<td>STX17</td>
</tr>
<tr>
<td></td>
<td>PI3K-Akt signaling pathway</td>
<td>0.021422</td>
<td>HSP90AB1; VEGFA</td>
</tr>
</tbody>
</table>
Figure 3. Identification of common genes. Venn diagram showed the number of genes that intersected between MDD’s DEGs, COVID-19 genes, and autophagy-related genes. 13 common genes (STX17, VEGFA, NRG1, RRAQD, NPC1, XPO1, HERC1, HSP90AB1, CAMP, ATM, EPHB2, S1PR3, TRIM13) were obtained by taking the intersection of the three.

Figure 4. Gene enrichment analysis. A, Biological process. B, Cellular component. C, Molecular function. D, KEGG enrichment analysis. A higher p-value indicates that there are more genes involved in a specific pathway.
Interaction of TF Genes with Common Genes

The TF-gene regulatory network was constructed using the JASPAR database in NetworkAnalyst 3.0. Figure 5 illustrates the interaction between TF genes and common genes. The constructed network contains 13 core targets, 56 nodes, and 108 edges. We have found here that NRG1 is regulated by about 17 TF genes whereas VEGFA, RRAGD as well as ATM are regulated by 10 TF genes. A high degree of interaction between TFs and common genes is indicated by the fact that these TFs regulate more than one common DEG in the network.

Identification of COVID-19-MDD-Related Comorbidities

Diseases are interconnected by common genes. According to the gene-disease interaction analysis using NetworkAnalyst 3.0, NRG1, ATM, VEGFA, and HSP90AB1 are the key genes involved in COVID-19-MDD related comorbidities (Figure 6). Psychiatric disorders and tumors were the most commonly associated comorbidities, followed by gastric ulcer, liver cirrhosis, experimental, heart failure, and contact dermatitis.

Identification of Drug Candidates

To screen gene-targeting candidates from the DSigDB database, common genes were imported into the Enrichr platform at a \( p \)-value <0.01 (Table II). The top ten drug candidates which were screened were 1-phosphatidyl-myo-inositol, lapatinib, dipyridamole, resveratrol, fludarabine, sphingosine, temozolomide, PD 98059, tetracycline, and 1,1,1-trifluorohenicosa-6,9,12,15-tetraen-2-one.


LASSO regression analysis was conducted to further screen common genes which can act as diagnostic biomarkers for COVID-19-MDD. Finally, eight genes, including STX17, NRG1, RRAGD, XPO1, HERC1, HSP90AB1, EPHB2, and S1PR3 (Figure 7) were screened as diagnostic biomarkers based on the above analysis.

Figure 5. Construction of TF-gene network and gene comorbidity network. Network of transcription factors (TFs) interacted with common genes. The highlighted yellow nodes represent common genes, while the other nodes represent TFs. The network consists of 56 nodes and 108 edges.
Figure 6. Network of gene comorbidity in COVID-19-MDD. A purple point represents a common gene, while the other points represent gene-related comorbidities.

Table II. Gene targeting drug candidates in the DSigDB database.

<table>
<thead>
<tr>
<th>Term</th>
<th>p-value</th>
<th>Combined Score</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Phosphatidyl-myoo-inositol BOSS</td>
<td>5.50E-06</td>
<td>566.85862</td>
<td>NRG1, ATM, EPHB2, VEGFA</td>
</tr>
<tr>
<td>Lapatinib BOSS</td>
<td>5.52E-06</td>
<td>1392.455472</td>
<td>NRG1, EPHB2, VEGFA</td>
</tr>
<tr>
<td>dipyridamole BOSS</td>
<td>7.94E-06</td>
<td>1190.023463</td>
<td></td>
</tr>
<tr>
<td>resveratrol BOSS</td>
<td>1.03E-05</td>
<td>456.4826769</td>
<td>NRG1, EPHB2, VEGFA</td>
</tr>
<tr>
<td>Fludarabine CTD 00000135</td>
<td>1.15E-05</td>
<td>1014.797902</td>
<td>ATM, EPHB2, VEGFA</td>
</tr>
<tr>
<td>sphingosine BOSS</td>
<td>1.20E-05</td>
<td>996.0632854</td>
<td>NPC1, ATM, EPHB2, VEGFA</td>
</tr>
<tr>
<td>temozolomide BOSS</td>
<td>1.65E-05</td>
<td>865.5738314</td>
<td>ATM, EPHB2, VEGFA</td>
</tr>
<tr>
<td>PD 98059 CTD 00003206</td>
<td>1.88E-05</td>
<td>369.9156486</td>
<td>NPC1, NRG1, ATM, VEGFA</td>
</tr>
<tr>
<td>tetracycline BOSS</td>
<td>2.96E-05</td>
<td>669.1933597</td>
<td>NRG1, ATM, EPHB2, VEGFA</td>
</tr>
<tr>
<td>1,1,1-trifluorohenicosa-6,9,12,15-tetraen-2-one CTD 00002973</td>
<td>3.03E-05</td>
<td>3435.359022</td>
<td>NRG1, VEGFA</td>
</tr>
</tbody>
</table>
An approach combining bioinformatics and machine learning

Discussion

The prevalence of MDD is higher among patients with COVID-19, which leads to a higher prevalence of MDD\(^2,3\). COVID-19 may lead to MDD due to immune dysfunction and increased levels of inflammatory factors\(^30\). Autophagy plays a crucial role in viral infections and neurodegenerative diseases and is directly related to the onset and progression of both COVID-19 and MDD. There are, however, insufficient studies focusing on the mechanisms involved in COVID-19-MDD. To the best of our knowledge, this study is the first to examine COVID-19-MDD from an autophagy perspective using machine learning and bioinformatics.

By differentially analyzing GSE98793 genes associated with MDD, we first obtained DEGs that were intersected with CRGs and ARGs to obtain 13 common genes. By using LASSO regression analysis, eight genes were identified as diagnostic biomarkers for COVID-19-MDD, including STX17, NRG1, RRAGD, XPO1, HERC1, HSP90AB1, EPHB2, and S1PR3. Among them, STX17, XPO1, and HSP90AB1 have been confirmed to be associated with COVID-19, whereas NRG1, EphB2, and S1PR3 are associated with MDD. However, the link between HERC and RRAGD with COVID-19 and MDD has not been elucidated and requires further research.

STX17, which was recently identified as an autophagosomal SNARE protein, is a key target for the treatment of COVID-19-MDD\(^31,32\). The SNARE protein complex, an essential molecular component of neural connectivity, catalyzes synaptic vesicle fusion. COVID-19 viruses inhibit autophagic activity by blocking the interaction of the HOPS complex with the autophagosomal SNARE protein STX17, thus evading an immune attack\(^33\). The SNARE protein is a key target protein for severe psychiatric disorders such as depression and even suicidal thoughts\(^33,34\). Therefore, COVID-19 viruses may contribute to the onset of depression by altering the expression of the SNARE protein STX17. KEGG enrichment analysis in this study also demonstrated that SNARE interactions in vesicular are an important pathway for the onset of both COVID-19 and MDD\(^33,34\). GO analysis revealed that lysosomes and endoplasmic reticulum membranes may be important sites for the regulation of cell signaling in the nervous system through autophagy in COVID-19-MDD\(^35\). For this reason, the core gene STX17 is a crucial target for the treatment of COVID-19-MDD.

HERC1, a ubiquitin ligase, contributes to autophagy/lysosome and proteasome pathways that lead to protein degradation\(^36-38\). In the peripheral nervous system, HERC1 mutations alter presynaptic membrane dynamics, resulting in delayed neurotransmission and impaired movement and
learning\textsuperscript{39}. According to GO enrichment analysis, the most common genes were involved in autophagy-related endomembrane systems such as lysosomes and endoplasmic reticulum membranes and neurological signaling. Therefore, \textit{HERC1} may be an important mechanism in the pathogenesis of COVID-19-MDD. The relationship between \textit{HERC1}, COVID-19, and MDD is unclear at present. The lysosomal pathway, however, is common to both COVID-19 and MDD. As a result, \textit{HERC1} may play a role in the onset of COVID-19-MDD.

\textit{NRG1}, a member of the epidermal growth factor family, is involved in the regulation of neurodevelopment and synaptic plasticity\textsuperscript{40}, which are important mechanisms involved in the pathogenesis of numerous psychiatric disorders such as depression, schizophrenia, bipolar affective disorder, and even suicidal thoughts and behaviors\textsuperscript{41,42}. Enrichment analysis revealed that molecular functions of common genes are closely associated with vascular endothelial growth factors. By affecting vascular endothelial growth factors and enhancing vascular permeability-induced plasma exudation and pulmonary edema, COVID-19 infection can further exacerbate tissue hypoxia. The anti-vascular endothelial growth factor drug, bevacizumab, inhibits this process, improves patient oxygen levels, and shortens the duration and demand of oxygen support\textsuperscript{43}. It is possible that COVID-19 infection affects the central nervous system via trans-synaptic transfer\textsuperscript{44}, causing depression. \textit{NRG1} gene may, therefore, play a significant role in the onset and progression of COVID-19-MDD. There is a high prevalence of comorbidities associated with COVID-19-MDD, and \textit{NRG1} may play an important role in comorbid tumors in COVID-19-MDD patients\textsuperscript{45}. The TF-gene interaction network suggested that \textit{NRG1} was the gene regulated by most TFs. Among them, Fox1 is not only a significant TF in COVID-19\textsuperscript{46} but also plays a key role in the pathogenesis of tumors\textsuperscript{37}.

\textit{SIPR3}, a G protein-coupled receptor, is considered a functional receptor for sphingosine 1-phosphate and may be involved in the regulation of angiogenesis and the function of vascular endothelial cells\textsuperscript{48}. \textit{SIPR3}, a modulator of stress resilience, is closely associated with psychiatric disorders. In the blood of veterans with PTSD, the mRNA expression of \textit{SIPR3} is reduced and negatively correlated with symptom severity\textsuperscript{49}. Additionally, pulmonary fibrosis is a common comorbidity associated with COVID-19\textsuperscript{40}. \textit{SIPR3} is a key gene that causes pulmonary fibrosis\textsuperscript{51}. Therefore, \textit{SIPR3} may play a crucial role in secondary depression and pulmonary fibrosis among patients after COVID-19 infection.

Exportin 1 (\textit{XPO1}) regulates the export of a series of “cargoes” (including proteins and several RNAs) from the nucleus to the cytoplasm and plays a crucial role in maintaining intracellular homeostasis. SARS and Middle East Respiratory Syndrome (MERS) virus replication are guided by several co-proteins, including \textit{XPO1}\textsuperscript{52,53}. In the morning, the clinical symptoms of depression are alleviated, and in the evening, they are aggravated. \textit{XPO1} is an important protein in the body that regulates circadian rhythms\textsuperscript{54}. COVID-19 virus replication affects the circadian rhythm of humans and induces depression by causing an increase in \textit{XPO1} protein expression.

As a member of the heat shock protein 90 families, \textit{HSP90AB1} is involved in signal transduction, protein folding and degradation, and morphological evolution. SARS-CoV-2 triggers cytokine storms by affecting genes such as \textit{HSP90AB1}, leading to lesions in the lung and extrapulmonary tissues\textsuperscript{55}. Cytokine storm-induced neuroinflammation may also be responsible for depression caused by COVID-19 infection\textsuperscript{56,57}.

Eph receptors are the largest family of receptor tyrosine kinases (RTKs). One of the important members of this family, Eph receptor B2 (\textit{EphB2}), may be involved in the onset of depression\textsuperscript{58,59} and plays a crucial role in hippocampal synaptic plasticity, especially in long-range enhancement and long-range inhibition\textsuperscript{60}. As a result of COVID-19-related respiratory distress, it can trigger corticosteroid release, hypothalamic stimulation, and glucocorticoid production, and thus interferes with brain metabolism and attributes to COVID-19-MDD\textsuperscript{44}. Enrichment analyses also revealed that tyrosine kinases are important targets for common gene functions. A key component of B-cell receptor (BCR) signaling is Bruton’s tyrosine kinase (BTK). As a selective BTK inhibitor, acalabrutinib inhibits the hyperinflammatory immune response and improves oxygenation levels in patients with COVID-19\textsuperscript{61}. Therefore, this gene plays a crucial role in the onset and progression of COVID-19-MDD.

\textit{RRAGD}, a monomeric guanine nucleotide-binding protein (G protein), serves as a molecular switch in many cellular processes and signaling pathways\textsuperscript{42}. A clear association exists between \textit{RRAGD} and COVID-19-MDD and other diseases. There is evidence that \textit{RRAGD} is associated with Crohn’s disease\textsuperscript{62}. It can, therefore, be studied further as a new diagnostic biomarker.
Comorbidity analysis of COVID-19-MDD was performed based on common genes. Based on our predictions, the possible comorbidities of COVID-19-MDD include various psychiatric disorders such as bipolar disorder and schizophrenia, various tumors such as breast and prostate tumors, and other diseases such as gastric ulcer, cirrhosis, heart failure, and contact dermatitis. Patients with psychiatric disorders such as schizophrenia are at high risk of SARS-CoV-2 infection\textsuperscript{64}. In patients with COVID-19-MDD, tumors are common comorbidity. Among cancer patients with SARS-CoV-2 infection, the incidence rates were 24.7\%, 20.5\%, 13.0\%, 7.6\%, 7.3\%, 6.1\% and 6.0\% for lung, colorectal, breast, esophageal, bladder, pancreatic, and cervical cancers, respectively\textsuperscript{65}. Dermatosis is also common comorbidity. A study by Jimenez et al\textsuperscript{66} described a skin rash as the clinical presentation found among 21 patients with COVID-19\textsuperscript{66}. The prevalence of COVID-19-MDD-related liver disease is also high. According to a report, 2–11\% of COVID-19 individuals have primary chronic liver disease\textsuperscript{67}. Furthermore, COVID-19-MDD-related comorbidities include heart failure and gastric ulcers\textsuperscript{68–70}. Drugs for COVID-19-MDD are lacking despite many commercially available depression medications. In the development of new drugs, there is a high cost and a long development period. To minimize the cost of drug development, we screened drugs based on the key genes using bioinformatics analysis. A total of ten drug candidates for COVID-19-MDD were screened from the DSigDB database. It is reported that resveratrol inhibits protease 3CL\textsubscript{\textalpha}, an important therapeutic target of SARS-CoV-2\textsuperscript{71,72} and that it has antidepressant properties similar to fluoxetine\textsuperscript{73}. SARS-CoV-2 replication is inhibited by lapatinib\textsuperscript{4}, dipyridamole\textsuperscript{5}, and tetracycline\textsuperscript{6}. Fludarabine inhibits type I interferon-induced expression of the SARS-CoV-2 receptor angiotensin-converting enzyme 2\textsubscript{\textbeta}. Sphingosine prevents the interaction of the viral spike protein of SARS-CoV-2 with the host cell receptor\textsuperscript{78}. Temozolomide has been used in the treatment of COVID-19. However, lymphopenia may predict a high mortality rate\textsuperscript{79}. The relationship between PD 98059, 1-phosphatidyl-mylo-inositol, 1,1,1-tri-fluorohenicosa-6,9,12, and 15-tetraen-2-one and COVID-19-MDD has not been revealed. One of the above drug candidates, resveratrol, can be used to treat both COVID-19 and depression and may be the preferred candidate for future therapeutic trials. There are some limitations to this study. First, it was based on secondary mining and analysis of a previously published dataset. Second, even though GSE98793 contains the most depression samples, its sample size is still relatively small. As a result, it is necessary to have a dataset with a larger sample size for further mining. Third, further validation of the diagnostic biomarkers screened in this study is needed in the future because of the lack of patients with COVID-19-MDD and their relevant gene sets.

Conclusions

A bioinformatics approach was used to screen 13 common genes associated with COVID-19-MDD. A subsequent GO and KEGG enrichment analysis identified specific mechanisms underlying COVID-19-MDD pathogenesis. Additionally, TF-gene regulatory networks were constructed. Comorbidities associated with COVID-19-MDD were also identified. Finally, the DSigDB database was searched to obtain candidate therapeutic agents. In summary, the eight diagnostic biomarkers of COVID-19-MDD screened by the machine learning algorithm provide further insight into its clinical diagnosis and treatment.

Conflicts of Interest

The authors report no conflict of interests associated with this manuscript.

Authors’ Contributions

The study’s conception and design were contributed by SW, LZ, and MP. The first draft of the manuscript was written by CY, FJZ, and LLZ. Material preparation, data collection, and analysis were performed by DXX, YL, and JJL. The final versions of the manuscript were revised by SXT, YL, and XJL. The final manuscript was read and approved by all authors.

Data Availability

The dataset GSE98793 for this study can be found in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo). All data generated or analyzed during this study are included in this published article.

Funding statement

This study was supported by the National Natural Science Foundation of China (82074388); the project of Shaoxing medical key discipline construction plan (2019SZD06); the Project of Health and Family Planning Commission of Zhejiang province (2021KY1139); Shaoxing Medical and Health Science and Technology Plan Project (2020A13026); Shandong Traditional Chinese Medicine Science and Technology Project (2020M044); and the Natural Science Foundation of Shandong Province (ZR2020QH306).
Acknowledgments
We would like to thank Dr. Guangli Sun, Chief Physician, for her critical comments on the manuscript. We thank Bulletin Edits Limited for the linguistic editing and proofreading of the manuscript.

ORCID IDs
Cheng Yu, https://orcid.org/0000-0002-3616-095X
Fengjun Zhang, https://orcid.org/0000-0003-3350-5710
Lili Zhang, https://orcid.org/0000-0001-7697-4854
Dexian Xian, https://orcid.org/0000-0002-6900-3993
Ying Li, https://orcid.org/0000-0003-0863-9467
Jiajing Li, https://orcid.org/0000-0002-6900-3993
Shuai Wang, https://orcid.org/0000-0002-1067-8347

References


An approach combining bioinformatics and machine learning

Poklar Ulrih N. Inhibition of the SARS-CoV-2 3CL (pro) main protease by plant polyphenols. Food Chem 2022; 373: 131594.

72) van Brummelen R, van Brummelen AC. The potential role of resveratrol as supportive antiviral in treating conditions such as COVID-19 - A formulator’s perspective. Biomed Pharmacother 2022; 148: 112767.


79) Tanriverdi O. Lymphopenia that may develop in patients treated with temozolomide and immune control check-point inhibitor may be a high risk for mortality during the COVID-19 outbreak. Med Oncol 2020; 37: 51.