

Network of “drug-target-SARS-CoV-2 Related Genes” Through Integrated Analysis of Pharmacology and Geo Database

Jin ping Hou

Chongqing Medical University Affiliated Children's Hospital

Yong heng Wang

Chongqing Medical University

Yu meng Chen

Chongqing Medical University

Yi hao Chen

Chongqing Medical University

Xiao Zhu

Chongqing Medical University

Rui si Qin

Chongqing Medical University

Tingting Chen (✉ cherrychen@cqmu.edu.cn)

Chongqing Medical University <https://orcid.org/0000-0002-0089-9402>

Research

Keywords: COVID-19, SARS-CoV-2, traditional Chinese medicine, big data mining, antivirus

Posted Date: December 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-117894/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Coronavirus Disease 2019 (COVID-19) respiratory disease rapidly caused a global pandemic and social and economic disruption. The combination of Traditional Chinese medicine (TCM) and Conventional Western medicine (CWM) is more effective for COVID-19 treatment. Moreover, TCM and CWM are important data source for developing new drug targets and promote strategies treat SARS-CoV-2 infections. However, many studies have analyzed the therapeutic mechanism of CWM or TCM alone for COVID-19, it is still unclear the interaction mechanism between TCM and CWM on COVID-19.

Methods

This paper integrates network pharmacology and GEO database to mine and identify COVID-19 molecular therapeutic targets, providing potential targets and new ideas for COVID-19 gene therapy and new drug development. It includes: 1) using TCMSP, TTD, PubChem and CTD databases to analyze drug interactions and associated phenotypes for SARS-CoV-2, to correlate drug and disease interaction mechanisms to screen key drug targets; 2) using GEO database to correlate differential genes and drug targets to screen potential antiviral gene therapy targets, to construct regulatory network and key points of SARS-CoV-2 therapeutic drugs; 3) using computer simulation of molecular docking to screen virus-related proteins for new drugs.

Results

Integrated analysis of network pharmacology discovered that baicalein, estrone and quercetin are the pivotal active ingredients in TCM and CWM. Combining drug target genes in pharmacology database and virus induced genes in GEO database, the result showed the core hub genes related to COVID-19: *STAT1*, *IL1B*, *IL6*, *IL8*, *PTGS2* and *NFKBIA*, and these genes were significantly downregulated in A549 and NHBE cells by SARS-CoV-2 infection. Moreover, chemical interaction and molecular docking analysis of hub genes showed that folic acid might as be potential therapeutic drug for COVID-19 treatment, and SARS-CoV-2 nucleocapsid phosphoprotein was a potential drug target. The network of “drug-target-SARS-CoV-2 related genes” provide noval potential compounds and targets for further studies of SARS-CoV-2.

Conclusions

Integrated analysis of network pharmacology and big data mining provided noval potential compounds and targets for further studies of SARS-CoV-2. Our research implied folic acid and SARS-CoV-2 N as therapeutic target in TCM and CWM. Our research also suggests that targeting SARS-CoV-2 N protein is likely to be a common mechanism of TCM and CWM. On the one hand, the identification of pivotal genes provides a target for COVID-19 molecular therapy, on the other hand, it provides ideas for the analysis of interaction mechanism between virus and host.

Introduction

The recently discovered beta-coronavirus severe acute respiratory syndrome (SARS-CoV-2), the causative agent of Coronavirus Disease 2019 (COVID-19) respiratory disease rapidly caused a global pandemic and social and economic disruption^{1,2}. As of October 23th, 2020, a total of over 41 million cases of COVID-19 were reported and more than 1100 thousand lives were claimed globally (<https://covid19.who.int>). SARS-CoV-2 is sense single-stranded genomic RNA virus, approximately 30 kb in length, encodes four structural proteins: the spike (S), membrane (M), envelope (E) and nucleocapsid (N) proteins³. S protein binds to angiotensin-converting enzyme (ACE2) receptor on the cytoplasmic membrane of type 2 lung cells and intestinal epithelium. After binding, S protein was cleaved by the host membrane serine protease TMPRSS2, which promoted the virus entry⁴. These structural proteins are components of the mature virus and play a crucial role in viral structure integrity, or as in the case of the spike protein, for viral entry into the host^{5,6}. The SARS-CoV-2 genome contains 14 open reading frames (ORFs), preceded by transcriptional regulatory sequences (TRSs)⁷. The structural and accessory proteins are translated from a set of nested sub-genomic (g) RNAs. Underlining how SARS-CoV-2 hijacks the host during infection absolutely could promote the development, characterization, and deployment of an effective vaccine or antibody prophylaxis or treatment against SARS-CoV-2 and could prevent morbidity and mortality and curtail its epidemic spread⁸.

Traditional Chinese medicine (TCM) formulations had been originally developed for earlier viral diseases and been used in COVID-19 treatment⁹. TCM was extensively widely used to control the epidemic in China¹⁰. Shufeng Jiedu Capsule and Lianhua Qingwen capsule have good clinical effect on patients with COVID-19^{11,12}. The further study of TCM may lead to the identification of novel antiviral compounds for the treatment of SARS-CoV-2 or other emerging fatal viral diseases⁹. The antiviral medicine mainly target on host protease or other cellular proteins, such as teicoplanin, arbidol, chloroquine and derivatives, and niclosamide; and target on viral proteins, such as niclosamide, lopinavir/ritonavir, remdesivir, favipiravir, and ribavirin¹³. Combination of TCM and conventional Western medicine (CWM) can effectively inhibit SARS-CoV-2 replication and prevent the production of proinflammatory cytokines induced by SARS-CoV-2 infection. Many studies have only analyzed the therapeutic mechanism of CWM or TCM alone for COVID-19, so this study integrated the mechanism of CWM and TCM together to explore their generality and characteristics. Based on the concept of "disease-target-compound-drug", network pharmacology systematically analyzes the mechanism of drug action and its interaction network from multiple levels and angles^{14,15}. Network pharmacology combines the information of genome, proteome, drugs, diseases and other related databases with experimental verification data through the method of big data mining, so as to establish an inter related network of "disease-target-compound-drug"^{16,17}. Therefore, this method can be used to systematically analyze the intervention and influence of TCM and CWM on COVID-19, reveal the mechanism of action of drugs in human body, correlate the pharmacophores and interactions of various drugs, and screen the core targets. In addition, the accumulation of big data resources is conducive to the screening of molecular markers for disease prevention¹⁸. Biomolecular

targets are the basis of accurate diagnosis and personalized drug design, which can be used to capture specific disease signals in the early stage to formulate appropriate treatment plans¹⁸. Screening of molecular targets for COVID-19 therapy is of great significance for understanding the interaction mechanism between SARS-CoV-2 and host. Therefore, it is of great significance to integrate network pharmacology and GEO database to screen COVID-19 molecular therapeutic targets with clinical value. Identifying the regulatory effects of core hub genes based on big data mining can provide potential targets and new ideas for COVID-19 gene therapy and new drug development.

Materials And Methods

2.1 Active compounds and drug target genes

Chemical ingredients and chemical target genes were collected from the Traditional Chinese Medicines for Systems Pharmacology Database and Analysis Platform (TCMSP) Database (<http://tcmospw.com/tcmosp.php>). In TCMSP database (<http://tcmospw.com/tcmosp.php>), the final candidate active ingredients were screened, according to the oral bioavailability (OB) $\geq 30\%$, drug-likeness (DL) ≥ 0.18 , drug half-life (HL) ≥ 10 . The compound target genes were imported into UniProt (<https://www.uniprot.org/>) that defining the species as human, and obtained the gene names corresponding to the target genes.

The COVID-19 treatment plan (issued in the sixth version) issued by the office of the National Health and Health Committee of China is selected as follows: *Ephedra Herba* (Mahuang), *Cinnammi Cortex* (Rougui), *Alisma Orientale*(Sam.) Juz.(Zexie), *Polyporus Umbellatus*(Pers)Fr.(Zhuqin), *Atractyloes Macrocephala* Koidz.(Baizhu), *Arum Ternatum* Thunb(Banxia), *Asteris Radix Et Rhizoma*(Zikou), *Farfarae Flos*(Donghua), *Belamcandae Rhizome*(Shegan), *Asari Radix Et Rhizoma*(Xixin), *Rhizoma Dioscoreae*(Shanyao), *Aurantii Fructus Immaturus*(Zhishi), *Fortunes Bossfern Rhizome*(Guangzong), *Radix Cynanchi Paniculati*(Xuchangqing), *Eupatorium Fortunei* Turcz(Peilan), *Scutellariae Radix*(Huangqin), *Radix Bupleuri*(Chaihu), *Artemisia Annu L.*(Qinghao), *Isatidis Folium*(Daqingye), *Polygoni Cuspidati Rhizoma Et Radix*(Huzhang), *Verbenae Herb*(Mabiancao), *Phragmitis Rhizoma*(Lugen), *Citri Grandis Exocarpium*(Huajuhong), *Notopterygii Rhizoma Et Radix*(Qianghuo), *Arecae Semen*(Binlang), *Amygdalus Communis Vas*(Xingren), *Licorice*(Gancao), *Pogostemon Cablin*(Blanco) Benth.(Huoxiang), *Magnolia Officinalis Rehd Et Wils*(Houpu), *Atractylodes Lancea* (Thunb.)Dc.(Cangshu), *Amomum Tsao-Ko Crevostet*(Caoguo), *Radix Rhei Et Rhizome*(Shengdahuang), *Anemarrhenae Rhizoma*(Zhimu), *Radix Paeoniae Rubra*(Chishao), *Figwort Root*(Xuansen), *Forsythiae Fructus*(Lianqiao), *Cortex Moutan*(Danpi), *Coptidis Rhizoma*(Huanglian), *Lepidii Semen Descurainiae Semen*(Tinglizi), *Panax Ginseng C. A. Mey.* (Rensen), *Cornus Officinalis Sieb. Et Zucc.*(Shanzhuyu), *Arum Ternatum Thunb.*(Fabanxia), *Citrus Reticulata*(Chenpi), *Codonopsis Radix*(Dangsen), *Glehniae Radix*(Beishasen), *Schisandre Chinensis Fructus*(Wuweizi), *Mori Follum*(Sangye), *Prragmitis Rhizoma*(Lugen), *Radix Salviae*(Dansen).

2.2 Differential expressed genes (DEGs) induced by SARS-CoV-2 infection NHEB and A549 cell lines

The differential expression genes (DEGs) between EGFR-TKI sensitive and resistance were calculated from public dataset of Gene Expression Omnibus (GEO). GSE147507 dataset include SARS-CoV-2 infected NHEB and A549 cell lines, and the uninfected group. SARS-004 group indicated that SARS-CoV-2 infected NHBE cells at MOI 2, Cov-002 group indicated that SARS-CoV-2 infected A549 cells at MOI 0.2, and the control group were the corresponding uninfected cell lines. For assay the deregulation gene expression, “limma” package of R software was employed to test the DEGs. mRNAs with log2 fold change ($|\log_2FC|$, $P < 0.01$) considered to be differentially expressed mRNAs.

2.3 Identify core sub-network from compound-target network

For identifying core targets from compounds-targets network, the overlap of DEGs and targets were extracted. There eight overlap targets were identified. We selected eight targets and their neighbors (only compounds) as core sub-network.

2.4 Construction of Gene Enrichment Analysis

To better understand the processes associated with COVID-19 and Chinese medicine treatment, we performed GO terms and KEGG pathway enrichment analyses. The GO enrichment analysis provides three structured networks of defined terms to describe gene attributes. Enriched GO terms are classified according to biological process (BP), molecular function (MF), and cellular component (CC). The KEGG (<http://www.genome.jp/kegg/>) is a database for large-scale systematic analysis of molecular interaction networks of genes or proteins. The DAVID bioinformatics resources consist of an integrated biological knowledge base analytic tools, aimed at systematically extracting biological meaning from large gene or protein lists. We used the web-based search engine, DAVID, to determine over-represented GO terms and KEGG pathways with thresholds of an enrichment score > 2 , count > 5 , and $P < 0.05$ and analyzed COVID-19-related pathways and GO terms. Venn diagram and bubble graphs of were performed using the OmicShare tools, a free online analysis tool (<http://www.omicshare.com/tools>). Significant pathway terms of KEGG were mapped into a bubble graph. The big and higher bubbles represent those highly significantly enriched pathway terms.

2.5 Protein-Protein Interaction Network Analysis

Protein-protein interaction (PPI) networks included information on the biological processes and molecular functions of cells. We used the online search tool for recurring instances of neighboring genes (STRING, Version 9.1) (<http://www.string-db.org>) to predict the interactions. The Cytoscape software 3.7.2 (<http://cytoscape.org/>) was used to visualize networks. To identify crucial relationships in the PPI network, potentially overlapping modules that were densely connected were subsequently identified using Molecular Complex Detection (MCODE) plugin in the Cytoscape program. The MCODE plugin was used to re-analyze the clusters among the network according to the k -core = 2. A value of $P < 0.01$ was considered the significant threshold.

2.6 Network Construction of KEGG Pathway

To better analysis the holistic mechanism of Chinese herbs in endometriosis treatment, the subnetworks pathway was compiled by following the procedures: All targets of Chinese herbs in endometriosis treatment were submitted to an online tool KEGG Search Pathway (https://www.genome.jp/kegg/tool/map_pathway1.html). Based on the mechanism of endometriosis, multiple pathways were integrated and overlapped according to cross-talk targets in these maps. Based on the cross-talk of the pathways, we further constructed a targets-pathways network of Chinese medicine treatment.

2.7 Network Construction

We established a network analysis of the relevant compound targets. The components of the target networks for the herbs were constructed using the Cytoscape software 3.7.2. The nodes in each network were evaluated based on three indices: degree, node betweenness, and node closeness. Degree indicated the number of edges between a single node and other nodes in a network. Node closeness represented the inverse of the sum of the distance from one node to other nodes. The importance of a node in a network was indicated by the values of these indices, with higher values indicating greater importance.

2.8 Component-molecular target docking

From RSCB PDB database (<https://www.rcsb.org/>) Download the 3D structure pdb format files of SARS-CoV-2 related protein, angiotensin converting enzyme II (*ACE2*) and angiotensin converting enzyme (*ACE*) remove ligands and non-protein molecules (such as water molecules) in target proteins by using discovery studio 2020 client software, and then save them as PDB suffix files. From PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) Download the SDF format file of compound 2D structure. Using pyrx software, we first upload the protein file after water removal and hydrogenation, convert it into pdbqt format file, then upload the compound file to minimize its energy, and convert it into pdbqt format file. Finally, Vina is used for docking. The binding energy is less than 0, which indicates that the ligand and the receptor can spontaneously bind. The active components with binding energy ≤ -5.0 kJ/mol were selected as the screening basis of antiviral particles for COVID-19 target.

Results

3.1 Identification of effective compounds and target genes of COVID-19 clinical TCM

Effective compounds and target genes were identified through integrated bioinformatics analysis based on TCMSP, TTD, PubChem, CTD and GEO datasets (Fig.S1). According to the COVID-19 pneumonia treatment plan (issued in the sixth version) issued by the office of the national health and Health Committee of China, TCMSP database was used to search the active ingredients in 49 kinds of TCM in COVID-19 prescription, and the parameters $OB > 30\%$, $DL > 0.18$, $HL > 8$ were used as the conditions to screen the active ingredients (gypsum, ginger, etc.) were the relevant results. A total of 374 active components and 8355 target genes were obtained by analyzing and summarizing the active components of drugs (Table S1). *Amygdalus Communis Vas*, *Radix Bupleuri*, *Scutellariae Radix*, *Citrus Reticulata*,

Pogostemon Cablin (Blanco) Benth, Ephedra Herba, Notopterygii Rhizoma Et Radix, Atractylodes Lancea (Thunb.)Dc, Magnolia Officinalis Rehd Et Wils, Arecae Semen, Amomum tsaoko Crevost et Lemarie, Anemarrhenae Rhizoma, Forsythiae Fructus, Licorice and Arum Ternatum Thunb were repeated drugs in 10 different prescriptions. Furthermore, the correlation analysis of these effective compounds showed that there was a hub effective component MOL000422 (kaempferol, C₁₅H₁₀O₆) presented in the 10 prescriptions (Fig.1A). PubChem analysis showed the hub compound structure (Fig.1B and C). There are 77 kinds of effective components associated with prescriptions 1, 3, 4, 6, 7, 9 and 10 (Table S2). Each prescription has its own unique active ingredients for patients with different diseases. Prescription 1 has 23 unique active ingredients, prescription 2 has 3, prescription 3 has 3, prescription 4 has 3, prescription 6 has 2, prescription 7 has 2, prescription 8 has 8, prescription 9 has 3, prescription 10 has 43 (Fig.1A). In addition, TCM target genes were collected for GO and KEGG pathway analysis, and the result implied that target genes significantly involved positive regulation of transcription from RNA polymerase II promoter, nucleus and protein binding category (Fig.1D), and mainly involved in pathway in cancer, TNF signaling and Osteoclast differentiation (Fig.1E).

3.2 Screening of joint effective compounds and drug-target genes by analysis of combination of Chinese and Western Medicine

We firstly constructed Venn graphs of TCM and CWM compounds and target genes based on TCMSP and TTD/PubChem database, and found 3 hub compounds and 64 target genes (Fig.2A and B). We further analyzed 3D structure of 3 hub compounds (Baicalein, Estrone and Quercetin) based on NCBI PubChem, and the result showed the three structures are similar, Baicalein and Quercetin have large conjugation systems, and Estrone only has conjugation of benzene rings (Fig.S2B). Hub target genes significantly involved in cytosol and protein binding category, and pathway in cancer and Hepatitis B, according to GO and KEGG pathway analysis (Fig. 2C and Fig.S2A). Moreover, Cytoscape was used to construct the regulatory network of “Traditional Chinese medicine-Hub compound-Target gens” (Fig.2D), which indicated that Quercetin had more complex network systems and Estrone had less target genes. *MAPK14, MAPK3, MCL1* connected Quercetin and Baicalein, *OPRM1, ADRB2, SCN5A, EGF* and *ACE* connected Estrone and Quercetin.

3.3 Identification of differentially expressed genes (DEGs) induced by SARS-CoV-2 infection based on GEO database: GSE147507

This project is based on GEO database to download the original data of GSE147507. The data set explored the response of NHBE cell lines and A549 cell lines to SARS-CoV-2 infection at the transcriptional level, with a total of 110 samples. The raw data was analyzed by R language and the conditions $P < 0.05$ and $|\log_2\text{-FC}| > 1$ defined as differential genes. The results showed that there were 1177 differentially expressed genes in A549 cells and 1407 genes in NHBE cells. The volcano plots of DEGs among Cov-set and SARS-set were shown in Fig.3A and B. Venn analysis showed that 8 up-regulated genes and 113 down regulated genes were the same in the two kinds of cells (Fig.3C and Table 1). NCBI database was used to analyze the tissue expression profile characteristics of hub differential

genes (Fig.S3). The results showed that the up-regulated differentially expressed genes induced by SARS-CoV-2 had high tissue specificity. *SYNE1* was highly expressed in ovary, *CDKN2C* in adipose tissue, *gnpda1* in kidney, *RPAGD* in goiter, *ANKH* in pancreas and *PRODH* in small intestine. The down-regulated genes induced by virus were highly expressed in bone marrow, lymph node, spleen, lung, cecum, gallbladder and bladder. In addition, the expression levels of pivotal genes in pancreas and liver tissues were significantly lower than those in other tissues. Among all DEGs, only *C3*, *C1R*, *C1S*, *ASS1*, *ERRF11*, *KYNU*, *HSD11B1*, *SERPINA3*, *SAA1* and *CFB* were significantly expressed in liver tissues, while *ERRF11*, *SERPINA3* and *BCL2A1* were significantly expressed in pancreas.

The DEGs KEGG pathway enrichment showed DEGs significantly involved in pathway of Herpes simplex infection, influenza A, Measles and TNF signaling (Fig3D). GO analysis results indicated DEGs were enrichment in type I interferon signaling pathway, defense response to virus, cytoplasm, cytosol and protein binding (Fig.3E). The above results indicated that SARS-CoV-2 infection wildly down-regulated the immune response of host.

3.4 Identification of key hub DEGs by integrating internet-pharmacology and GEO database

In order to further screen molecular targets for COVID-19 regulation, and provide targets for exploring the mechanism of interaction between virus and host and screening new drugs, we further conduct molecular docking between drug targeted genes (DTGs) and differential expressed genes (DEGs) induced by SARS-CoV-2 infection, and screen core pivotal genes. Venn analysis was carried out on the target genes of ten TCM prescriptions and the DEGs induced by virus infection. The results showed that the number of key genes in the target genes of SARS-004 group was more than that of Cov-002 group (Fig.4A). These joint genes were down regulated by viral infection in both Cov and SARS groups (Fig.4B and C). Furthermore, Venn analysis was carried out on the related genes of each prescription and GEO set, and found the common associated genes (Fig.S4A and B). *ICAM1*, *IL1A*, *PTGS2*, *SLPI*, *STAT1*, *TOP2A* were the common key genes in CoV and SARS groups. Except *TOP2A*, they were all down regulated by virus infection (Fig. S4C).

Besides, based on TTD (Therapeutic Target Database), GEO and TCMSP databases, the drug targets of Chinese and Western medicine and the differential expressed genes induced by virus were analyzed, and 14 core pivotal genes were obtained. *IL6*, *IL8*, *IL1B*, *NFKBIA*, *PTGS2*, *STAT1* and *THBD* were all present in both Cov and SARS groups. Except *THBD*, other genes were down regulated by viral infection in both groups (Fig.4D and E). Finally, the interaction compounds of core hub genes were analyzed by CTD database, and two common compounds, folic acid and ozone, were screened (Fig.S4E). Folic acid can down regulate the expression of core hub genes, and zone can up regulate the expression of core hub genes.

3.5 Identification of Component-molecular target docking

The hub genes and SARS-CoV-2 related genes were respectively done component-molecular target docking with compounds. The lower the Binding energy is, the firmer the binding between compound and

molecular target. This means the compound have greater ability to make a difference. The binding energy was lower than 5KJ/mol as the screening condition, and the docking results were listed (Table 2 and Table 3), indicating that these compounds had good binding ability to target proteins. Among the four compounds, the binding energy between folic acid and target protein were -7.2/kcal.mol, -6.8/kcal.mol, -6.5/kcal.mol, -8.5/kcal.mol, -10.0/kcal.mol and -6.8/kcal.mol, generally lower than that of other compounds. *NFKBIA* and *PTGS2* also have low binding energy with compounds (Table 2 and Fig. 5A). All the four compounds had lower binding energy with *ACE2*, among which folic acid was the most significant. The binding energy between the compounds and *ACE2* is slightly higher than that of the compounds with *ACE*. Among the eight proteins of SARS-COV-2, Nucleocapsid Phosphoprotein (N) had the lowest binding energy with compounds. The results with the minimum binding energy for each protein were selected for visualization (Table 3 and Fig. 5B).

Discussion

SARA-Cov-2 infection molecular therapeutic network model is built based on intelligent computer and big data integration bioinformatics. The project mainly relies on database analysis to analyze the interactions and related phenotypes among SARA-Cov-2 drugs, and analyze the key drug targets by correlation analysis. Infected cells and normal cells were investigated by GEO database. This research mainly relies on the published data of the database to integrate bioinformatics analysis to obtain the key drug targets and potential targets of gene therapy, explore the mechanisms of effective compounds in the treatment of COVID-19.

The category of "epidemic disease" in traditional Chinese medicine has accumulated rich experience in fighting epidemics for thousands of years. After analyzing the active compounds of ten TCM prescriptions, a common core compound, kaempferol, was found. It has been shown that kaempferol not only has some anti-inflammatory, anti-cough and expectorant effects, but also lowers blood glucose and prevents vascular complications of diabetes¹⁹. Coronavirus infection can induce COVID-19 patients with normal glucose tolerance to increase blood glucose, sustained hyperglycemia is not only conducive to virus replication in the body, but also reduce the body's ability to resist infection, hyperglycemia patients develop serious illness and face a higher risk of death^{20,21}. This indicates that TCM can play a full therapeutic role in COVID-19. In addition, GO function was performed on the target sites of TCM, and the results showed that the target genes were enriched in the positive regulation of transcription from RNA polymerase II promoter, oxidation-reduction process, protein binding. Conducting KEGG pathway analysis revealed that target genes were mainly involved in pathway in cancer, TNF signaling pathway, and osteoclast differentiation. It was shown that these terms play a key role in the treatment of COVID-19 by TCM.

In the association analysis of TCM and CWM, three central compounds were identified, namely, baicalein, estrone and quercetin. Baicalein is a flavonoid with antihyperglycemic activity²², which, like kaempferol, can be useful in patients with diabetes mellitus. Quercetin, like baicalein, is a flavonoid that has the beneficial effects on inflammation and immunity²³. Study had shown that quercetin may decreases the

frequency and duration of respiratory tract infections as an effective intervention; however, more research is needed²⁴. Natural estrogens are mainly estradiol, estrone and estriol. Numerous studies had shown that estrogens are associated with the development of liver cancer and work by binding to estrogen receptors²⁵. What's more, it had shown that estrogen treatment silences the inflammatory reactions and decreases virus titers leading to improved survival rate in animal experiments²⁶. It is evident that TCM and CWM are effective in targeting patients with comorbid diabetes, liver damage, and kidney damage. GO analysis was performed on the overlapping fraction of targets obtained from target association analysis of TCM and CWM to obtain target enrichment in cytosol and protein binding category. KEGG analysis was performed to obtain target enrichment in the pathway in cancer and Hepatitis B. The commonality between TCM and CWM in the treatment of COVID-19 was shown.

In the analysis of differential genes induced by viral infection, up-regulated genes such as *gnpda1* were found to be highly expressed in the kidney, *RPAGD* in the thyroid, *ANKH* in the pancreas, *PRODH* in the small intestine, and down-regulated genes were highly expressed in the bone marrow, lymph nodes, spleen, lung, cecum, gallbladder, and bladder. In addition, key genes were expressed at significantly lower levels in pancreatic and liver tissues. It is hypothesized that these genes are key targets in the clinical symptoms of COVID-19 that cause fever, cough, chest tightness, and diarrhea and sore throat symptoms. After GO functional analysis of the differential genes, it was found that the differential genes were mostly enriched in the type I interferon signaling pathway, defense response to virus, cytoplasm, cytosol, and protein binding. KEGG analysis revealed that the differential genes were mostly enriched in herpes simplex infection, influenza A, and TNF signaling pathways. It is evident that SARS-CoV-2 infection affects the immune response of the host. It had shown that after infection with SARS-CoV-2, uncontrolled inflammatory mediators in the body, causing a cytokine storm, will lead to an excessive immune response, consistent with the results of differential gene enrichment analysis²⁷. In addition, the results of GO classification showed that both TCM drug target genes and virus induced genes were significantly involved in protein binding of molecular functions. Whether TCM can inhibit protein synthesis more effectively, thus hindering the replication and proliferation of virus in the host?

After analyzing the drug targets and differential genes separately, we obtained six core hub genes, namely *IL6*, *CXCL8*, *IL1B*, *NFKBIA*, *PTGS2*, *STAT1*, by correlating them to establish a "disease-target-compound-drug" regulatory network. SARS-CoV-2 causes activation of different immune cells that helps to secrete proinflammatory cytokine *IL6* and other inflammatory cytokines then causes cytokine storm²⁸. Both *IL8* and *IL1B* encode proteins that are cytokines and, like *IL6*, are more abundant in critically ill patients and statistically different ($P < 0.05$)^{27,29}. Analysis of the KEGG pathway for six core hub genes yielded four more important nodes: legionellosis, interleukin-6-mediated signaling pathway, IL-17 signaling pathway, and Leishmaniasis. Among them Legionellosis may involve pneumonia. IL-17 signaling pathway plays an important role in both acute and chronic inflammatory responses. This reveals the mechanism of herbal and Western medicine in the treatment of COVID-19.

Identification of potential molecular targets is essential for drug repurposing³⁰⁻³². Through the construction of molecular docking model between hub host gene and hub compound, it was found that folic acid (FA) and protein ILB, PTGS2, STAT1 were stable binding, and estrogen was stable binding to IL6, IL8. Estrone has good binding properties to proteins, and exogenous estrogen is recommended as a drug for the prevention and treatment of COVID-19²⁶. Folic acid may be a potential drug target. FA had a relatively lower binding energy to the protein, suggesting that it works better as a drug to treat COVID-19. This suggests that folic acid is a potential drug to treat COVID-19. FA have preventive effects of on Zika virus-associated poor pregnancy outcomes in immunocompromised mice. Mice with FA treatment showed lower viral burden and better prognostic profiles in the placenta including reduced inflammatory response, and enhanced integrity of BPB³³. It had indeed been shown to related to pregnant women with COVID-19³⁴. However, the mechanism by which folic acid has apparently protected pregnant women during the COVID-19 pandemic has not been determined.

For the precise treatment of COVID-19, our research performed molecular docking between drug compounds and viral proteins. The role of N in RNA recognition, replicating, transcribing the viral genome, and modulating the host immune response is indispensable³⁵. The *3CL* protease and the *NSP13* helicase are crucial to viral replication^{36,37}. The spike proteins help in fusing into the host and the *NSP9* replicase plays a major role in viral replication³⁸. The *ORF3a* protein can activate the *NLRP3* inflammasome by promoting TRAF3-dependent ubiquitination of *ASC*³⁹. The *ORF7A* protein blocks cell cycle progression at G0/G1 phase via the cyclin D3/pRb pathway⁴⁰. The *ORF8* can mediate immune suppression and evasion activities potentially⁴¹. These proteins play an important role in the viral infection of the host, so they are selected for molecular docking with the effective compounds obtained from the analysis. 3CL Protease and ORF3a protein, ORF8 also have low binding energy with the compounds. The binding of Phosphoprotein N to folic acid was stable, indicating that it can be targeted for molecular therapy. Together, above results showed that the binding energy between FA and N is the lowest, so it was speculated that N might be the target of folic acid in the treatment of COVID-19.

Conclusions

In general, through the analysis of the effective compounds and drug targets of COVID-19 clinical TCM and CWM, we found the pivotal genes and effective compounds, and further established the "drug-target-virus infection" molecular network through the GEO database combined with the virus induced differential genes, and screened out the potential new drugs (FA) and new molecular targets (SARS-CoV-2 N protein and PTGS2) for COVID-19 treatment, which promote underlying the interaction mechanism between virus and host, and provide a new insight for the development of new drugs.

Abbreviations

COVID-19: Coronavirus Disease 2019

TCM: Traditional Chinese medicine

CWM: Conventional Western medicine

SARS-CoV-2: severe acute respiratory syndrome

S: spike

M: membrane

E: envelope

N: nucleocapsid

ACE2: angiotensin-converting enzyme

ORFs: open reading frames

TRSs: transcriptional regulatory sequences

TCMSP: Traditional Chinese Medicines for Systems Pharmacology Database and Analysis Platform

DEGs: differential expression genes

GEO: Gene Expression Omnibus

BP: biological process

MF: molecular function

CC: cellular component

PPI: Protein-protein interaction

ACE: angiotensin converting enzyme

DTGs: drug targeted genes

N: SARS-COV-2, Nucleocapsid Phosphoprotein

TTD: Therapeutic Target Database

FA: folic acid

Declarations

Availability of data and materials

All data and materials in the current study are included in this published article.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Ethics statements

No animal studies are presented in this manuscript. No human studies are presented in this manuscript. No potentially identifiable human images or data is presented in this study.

Consent to publish

All authors agree to publish.

Funding

This work was supported by Science and Technology Planning Project of Yuzhong District of Chongqing City, 2020050, and the Chongqing Municipal Education Commission Foundation, KJQN202000403.

ACKNOWLEDGMENTS

We gratefully thank American Journal Experts for editing this manuscript.

References

- 1 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature***579**, 265-269, doi:10.1038/s41586-020-2008-3 (2020).
- 2 Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet***395**, 470-473, doi:10.1016/S0140-6736(20)30185-9 (2020).
- 3 Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell***181**, 914-921 e910, doi:10.1016/j.cell.2020.04.011 (2020).
- 4 Sekimukai, H. *et al.* Gold nanoparticle-adjuvanted S protein induces a strong antigen-specific IgG response against severe acute respiratory syndrome-related coronavirus infection, but fails to induce protective antibodies and limit eosinophilic infiltration in lungs. *Microbiol Immunol***64**, 33-51, doi:10.1111/1348-0421.12754 (2020).
- 5 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science***367**, 1260-1263, doi:10.1126/science.abb2507 (2020).
- 6 Luan, J., Lu, Y., Jin, X. & Zhang, L. Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochem Biophys Res Commun***526**, 165-169,

doi:10.1016/j.bbrc.2020.03.047 (2020).

7 Wu, F. *et al.* Author Correction: A new coronavirus associated with human respiratory disease in China. *Nature***580**, E7, doi:10.1038/s41586-020-2202-3 (2020).

8 Gao, Q. *et al.* Development of an inactivated vaccine candidate for SARS-CoV-2. *Science***369**, 77-81, doi:10.1126/science.abc1932 (2020).

9 Akalin, E. *et al.* Traditional Chinese medicine practices used in COVID-19 (Sars-cov 2/Coronavirus-19) treatment in clinic and their effects on the cardiovascular system. *Turk Kardiyol Dern Ars***48**, 410-424, doi:10.5543/tkda.2020.03374 (2020).

10 Girija, P. L. T. & Sivan, N. Ayurvedic treatment of COVID-19/SARS-CoV-2: A case report. *J Ayurveda Integr Med*, doi:10.1016/j.jaim.2020.06.001 (2020).

11 Liu, Y. *et al.* Clinical and biochemical indexes from 2019-nCoV infected patients linked to viral loads and lung injury. *Sci China Life Sci***63**, 364-374, doi:10.1007/s11427-020-1643-8 (2020).

12 Commission, G. O. o. C. H. Pneumonia diagnosis and treatment plan for new coronavirus infection_Trial fifth edition. **5** (2020).

13 Li, C., Wang, L. & Ren, L. Antiviral mechanisms of candidate chemical medicines and traditional Chinese medicines for SARS-CoV-2 infection. *Virus Res***286**, 198073, doi:10.1016/j.virusres.2020.198073 (2020).

14 Iorio, F. *et al.* Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proceedings of the National Academy of Sciences of the United States of America***107**, 14621-14626, doi:10.1073/pnas.1000138107 (2010).

15 Jia, Z. *et al.* Cogena, a novel tool for co-expressed gene-set enrichment analysis, applied to drug repositioning and drug mode of action discovery. *BMC Genomics***17**, 414, doi:10.1186/s12864-016-2737-8 (2016).

16 Alaimo, S. & Pulvirenti, A. Network-Based Drug Repositioning: Approaches, Resources, and Research Directions. *Methods Mol Biol***1903**, 97-113, doi:10.1007/978-1-4939-8955-3_6 (2019).

17 Zhang, T. T. *et al.* Network-based drug repositioning: A novel strategy for discovering potential antidepressants and their mode of action. *Eur Neuropsychopharmacol***28**, 1137-1150, doi:10.1016/j.euroneuro.2018.07.096 (2018).

18 Lin, Y. *et al.* Computer-aided biomarker discovery for precision medicine: data resources, models and applications. *Briefings in bioinformatics***20**, 952-975, doi:10.1093/bib/bbx158 (2019).

- 19 Xiaoqing, L. Research progress on pharmacological effects of kaempferol. *Research on Trace Elements and Health***34**, 61-62 (2017).
- 20 Giuseppe, P., Martina, V., Veronica, R. & Emanuela, O. Is diabetes mellitus a risk factor for COVirus Disease 19 (COVID-19). *J Acta diabetologica*. (2020).
- 21 Kumar, G. S., Purushothaman, V., Jaideep, M. & Nalini, D. Increased mortality of COVID-19 infected diabetes patients: role of furin proteases. *J International journal of obesity* (2005). (2020).
- 22 Dinda, B. *et al.* Therapeutic potentials of baicalin and its aglycone, baicalein against inflammatory disorders. *Eur J Med Chem***131**, 68-80, doi:10.1016/j.ejmech.2017.03.004 (2017).
- 23 Li, Y. *et al.* Quercetin, Inflammation and Immunity. *J Nutrients*. **8** (2016).
- 24 Aucoin, M. *et al.* The effect of quercetin on the prevention or treatment of COVID-19 and other respiratory tract infections in humans: A rapid review. *J Advances in Integrative Medicine*. (2020).
- 25 Ezhilarasan, D. Critical role of estrogen in the progression of chronic liver diseases %J *Hepatobiliary & Pancreatic Diseases International*. (2020).
- 26 Zsuzsanna, S. Prevention and therapy of COVID-19 via exogenous estrogen treatment for both male and female patients. *J Journal of pharmacy & pharmaceutical sciences : a publication of the Canadian Society for Pharmaceutical Sciences, Societe canadienne des sciences pharmaceutiques*. **23** (2020).
- 27 Miguel, Q., Jeannie, L. & Shabnam, R. Hazards of the cytokine storm and cytokine-targeted therapy in COVID-19 patients: A Review. *J Journal of medical Internet research*. **22** (2020).
- 28 Chiranjib, C. *et al.* COVID-19: Consider IL-6 receptor antagonist for the therapy of cytokine storm syndrome in SARS-CoV-2 infected patients. *J Journal of medical virology*. (2020).
- 29 Lorenz, C. R. *et al.* COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *J Nature biotechnology*. **38** (2020).
- 30 Wang, J. Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study. *J Chem Inf Model***60**, 3277-3286, doi:10.1021/acs.jcim.0c00179 (2020).
- 31 Skariyachan, S. *et al.* Structural and molecular basis of the interaction mechanism of selected drugs towards multiple targets of SARS-CoV-2 by molecular docking and dynamic simulation studies-deciphering the scope of repurposed drugs. *Comput Biol Med***126**, 104054, doi:10.1016/j.combiomed.2020.104054 (2020).

- 32 Lokhande, K. B., Doiphode, S., Vyas, R. & Swamy, K. V. Molecular docking and simulation studies on SARS-CoV-2 M(pro) reveals Mitoxantrone, Leucovorin, Birinapant, and Dynasore as potent drugs against COVID-19. *J Biomol Struct Dyn*, 1-12, doi:10.1080/07391102.2020.1805019 (2020).
- 33 Simanjuntak, Y., Ko, H. Y., Lee, Y. L., Yu, G. Y. & Lin, Y. L. Preventive effects of folic acid on Zika virus-associated poor pregnancy outcomes in immunocompromised mice. *PLoS Pathog***16**, e1008521, doi:10.1371/journal.ppat.1008521 (2020).
- 34 Jesus, A.-E. & Ricardo, E.-T. The Folate Concentration and/or Folic Acid Metabolites in Plasma as Factor for COVID-19 Infection. *J Frontiers in pharmacology*. **11** (2020).
- 35 Shan, L. *et al.* The SARS-CoV-2 Nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *J bioRxiv : the preprint server for biology*. (2020).
- 36 Fernandes, S. Z., Anson, B. & Mesecar, A. Crystallization of SARS Coronavirus 3CL Protease to Identify Inhibitor Targets. *J The FASEB Journal*. **33** (2019).
- 37 Keum, Y.-S. & Jeong, Y.-J. Development of chemical inhibitors of the SARS coronavirus: Viral helicase as a potential target. *J Biochemical Pharmacology*. **84** (2012).
- 38 Vaishali, C. *et al.* Structure-based drug repurposing for targeting Nsp9 replicase and spike proteins of severe acute respiratory syndrome coronavirus 2. *J Journal of biomolecular structure & dynamics*. (2020).
- 39 Siu, K. L. *et al.* Severe acute respiratory syndrome Coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *J The FASEB Journal*. **33** (2019).
- 40 Schaecher, S. R. *et al.* An immunosuppressed Syrian golden hamster model for SARS-CoV infection. *J Virology*. **380** (2008).
- 41 Structure of SARS-CoV-2 ORF8, a rapidly evolving coronavirus protein implicated in immune evasion. *J bioRxiv : the preprint server for biology*. (2020).

Tables

Table 1 Total number of differentially expressed genes in GSE147507 dataset

Category	Numbers of DEGs		
	Cov002_mock_vs_Co002_CoV2	SARS004_mock_vs_SARS004_CoV2	Venn
Up-regulated	349	677	8
Down-regulated	828	730	105
Total number	1177	1407	113

Table 2 Binding energy (/kcal.mol) of effective compounds to hub genes

Compounds	Chemical formula	IL1B	IL6	IL8	NFKBIA	PTGS2	STAT1
Baicalein	C ₁₅ H ₁₀ O ₅	-7.1	-6.9	-6.1	-7.9	-8.8	-6.2
Estrone	C ₁₈ H ₂₂ O ₂	-6.4	-7.4	-6.5	-8.1	-8.4	-6.5
Quercetin	C ₁₅ H ₁₀ O ₇	-7.1	-7.0	-6.1	-8.4	-9.1	-6.5
Folic acid	C ₁₉ H ₁₉ N ₇ O ₆	-7.2	-6.8	-6.5	-8.5	-10.0	-6.8

Table 3 Binding energy (/kcal.mol) of effective compounds to SARS-Cov-2 related genes

Compounds	Baicalein	Estrone	Quercetin	Folic acid
Chemical formula	C ₁₅ H ₁₀ O ₅	C ₁₈ H ₂₂ O ₂	C ₁₅ H ₁₀ O ₇	C ₁₉ H ₁₉ N ₇ O ₆
ACE	-7.9	-8.3	-8.0	-8.6
ACE2	-6.5	-7.0	-6.0	-7.5
SARS-Cov-2 3CL	-7.8	-8.3	-8.0	-7.7
SARS-Cov-2 helicase-NSP13	-7.6	-7.4	-7.0	-7.9
SARS-Cov-2 Spike	-5.8	-5.5	-5.6	-6.6
SARS-Cov-2 NSP9	-7.3	-7.4	-7.0	-7.4
SARS-Cov-2 Nucleocapsid Phosphoprotein	-8.0	-7.8	-7.8	-8.3
SARS-Cov-2 ORF3a	-7.7	-8.0	-7.5	-7.6
SARS-Cov-2 ORF7a	-5.1	-5.8	-5.6	-6.5
SARS-Cov-2 ORF8	-7.6	-7.9	-7.6	-7.6

Figures

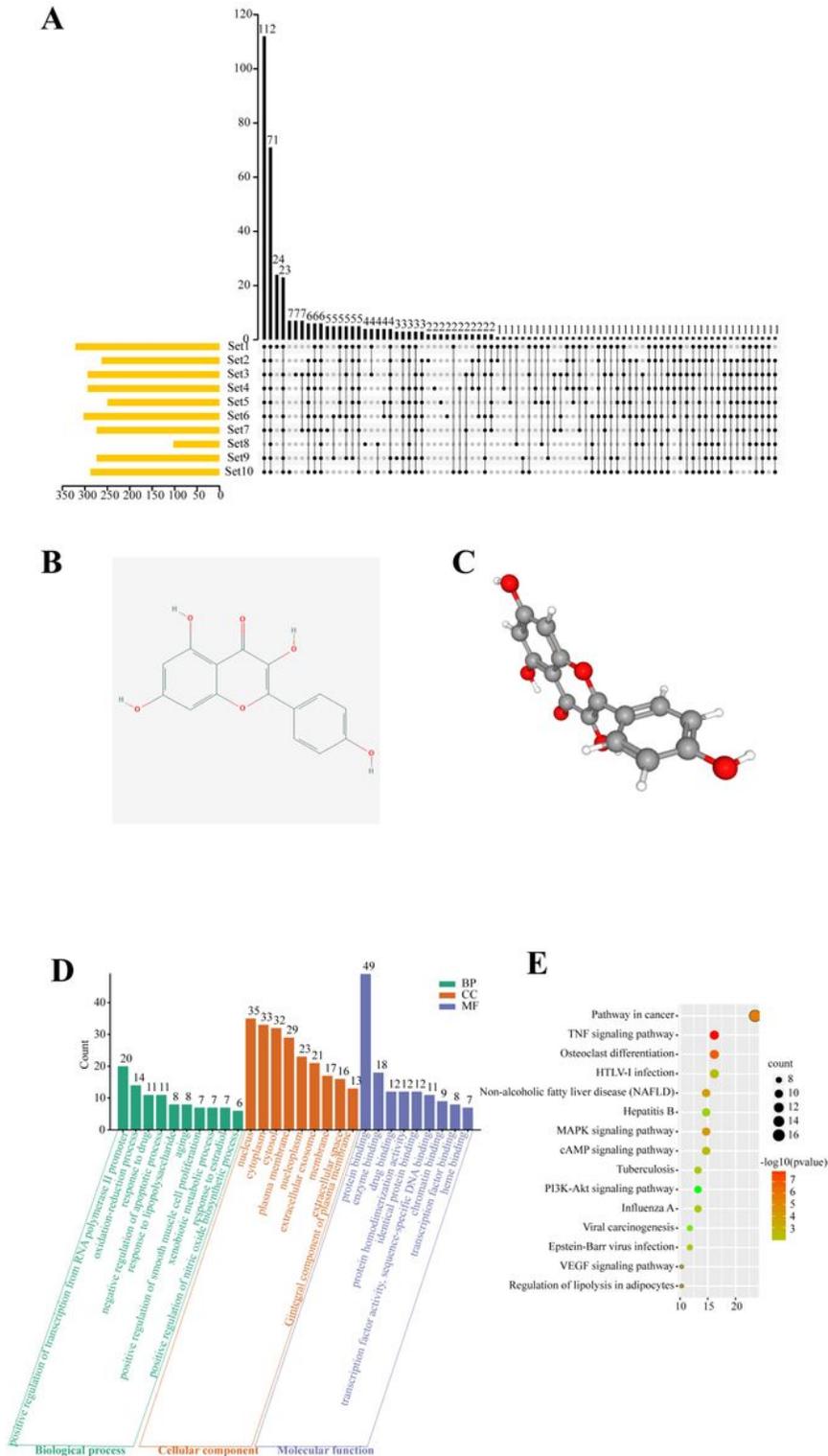


Figure 1

Gene_Ontology (GO) analysis of target genes. (E) KEGG pathway of target genes.

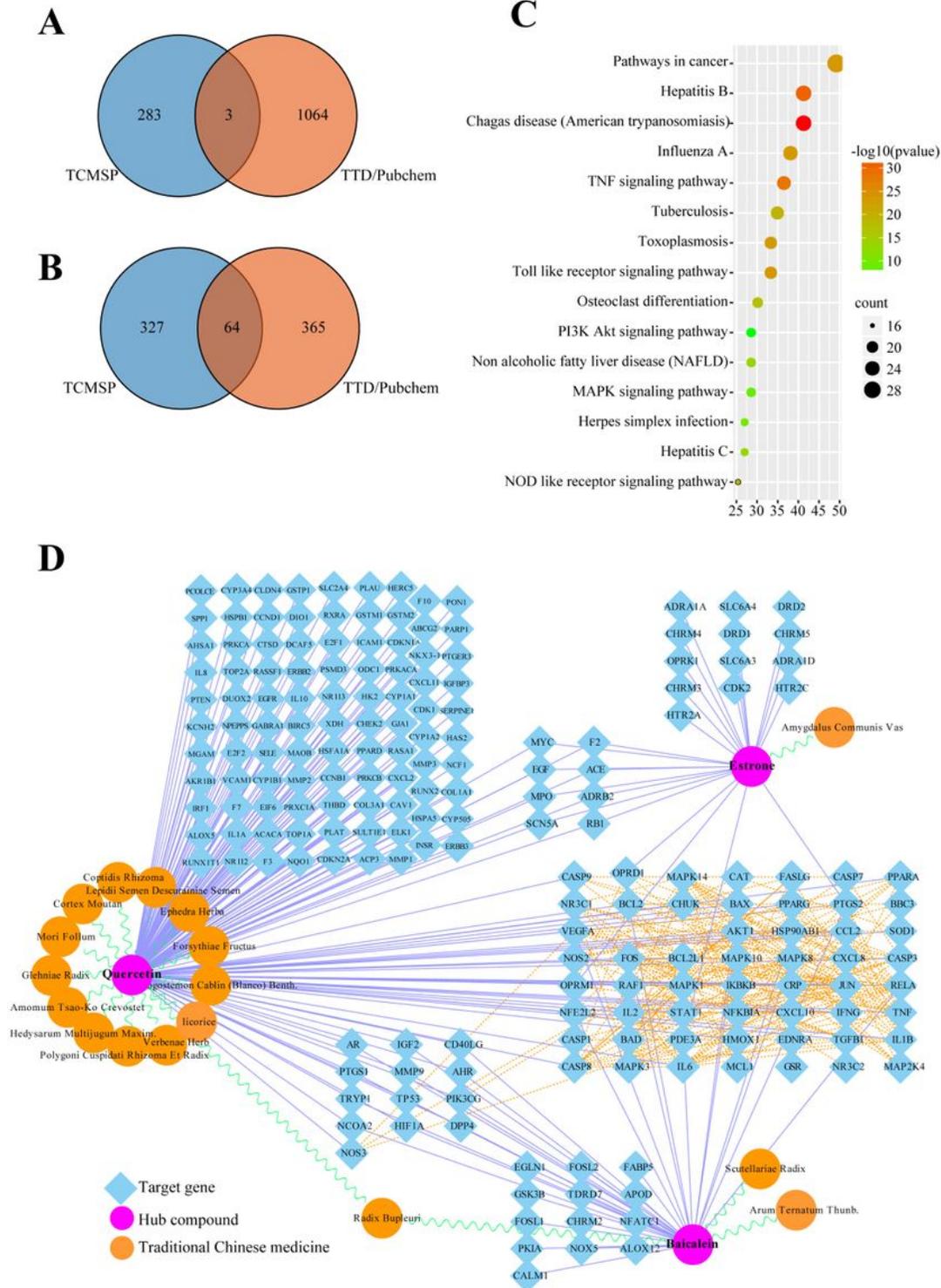


Figure 2

Integrated analysis of COVID-19 clinical TCM and CWM (A) Venn analysis of effective compounds based on TCMSP, TTD and PubChem database. (B) Venn analysis of drug target genes based on TCMSP, TTD and PubChem. (C) KEGG pathway of target genes. (D) Network of “TCM-compounds-target genes”.

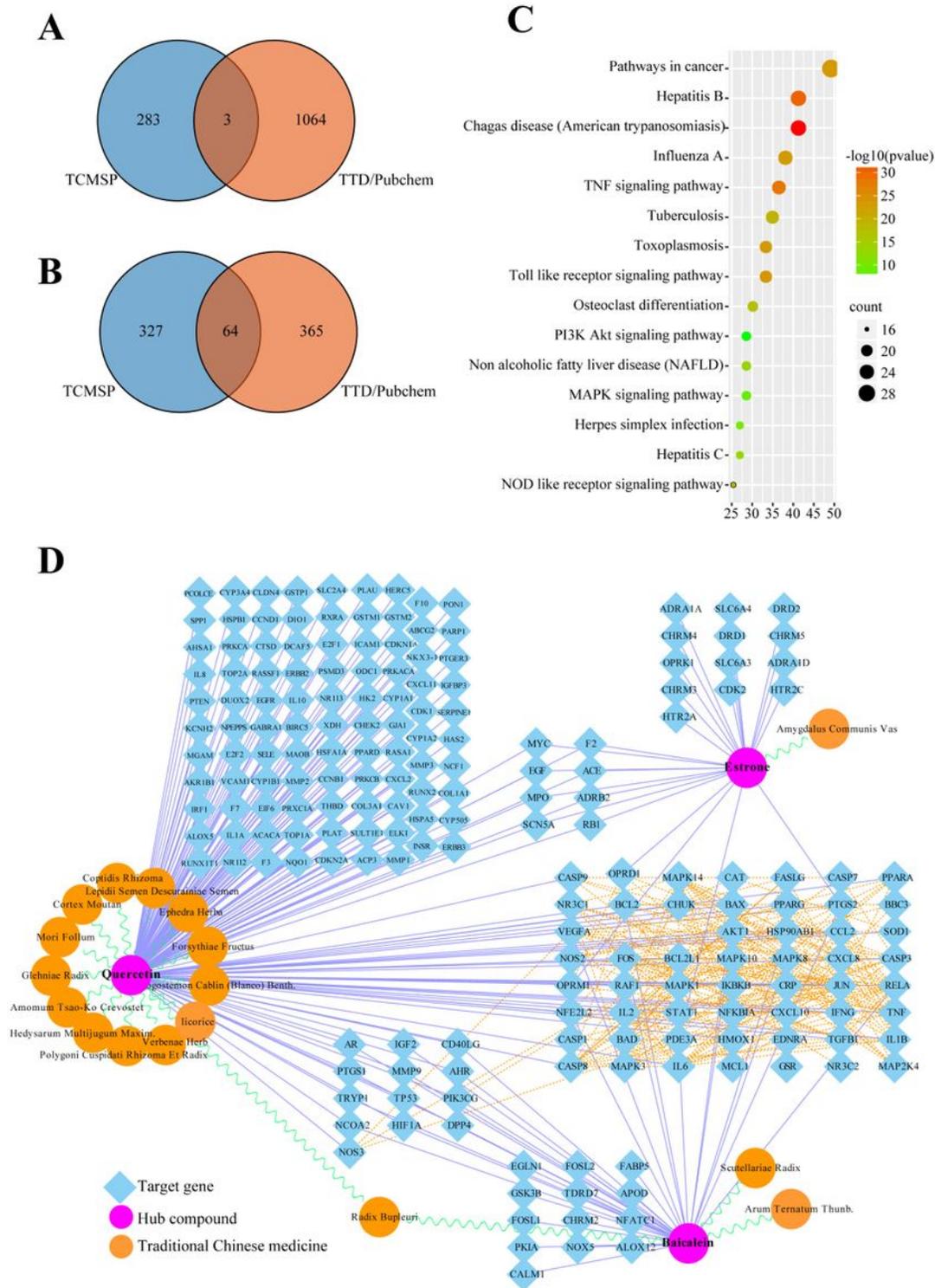


Figure 2

Integrated analysis of COVID-19 clinical TCM and CWM (A) Venn analysis of effective compounds based on TCMSP, TTD and PubChem database. (B) Venn analysis of drug target genes based on TCMSP, TTD and PubChem. (C) KEGG pathway of target genes. (D) Network of “TCM-compounds-target genes”.

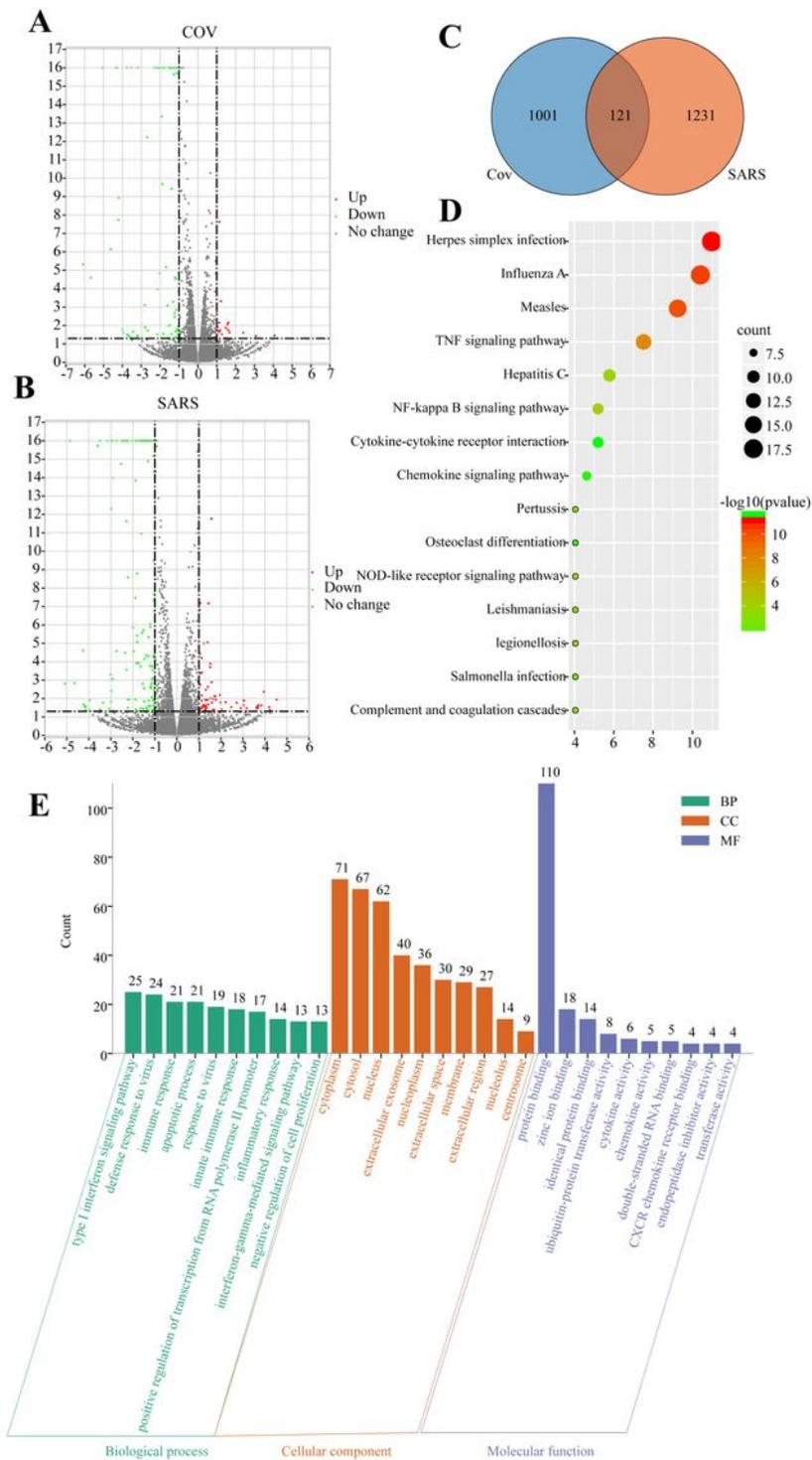


Figure 3

Identification of differentially expressed genes (DEGs) in GSE147507 dataset (A) The volcano plots of the distribution of DEGs in Cov-group. (B) The volcano plots of the distribution of DEGs in SARS-group. (C) Venn analysis of DEGs in Cov- and SARS- groups. (D) KEGG pathway of intersection genes. (E) GO analysis of intersection genes.

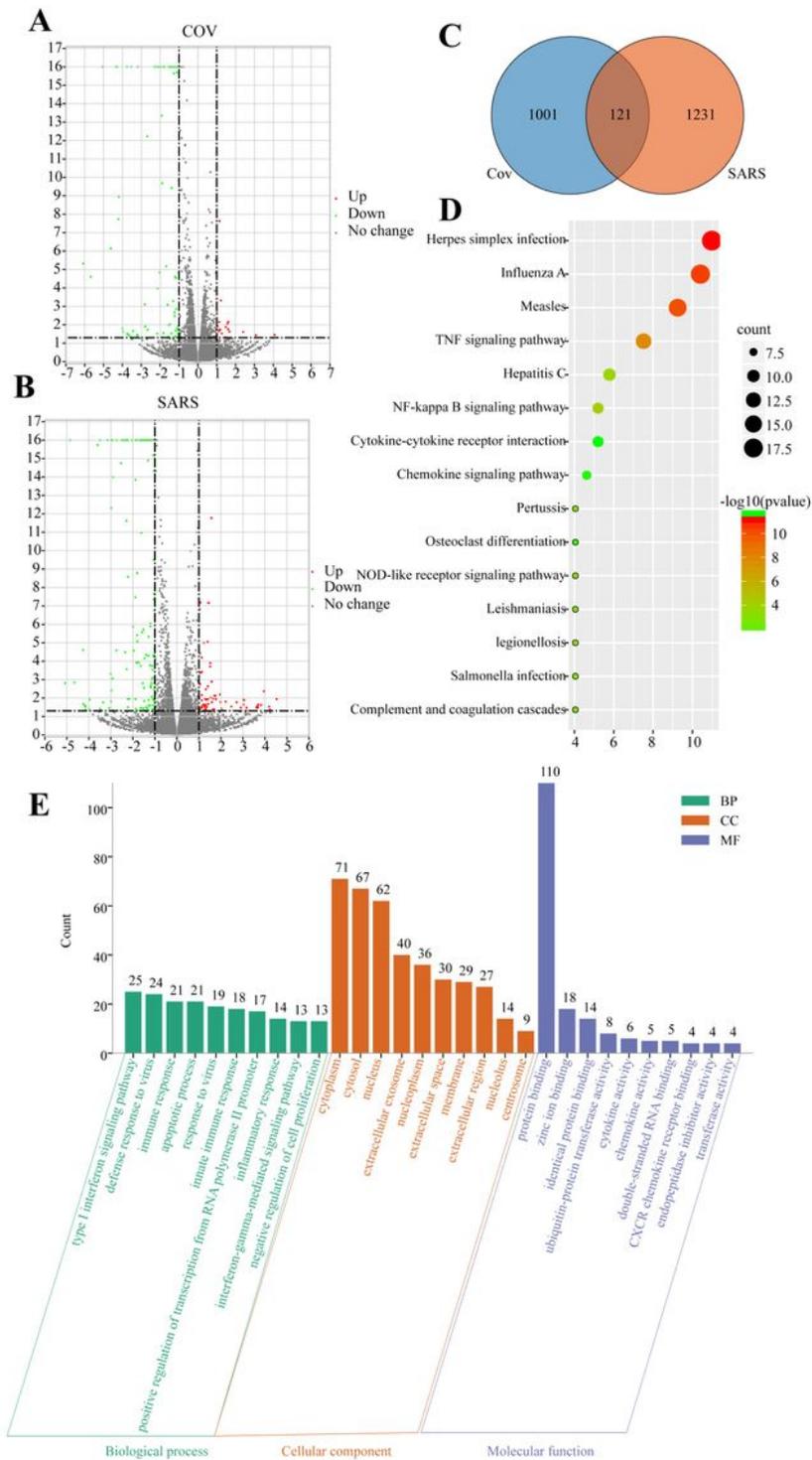


Figure 3

Identification of differentially expressed genes (DEGs) in GSE147507 dataset (A) The volcano plots of the distribution of DEGs in Cov-group. (B) The volcano plots of the distribution of DEGs in SARS-group. (C) Venn analysis of DEGs in Cov- and SARS- groups. (D) KEGG pathway of intersection genes. (E) GO analysis of intersection genes.

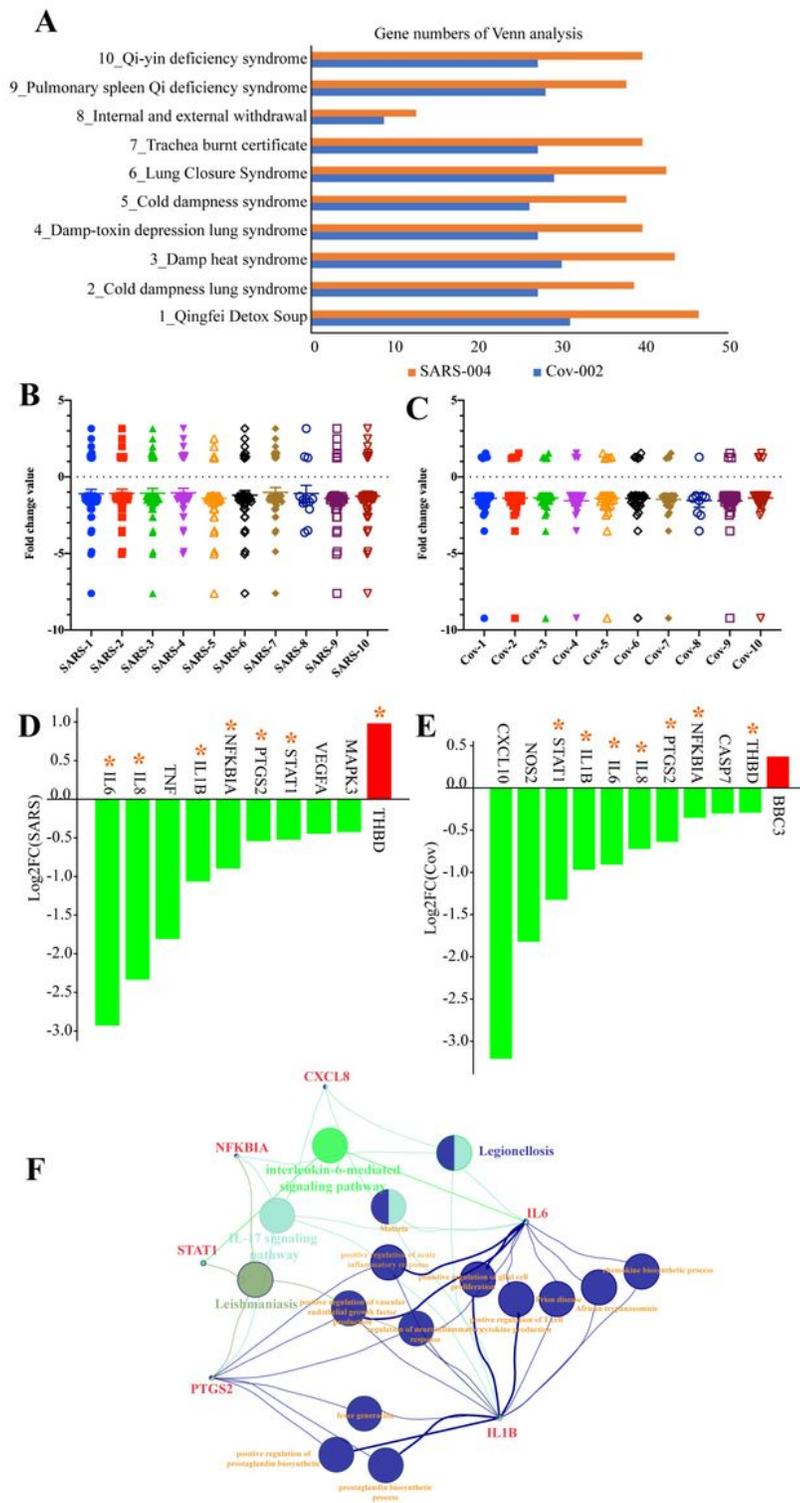


Figure 4

Molecular docking of drug target genes (DTG) and virus induced differential expressed genes (VIDEGs) (A) Venn analysis of DTG and VIDEGs in different TCM prescriptions. (B) The fold change value of “TCM-DTG-VIDEGs” genes in SARS-group. The integrated genes were screened based on TCM target genes and viral induced genes in SARS-group, and the fold change of hub genes were analyzed in GSE147507. (C) The fold change value of “TCM-DTG-VIDEGs” genes in Cov-group. The integrated genes were screened

based on TCM target genes and viral induced genes in Cov-group, and the fold change of hub genes were analyzed in GSE147507. (D) The fold change value of “TTD-TCMSP-GEO” genes in SARS-group. (E) The fold change value of “TTD-TCMSP-GEO” genes in Cov-group. * indicates that the gene is present in both Cov-group and SARS group. (F) The Chemical interaction of key hub genes. We screened chemical interaction of genes that present in venn analysis of “TTD-TCMSP-GEO” and both SARS-group and Cov-group.

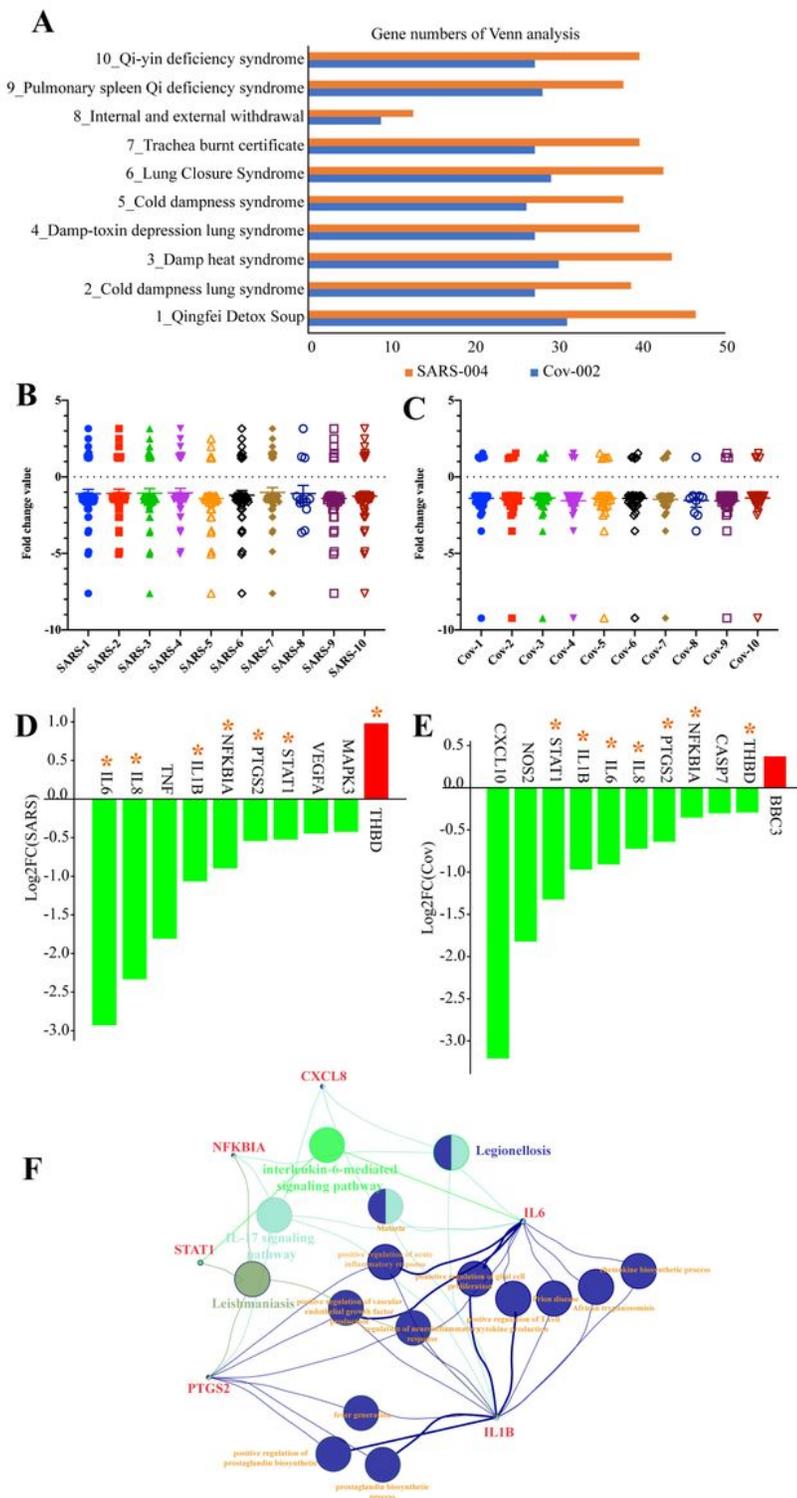


Figure 4

Molecular docking of drug target genes (DTG) and virus induced differential expressed genes (VIDEGs) (A) Venn analysis of DTG and VIDEGs in different TCM prescriptions. (B) The fold change value of “TCM-DTG-VIDEGs” genes in SARS-group. The integrated genes were screened based on TCM target genes and viral induced genes in SARS-group, and the fold change of hub genes were analyzed in GSE147507. (C) The fold change value of “TCM-DTG-VIDEGs” genes in Cov-group. The integrated genes were screened based on TCM target genes and viral induced genes in Cov-group, and the fold change of hub genes were analyzed in GSE147507. (D) The fold change value of “TTD-TCMSP-GEO” genes in SARS-group. (E) The fold change value of “TTD-TCMSP-GEO” genes in Cov-group. * indicates that the gene is present in both Cov-group and SARS group. (F) The Chemical interaction of key hub genes. We screened chemical interaction of genes that present in venn analysis of “TTD-TCMSP-GEO” and both SARS-group and Cov-group.

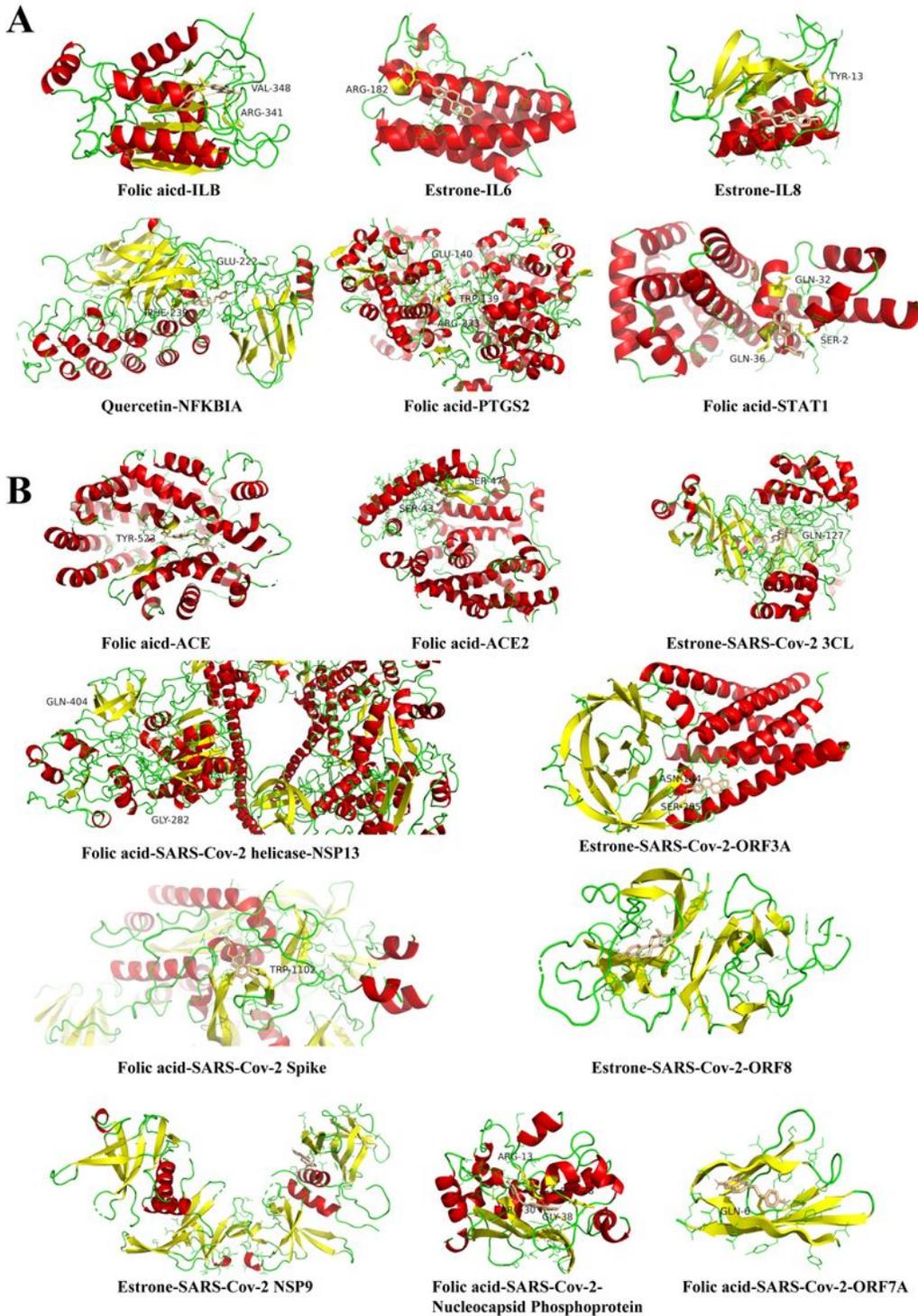


Figure 5

Molecular docking of effective compounds to hub genes and SARS-CoV-2 related genes (A) The three-dimensional pattern that shows the hub genes of effective compounds on the associated proteins. (B) The three-dimensional pattern that shows the molecular docking of the SARS-CoV-2 related genes and effective compounds.

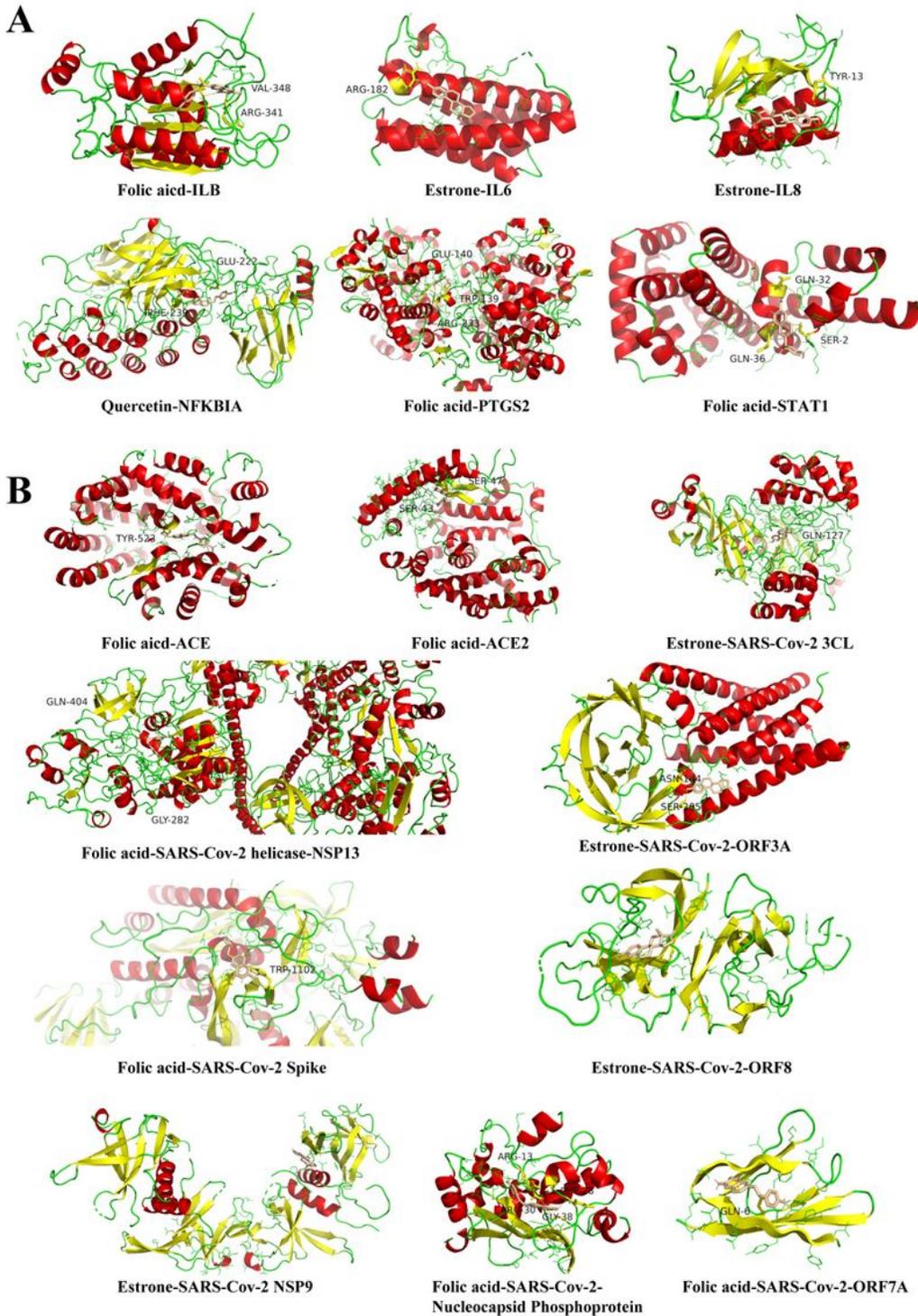


Figure 5

Molecular docking of effective compounds to hub genes and SARS-CoV-2 related genes (A) The three-dimensional pattern that shows the hub genes of effective compounds on the associated proteins. (B) The three-dimensional pattern that shows the molecular docking of the SARS-CoV-2 related genes and effective compounds.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Figure.S1.tif](#)
- [Figure.S1.tif](#)
- [Figure.S2.tif](#)
- [Figure.S2.tif](#)
- [Figure.S3.tif](#)
- [Figure.S3.tif](#)
- [Figure.S4.tif](#)
- [Figure.S4.tif](#)
- [Figure.S5.tif](#)
- [Figure.S5.tif](#)
- [SupplementaryTables.docx](#)
- [SupplementaryTables.docx](#)
- [GraphicalAbstract.tif](#)
- [GraphicalAbstract.tif](#)