ANNALS OF MEDICINE
2025, VOL. 57, NO. 1, 2477301
https://doi.org/10.1080/07853890.2025.2477301

INFECTIOUS DISEASES

OPEN ACCESS Check for updates

Identification and mechanism analysis of biomarkers related to butyrate metabolism in COVID-19 patients

Wenchao Zhou[a,b‡], Hui Li[a‡], Juan Zhang[a‡], Changsheng Liu[c‡], Dan Liu[a], Xupeng Chen[b], Jing Ouyang[d], Tian Zeng[c], Shuang Peng[d], Fan Ouyang[e‡], Yunzhu Long[d] and Yukun Li[a]

[a]Department of Assisted Reproductive Centre, Xiangya Hospital Zhuzhou Central South University, Central South University, Zhuzhou, China; [b]Department of Gynecology, The Second Affiliated Hospital, Hunan Province Key Laboratory of Tumor Cellular & Molecular Pathology, Cancer Research Institute, Hengyang Medical School, University of South China, Hengyang, China; [c]Department of Scientific Research, Xiangya Hospital Zhuzhou Central South University, Central South University, Zhuzhou, China; [d]Department of Infectious Disease, Xiangya Hospital Zhuzhou Central South University, Central South University, Zhuzhou, China; [e]Department of Cardiology, Xiangya Hospital Zhuzhou Central South University, Central South University, Zhuzhou, China

ABSTRACT

Background: Butyrate may inhibit SARS-CoV-2 replication and affect the development of COVID-19. However, there have been no systematic comprehensive analyses of the role of butyrate metabolism-related genes (BMRGs) in COVID-19.

Methods: We performed differential expression analysis of BMRGs in the brain, liver and pancreas of COVID-19 patients and controls in GSE157852 and GSE151803. The differentially expressed genes (DEGs) and module genes between COVID-19 patients and healthy controls in GSE171110 were screened through 'limma' and 'WGANA' R package, respectively, followed by an intersection with BMRGs via 'ggvenn' R package. Six machine learning algorithms were employed to determine the best model for identifying biomarkers, and receiver operating characteristic (ROC) curves were plotted to evaluate the diagnostic value of the biomarkers in COVID-19. Moreover, the differences in immune-infiltrating cells between the COVID-19 and control groups were compared using CIBERSORT. The differences in immune cells and expression levels of biomarkers in immune cells among different tissues were analysed using GSE171668.

Results: The BMRGs were the most different in the brain between the COVID-19 and control groups, including 21 upregulated and 16 downregulated genes. Five important common BMRGs were screened as biomarkers for COVID-19 using XGBoost, namely CCNB1, CCNA2, BRCA1, HBB and HSPA5, with increased diagnostic performance. Enrichment analysis revealed that these five genes were related to the cell cycle, cell proliferation and cell senescence. The infiltrating abundance of 12 immune cells was different between the COVID-19 and control groups. Finally, the expression levels of HSPA5, BRCA1 and HBB were higher in annotated cells than in CCNB1 and CCNA2, and there were four different types of immune cells in the liver, heart, lungs and kidneys.

Conclusions: These five genes may be potential biomarkers of butyrate metabolism in COVID-19 patients. These findings provide a direction for further studies on the molecular mechanisms underlying COVID-19.

ARTICLE HISTORY
Received 1 February 2024
Revised 8 February 2025
Accepted 10 February 2025

KEYWORDS
COVID-19; butyrate metabolism; biomarkers; organoid; single-cell sequencing

1. Background

The advent of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has precipitated the global outbreak of coronavirus disease 2019 (COVID-19) [1]. The genetic makeup of SARS-CoV-2 exhibits the closest resembles that of SARS-CoV, which first appeared in 2003 [2–4]. The primary modes of transmission of this infection include direct contact, inhalation of droplets, and contact with oral, nasal and ocular mucosal surfaces. Typically, respiratory infections manifest as mild to moderate symptoms such as fever, cough and myalgia [1]. However, in severe cases, these symptoms can progress to pneumonia, potentially resulting in fatality [5]. The elevated incidence of COVID-19 poses a significant threat to the well-being and survival of

CONTACT Yunzhu Long 2863454482@qq.com Department of Infectious Disease, Zhuzhou Central Hospital, Xiangya Hospital Zhuzhou Central South University, Central South University, Zhuzhou, Hunan 412000, PR China; Yukun Li yukun_li@foxmail.com; yukun_li@csu.edu.cn Department of Assisted Reproductive Centre, Zhuzhou Central Hospital, Xiangya Hospital Zhuzhou Central South University, Central South University, Zhuzhou, Hunan 412000, PR China
‡These authors contributed equally to this work.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07853890.2025.2477301.

individuals [6], leading to considerable distress. Individuals with underlying health conditions experience more severe disease progression and an increased likelihood of mortality [7]. Notably, diabetes, hypertension, advanced age, immunodeficiency and cardiovascular ailments are all risk factors associated with severe COVID-19 [8]. SARS-CoV-2 consists of 20 distinct proteins, encompassing four primary structural proteins (S: spike, E: envelope, M: membrane and N: nucleocapsid), as well as numerous nonstructural proteins. For instance, RNA-dependent RNA polymerase, coronavirus main protease and papain-like protease (PLpro) have been identified as potential targets for antiviral drug development [9–11]. Despite the initial use of lopinavir and ritonavir as therapeutic drugs in clinical studies, their efficacy in treating COVID-19 patients has not gained public recognition [12].

Butyrate, a short-chain fatty acid (SCFA) with four carbon atoms, is produced by the metabolism of intestinal microbial fibres. It can accumulate in the colon at concentrations as high as 140 mm/L and is found in venous blood and peripheral tissues [13]. Butyrate plays an important role in regulating the immune system and protecting the host from various immune-related diseases by modulating intestinal immunity, exerting anti-inflammatory effects and enhancing intestinal barrier function [14,15]. However, recent studies have shown that butyrate also has certain effects on other viral infections. For example, Zhong et al. found that butyrate inhibited porcine epidemic diarrhea virus (PEDV) infection by regulating cellular transcriptional responses and signalling pathways [16]. In addition, butyrate helps improve the overall health of HIV-infected people and reduces their risk of developing AIDS complications [17]. Studies have also found that butyrate has the potential to enhance macrophage immune function, especially in anti-fungal infections [18]. In the microglial cell model infected with Japanese encephalitis virus (JEV), butyrate played an anti-inflammatory role, providing new ideas for the potential therapeutic role of SCFAs in viral neuroinflammation [19]. In addition, butyrate exhibited anti-tumour effects in inhibiting Marek's disease virus (MDV) replication, reducing tumour formation and regulating lymphoma cell growth [20]. Similarly, butyrate also has potential in the prevention and control of bovine viral diarrhea virus (BVDV) [21]. Conversely, experimental evidence has demonstrated that butyrate confers protection against influenza virus-induced pathology in mice [22] while also exhibiting the capacity to modulate Th9 cells, thereby reducing lung inflammation and mucus production [23].

To comprehend the intricacies of butyrate metabolism in individuals afflicted with COVID-19, machine learning and other bioinformatics analysis techniques were employed to identify biomarkers in COVID-19 patients. This endeavour may yield novel insights into the judicious management of COVID-19 patients and the advancement of therapeutic interventions.

2. Materials and methods

2.1. Acquisition of sequencing data and target genes

The expression data from high-throughput sequencing and clinical information of the GSE171110 [24] training set and GSE157344 [25] validation set were acquired from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/). In the training set, whole blood samples from 44 COVID-19 patients and 10 healthy donors were selected for follow-up analysis. Meanwhile, blood cell samples from 27 COVID-19 patients and six healthy donors were analysed in the external validation set GSE157344. Furthermore, GSE157852 (brain organoids of six SARS-CoV-2 infected and three controls) [26] and GSE151803 (six pairs of livers and three pairs of pancreases with mock-infected versus SARS-CoV-2 infected) were singled out from the GEO database and were utilized to analyse the differential expression of butyrate metabolism-related genes (BMRGs) in different organs. The GSE171668 dataset was also retrieved from the GEO database for single-cell analysis of different organs, including 24 lung, 16 liver, 19 heart and 16 kidney tissue with COVID-19 [27]. Finally, 346 BMRGs (score >2) were collected from the GeneCards database (https://www.genecards.org/). The data used in this study was obtained from public database and the ethnicity was European (GSE171110 was released on 27 May 2021; GSE157344 was released on 27 January 2021; GSE157852 was released on 23 September 2020; GSE151803 was released on 12 June 2020; and GSE171668 was released on 9 April 2021). The analysis of this study started on 17 January 2023 and ended on 9 June 2023.

2.2. Enrichment analysis of differentially expressed BMRGs in different organs

In GSE157852 and GSE151803, the differential expression analysis of BMRGs in brain, liver and pancreas between infection and control were recognized by 'limma' package in R software (R Foundation for Statistical Computing, Vienna, Austria), with $|\log_2 FC| > 0.5$ and adj.$p < .05$ as the thresholds. A volcano plot

was drawn for all BMRGs based on log_2FC and *P*, using the R package 'ggplot2.' The R 'ggpubr' package was used for visualizing gene expression pattern of the top 10 up- and down-regulated BMRGs. To probe the biological features and signalling pathways of differentially expressed BMRGs, the 'clusterProfiler' R package was utilized for enrichment analysis of the strength of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (adj.$p < .05$) [27]. Then, the R package 'GOplot' was used to show the first 10 entries in the enrichment results using the gossip graph.

2.3. Enrichment analysis of common DE-BMRGs by differential expression analysis and weighted correlation network analysis (WGCNA) in GSE171110

First, we authenticated the differentially expressed genes (DEGs) between COVID-19 patients and healthy controls in GSE171110 by 'limma' R package with $|log_2FC| > 0.5$ and $p < .05$ as the thresholds. The R package 'ggplot2' and 'pheatmap' were used for volcano plot and heatmap, respectively. Second, the R package 'WGCNA' was used to detect outlier samples in GSE171110 and construct a co-expression network based on the gene expression matrix and sample grouping information. The module genes with the highest correlation with COVID-19 were selected based on the heatmap of module–trait relationships. A scatter plot was used to display the correlation between module membership (MM) and gene significance (GS). In the next moment, the intersection of DEGs, module genes and 346 BMRGs was obtained via 'ggvenn' R package, named DE-BMRGs. In the same way, the enrichment analysis of DE-BMRGs was performed by 'clusterProfiler' R package [27].

2.4. Screening and diagnostic value assessment of the biomarkers

Based on the expression matrix of DE-BMRGs and sample grouping information, six machine learning algorithms (AdaBoost [28], ExtraTrees [29], logistic regression [30], naïve Bayes classifier [31], random forest [32] and XGBoost [33]) were used to identify biomarkers. The receiver operating characteristic (ROC) curves of six algorithms were measured using 'pROC' R package, and the area under the curve (AUC) was estimated as well. By adjusting the parameter combinations of each model, the accuracy of each combination was calculated and compared to determine the best performing model. Meanwhile, the performance of these six machine learning models was compared and validated using 5-fold cross-validation (5-fold CV) and leave-one-out cross-validation (LOOCV) methods. Subsequently, the best performing model was screened based on the metric values of the six models for subsequent studies. Next, the importance degree of DE-BMRGs was then calculated, and the top five genes were defined as biomarkers. For assessing the diagnostic value of biomarkers in COVID-19, the ROC curves were measured using 'pROC' R package in the GSE171110 accompanied with estimated AUC values. Simultaneously, the expression levels of the aforementioned biomarkers among COVID-19 and control samples were compared in the GSE171110 training and GSE157344 validation sets. Finally, based on the biomarkers expression matrices and samples grouping information in GSE171110 and GSE157344, the data were first resampled using the bootstrapping method with a sampling number of 100 times. Then, artificial neural network (ANN) analysis was performed using the R package 'neuralnet' (v 1.33), the number of hidden layers was set to 4, and ROC curves were plotted to evaluate the classification performance of the ANN model.

2.5. Immune infiltration analysis

In GSE171110, we determined the abundance of 22 immune infiltrating cells from each sample in COVID-19 patients and controls using the CIBERSORT algorithm. The differential immune cells between the COVID-19 and control groups were identified, and a correlation analysis was performed for COVID-19.

2.6. Single cell analysis

To further verify the composition of different immune cells in different tissues, we collected single-cell data of the lung, liver, heart and kidney tissues in GSE171110. First, the scRNA-seq dataset GSE171110 of COVID-19 was filtered by 'Seurat' package according to the following criteria: (1) the number of covered cells less than 3; (2) the amount of expressed genes was lower than 100 or higher than 10,000. After data standardization, the top 2000 highly variable genes were identified and the data were subjected to principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) dimensionality reduction analysis. Following search for the marker genes of each cell cluster, R package 'SingleR' was used for cell annotation and cell type identification. Subsequently, the gene expression levels of the biomarkers in the annotated cells in different tissues were analysed. Finally, we

implemented an inter-organ differential analysis of discrepant immune cells based on the proportion of cells annotated in each organ.

2.7. The serum sample of COVID-19 patients

Serum samples were collected from 312 patients with COVID-19 and 50 normal volunteers from Xiangya Hospital Zhuzhou Central South University, Central South University (Zhuzhou, China) from December 2022 to April 2023, which was approved by the Research Ethics Committee of Xiangya Hospital Zhuzhou Central South University. Patients with COVID-19 were recruited if they tested positive for COVID-19 and were not receiving any treatment. Protocols were approved under Xiangya Hospital Zhuzhou Central South University IRB protocols # ZZCHEC2022042-01.

2.8. ELISA analysis

The operation was carried out strictly according to the instructions.

The kit employs purified Human HSPA5, CCNB1, CCNA2, BRCA1 or HBB antibodies to immobilize the protein in microtiter plate wells, forming a solid-phase antibody. Subsequently, it was added to the wells, followed by the addition of an HRP-labelled antibody, resulting in the formation of an antibody–antigen–enzyme–antibody complex. After thorough washing, TMB substrate solution was added, causing the HRP enzyme to catalyse a reaction that turns the substrate blue. The reaction was then terminated by the addition of sulphuric acid solution, and the resulting colour change was measured spectrophotometrically at a wavelength of 450 nm. The levels of HSPA5, CCNB1, CCNA2, BRCA1 and HBB in the samples were determined by comparing the optical density (O.D.) of the samples to the standard curve. The ELISA kits used for this analysis were procured from Shanghai Enzyme-linked Biotechnology Co., Ltd. (Shanghai, China), with specific catalogue numbers assigned to each biomarker kit (HSPA5, YJ365599; CCNB1, YJ304452; CCNA2, YJ960233; BRCA1, YJ330515; and HBB, YJ023056).

2.9. Statistical analysis

R software (R Foundation for Statistical Computing, Vienna, Austria) was used for statistical analysis. Differences were analysed using the Wilcoxon test. If not specifically stated, $p < .05$.

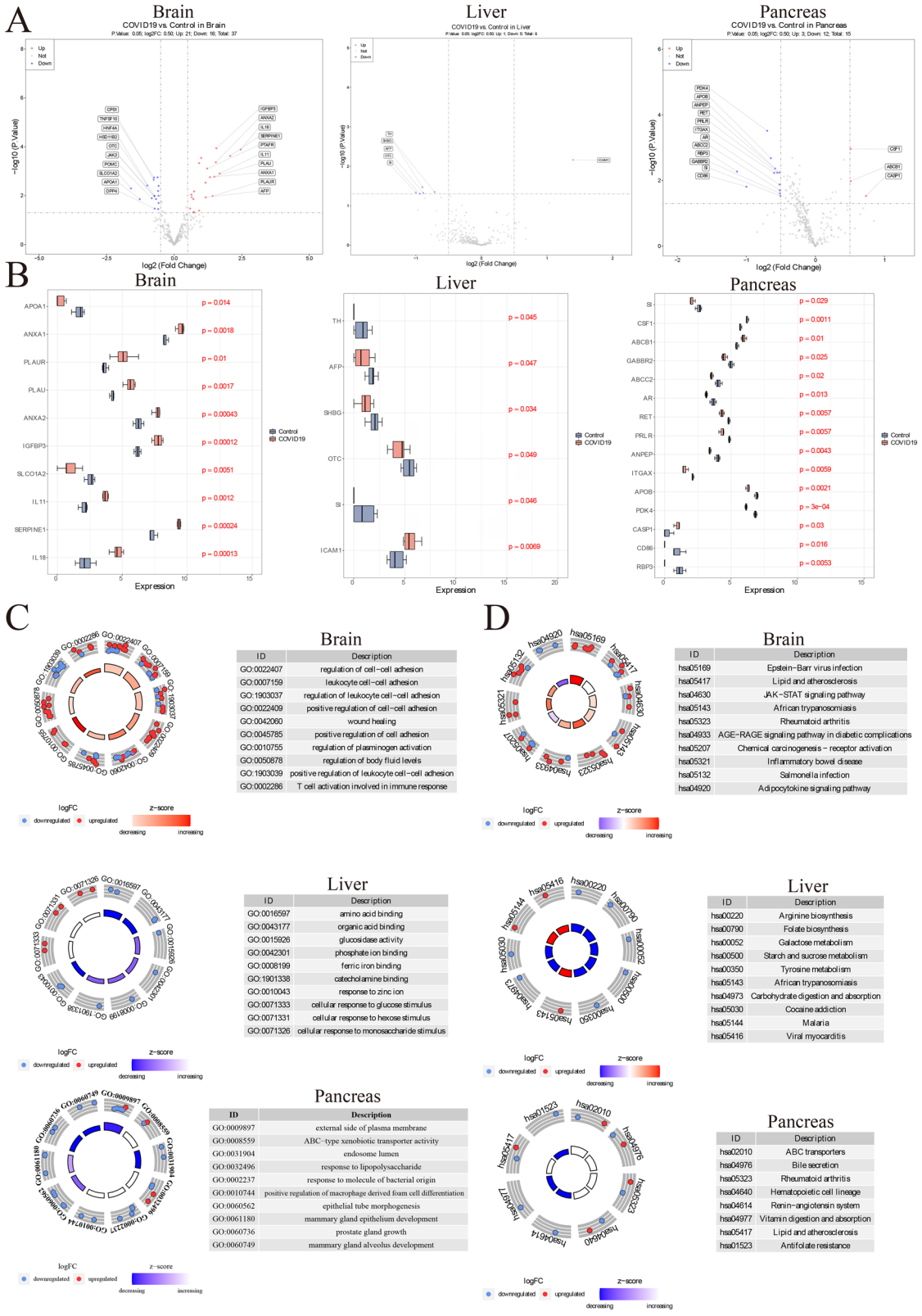
3. Results

3.1. Differential expression analysis of BMRGs in different organs

There were 37, 6 and 15 differentially expressed BMRGs in brain, liver and pancreas between COVID-19 infection and control respectively (Supplementary Table S1). The top 10 altered genes in the brain and all altered genes in the liver and pancreas were labelled within the volcano plot (Figure 1(A)), and their gene expression profiles are shown via box plots (Figure 1(B)). For differentially expressed BMRGs in the brain, 1076 GO items (such as regulation of cell–cell adhesion, leukocyte cell–cell adhesion and regulation of leukocyte cell–cell adhesion) and 50 KEGG signalling pathways (such as Epstein–Barr virus infection, lipid and atherosclerosis, and JAK-STAT signalling pathways) were significantly enriched (Figure 1(C,D)). For differentially expressed BMRGs in the liver, 213 GO items (such as amino acid binding, organic acid binding and ferric iron binding) and 13 KEGG signalling pathways (such as arginine biosynthesis, folate biosynthesis and galactose metabolism) were significantly enriched (Figure 1(C,D)). For differentially expressed BMRGs in the pancreas, 432 GO items (such as the external side of the plasma membrane, ABC-type xenobiotic transporter activity and endosome lumen) and eight KEGG signalling pathways (such as ABC transporters, bile secretion and rheumatoid arthritis) were significantly enriched (Figure 1(C,D)).

3.2. Screening results of DE-BMRGs

Two thousand and sixty-eight DEGs between COVID-19 and control samples from the GSE171110 dataset were screened via differential expression analysis, including 1207 upregulated and 861 downregulated genes. In the volcano plot, we identified the top 10 upregulated and downregulated genes (Figure 2(A)), and the expression trend of these genes is shown via a heat map (Figure 2(B)). WGCNA was performed using the GSE171110 dataset. Sample clustering analysis revealed an outlier (Figure 2(C)). The β value was 8 (Figure 2(D)). The cluster dendrogram of module distribution and co-expression network heatmap are shown in Figure 2(E,F). The green module was markedly correlated with COVID-19 (cor = 0.68, p.adj = 1×10^{-7}), contained 943 genes (Figure 2(F)). In addition, we found a strong positive correlation between MM and GS (Figure 2(G)). Eighteen DE-BMRGs were identified by intersecting 2068 DEGs, 943 modular genes and 346 BMRGs, including CCNB1, CCNA2, HSP90B1, HSPA5, CXCL8,



Brain	
ID	Description
GO:0022407	regulation of cell−cell adhesion
GO:0007159	leukocyte cell−cell adhesion
GO:1903037	regulation of leukocyte cell−cell adhesion
GO:0022409	positive regulation of cell−cell adhesion
GO:0042060	wound healing
GO:0045785	positive regulation of cell adhesion
GO:0010755	regulation of plasminogen activation
GO:0050878	regulation of body fluid levels
GO:1903039	positive regulation of leukocyte cell−cell adhesion
GO:0002286	T cell activation involved in immune response

Liver	
ID	Description
GO:0016597	amino acid binding
GO:0043177	organic acid binding
GO:0015926	glucosidase activity
GO:0042301	phosphate ion binding
GO:0008199	ferric iron binding
GO:1901338	catecholamine binding
GO:0010043	response to zinc ion
GO:0071333	cellular response to glucose stimulus
GO:0071331	cellular response to hexose stimulus
GO:0071326	cellular response to monosaccharide stimulus

Pancreas	
ID	Description
GO:0009897	external side of plasma membrane
GO:0008559	ABC−type xenobiotic transporter activity
GO:0031904	endosome lumen
GO:0032496	response to lipopolysaccharide
GO:0002237	response to molecule of bacterial origin
GO:0010744	positive regulation of macrophage derived foam cell differentiation
GO:0060562	epithelial tube morphogenesis
GO:0061180	mammary gland epithelium development
GO:0060736	prostate gland growth
GO:0060749	mammary gland alveolus development

Brain	
ID	Description
hsa05169	Epstein−Barr virus infection
hsa05417	Lipid and atherosclerosis
hsa04630	JAK−STAT signaling pathway
hsa05143	African trypanosomiasis
hsa05323	Rheumatoid arthritis
hsa04933	AGE−RAGE signaling pathway in diabetic complications
hsa05207	Chemical carcinogenesis − receptor activation
hsa05321	Inflammatory bowel disease
hsa05132	Salmonella infection
hsa04920	Adipocytokine signaling pathway

Liver	
ID	Description
hsa00220	Arginine biosynthesis
hsa00790	Folate biosynthesis
hsa00052	Galactose metabolism
hsa00500	Starch and sucrose metabolism
hsa00350	Tyrosine metabolism
hsa05143	African trypanosomiasis
hsa04973	Carbohydrate digestion and absorption
hsa05030	Cocaine addiction
hsa05144	Malaria
hsa05416	Viral myocarditis

Pancreas	
ID	Description
hsa02010	ABC transporters
hsa04976	Bile secretion
hsa05323	Rheumatoid arthritis
hsa04640	Hematopoietic cell lineage
hsa04614	Renin−angiotensin system
hsa04977	Vitamin digestion and absorption
hsa05417	Lipid and atherosclerosis
hsa01523	Antifolate resistance

Figure 1. Differential expression analysis of BMRGs in different organs. (A) The expression level of BMRGs in brain, liver and pancreas by volcano plots. Each point in the graph represented a gene, grey points indicated genes with insignificant differential expression, red points indicated significantly up-regulated genes and blue points indicated significantly down-regulated genes. (B) The significantly differential expression level of BMRGs in brain, liver and pancreas. (C) GO enrichment for these differential expression BMRGs in brain, liver and pancreas. (D) KEGG enrichment for these differential expression BMRGs in brain, liver and pancreas.

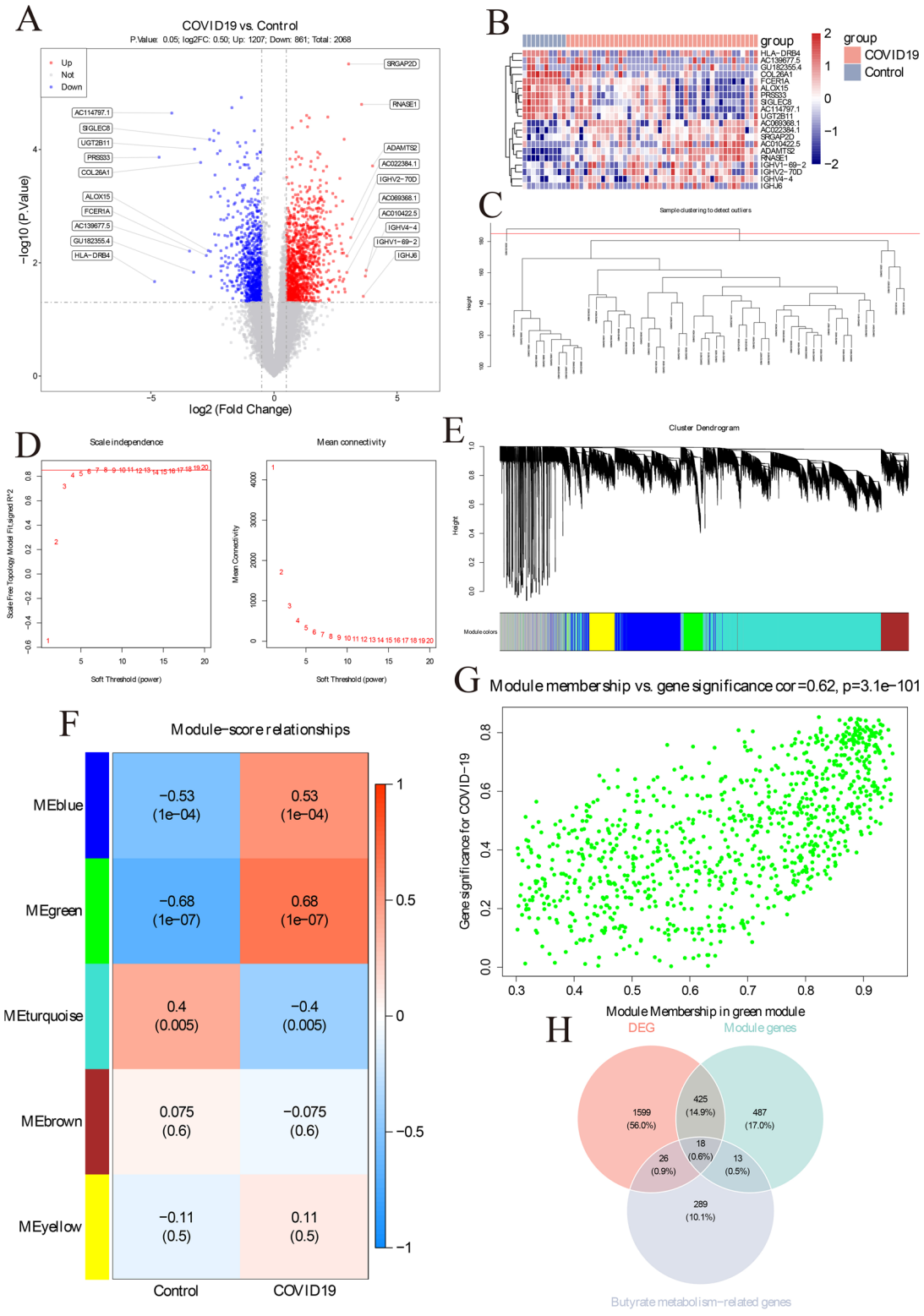


Figure 2. The screening of DE-BMRGs in COVID-19 by WGCNA and differential expression analysis. (A) The expression level of genes between COVID-19 patients and normal people by volcano plots. Each point in the graph represented a gene, grey points indicated genes with insignificant differential expression, red points indicated significantly up-regulated genes and blue points indicated significantly down-regulated genes. (B) Differential expression genes of COVID-19 were shown by a heat map. Different colours indicated different levels of gene expression, the closer to red the higher the level of gene expression, the closer to blue the lower the level of expression. (C) Outlier detection algorithm for GSE171110. (D) Soft threshold screening plot for GSE171110. (E) Construction of a weighted co-expression network for GSE171110. (F) Module-disease correlation analysis based on WGCNA network. Red squares indicated positive correlation and blue indicated negative correlation. (G) Gene importance (GS) – module membership (MM) scatter plot for hub module. (H) Acquisition of shared genes among DEG, module genes and butyrate metabolism-related genes by Venn plot.

HBA2, TERT, IGF1, HBB, CDKN2A, BRCA1, CDK1, BIRC5, CDKN3, TYMS, RAD51, E2F1 and IL2 (Figure 2(H), Supplementary Table S2).

3.3. Enrichment analysis of DE-BMRGs

We performed GO and KEGG enrichment analyses of the above 18 DE-BMRGs to investigate their involvement in biological functions and related signalling pathways. A total of 266 GO items were significantly enriched, including 216 BP items, 36 CC items and 14 MF items (Supplementary Table S3). DE-BMRGs related to BP were primarily enriched in mitotic cell cycle phase transition, positive regulation of fibroblast proliferation, and responses to toxic substances. DE-BMRGs related to CC mainly focused on the endocytic vesicle lumen, cyclin-dependent protein kinase holoenzyme complex and chromosomal region. In the MF term, DE-BMRGs were significantly concentrated in cyclin-dependent protein serine/threonine kinase regulatory activity, haptoglobin binding and kinase regulatory activity (Figure 3(A)). A total of 42 KEGG signalling pathways were significantly enriched for 18 DE-BMRGs (Supplementary Table S4), including cellular senescence, the cell cycle and human T-cell leukaemia virus 1 infection (Figure 3(B)).

3.4. Biomarkers screening and evaluation

All the AUC values of the ROC curves of the six machine learning models were greater than 0.95, indicating that these models possessed satisfactory prediction accuracy (Supplementary Figure S1). Based on the performance indices and metric values of the six models, we selected XGBoost for follow-up biomarker identification (Table 1, Figure 4(A)). By calculating the importance of the 18 DE-BMRGs, the top five genes (CCNB1, CCNA2, BRCA1, HBB and HSPA5) were screened as biomarkers (Figure 4(B)). Subsequently, the AUC values of all five genes (CCNB1: 0.990, CCNA2: 0.993, BRCA1: 0.988, HBB: 0.990 and HSPA5: 0.911) in GSE171110 were greater than 0.9. This signified excellent diagnostic value for COVID-19 (Figure 4(C)). More, the sensitivity, specificity, accuracy and 95% confidence intervals of the ROC analyses for each biomarker were CCNB1 (1, 0.977, 0.988, 0.965–0.993), CCNA2 (1, 0.954, 0.977, 0.975–0.995), BRCA1 (1, 0.909, 0.954, 0.959–0.990), HBB (1, 0.909, 0.954, 0.965–0.993) and HSPA5 (1, 0.795, 0.897, 0.820–0.913). The expression trends of HBB, CCNB1, CCNA2 and BRCA1 were consistent in both the GSE171110 training set and the GSE157344 validation set, except for HSPA5 (Figure 4(D)). Finally, an ANN diagnostic model was constructed based on the GSE171110 and GSE157344 datasets, which contains an input layer of five nodes, a hidden layer of four nodes, and an output layer of two nodes (Figure 4(E)). The AUC values were 1.00 and 0.85 in the training dataset and validation dataset, respectively (Figure 4(F)), indicating that the ANN model has a good performance in diagnosing COVID-19.

3.5. Enrichment analysis of biomarkers

To investigate the possible biological functions of five biomarkers, GO and KEGG enrichment analysis were also executed via R package 'clusterProfiler.' A total of 143 BP, 14 CC and nine MF items were significantly enriched in the GO database (Supplementary Table S5). G2/M transition of mitotic cell cycle, cell cycle G2/M phase transition and cellular response to inorganic substances were included in the BP items. In the CC term, the five biomarkers were significantly concentrated in the cyclin-dependent protein kinase holoenzyme complex, serine/threonine protein kinase complex and protein kinase complex. Biomarkers related to MF were mainly focused on ubiquitin-like protein ligase binding, cyclin-dependent protein serine/threonine kinase regulatory activity and protein kinase regulatory activity (Figure 5(A)). Furthermore, progesterone-mediated oocyte maturation, cell cycle and cellular senescence were associated with these biomarkers in the KEGG signalling pathways (Figure 5(B)).

3.6. Immune cell infiltration analysis between COVID-19 and control groups

We compared the infiltrating abundance of immune cells between COVID-19 patients and controls using a box plot. In summary, the infiltrating abundance of memory B cells, naïve B cells, resting dendritic cells, M0 macrophages, monocytes, neutrophils, resting NK cells, plasma cells, resting CD4 memory T cells, Naive CD4T cells, CD8 T cells and gamma delta T cells were significantly different (Figure 6(A)). The correlation between 12 differential immune infiltrating cells and five biomarkers showed that plasma cells were strongly correlated with CCNA2, CCNB1 and HSPA5, similar to resting CD4 memory T cells and CCNB1 (Figure 6(B)).

3.7. Cellular expression pattern of biomarkers in GSE171668

The cell type identification results of the heart, lung, liver and kidney tissues are shown in Figure 7(A). We

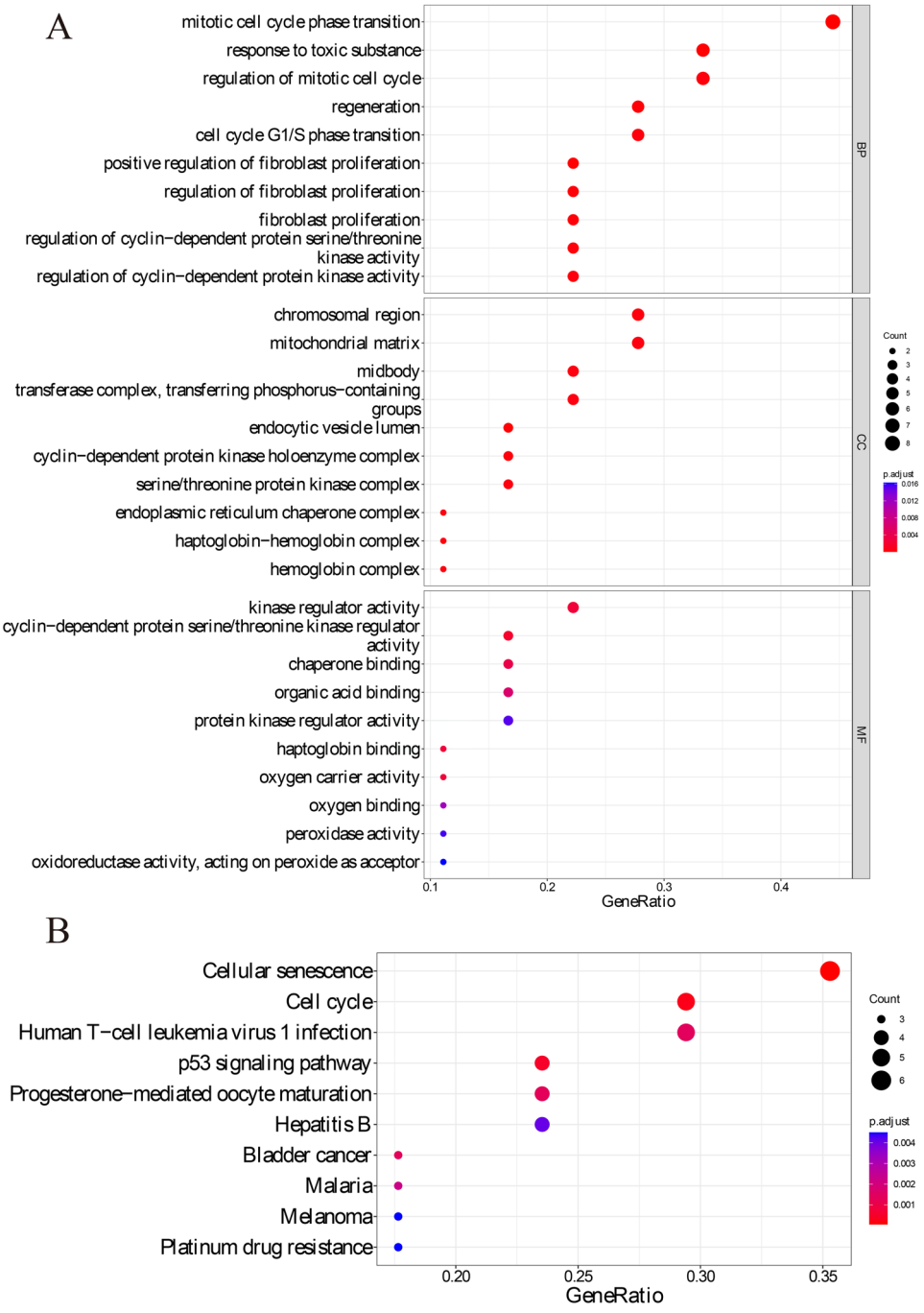


Figure 3. Enrichment analysis of DE-BMRGs. The GO (A) and KEGG (B) enrichment analysis for these DE-BMRGs.

Table 1. Performance metrics for the six models.

Classifier	Precision	Recall	F1	Sensitivity	Specificity	AUROC	Accuracy	Accuracy: 5-fold CV	Accuracy: LOOCV	
Logistic regression	Logistic regression	1	0.614	0.761	0.614	0	1	0.685	0.982	0.981
Random forest	Random forest	1	0.614	0.761	0.614	0	0.995	0.685	0.945	0.944
XGBoost	XGBoost	1	1	1	1	0	1	1	1	0.976
AdaBoost	AdaBoost	1	0.795	0.886	0.795	0	0.999	0.833	0.925	0.976
Naive Bayes classifier	Naive Bayes classifier	1	1	1	1	0	1	1	0.982	0.981
Support vector machine	Support vector machine	1	1	1	1	0	1	1	0.964	0.963

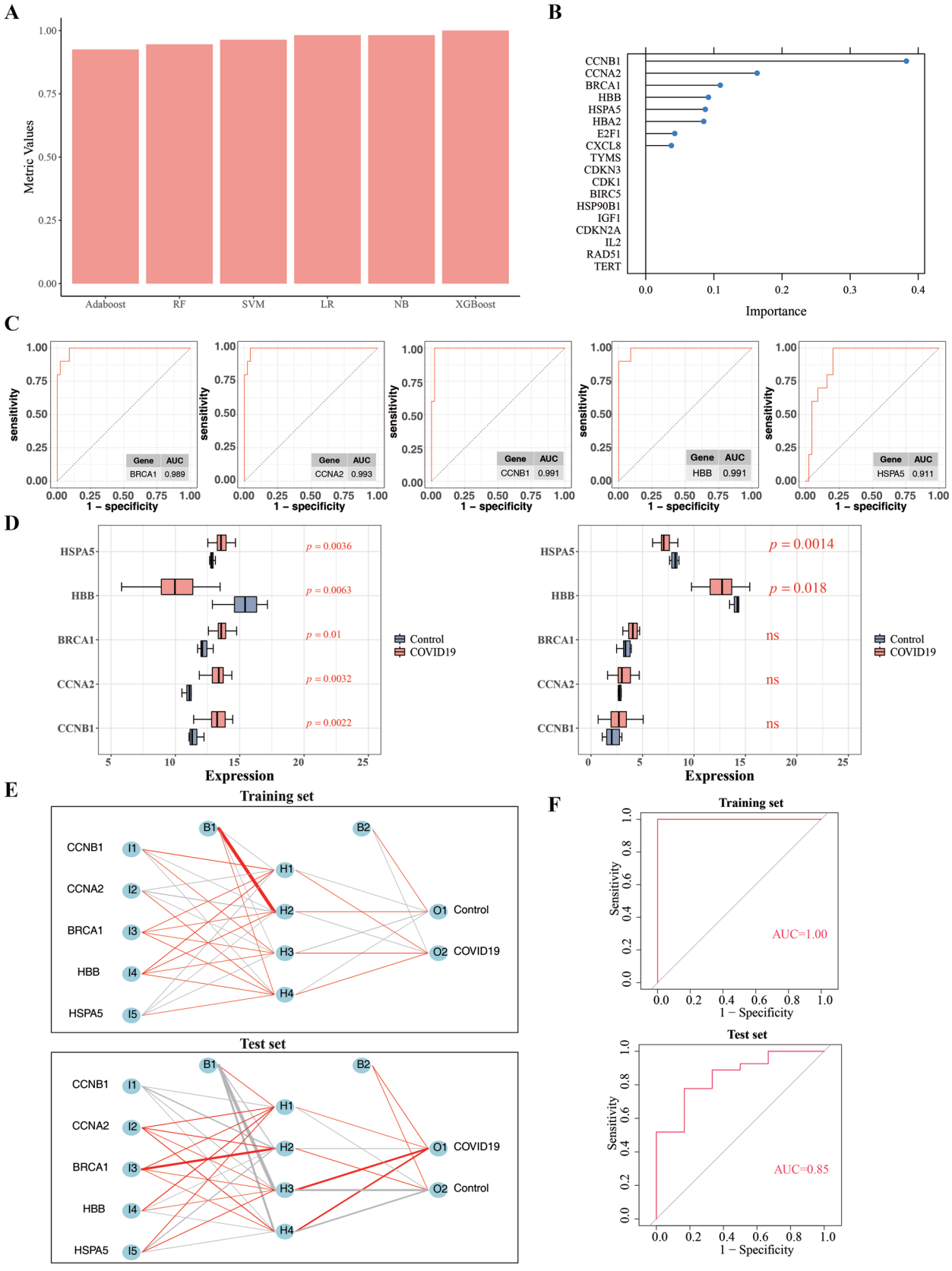


Figure 4. Biomarkers screening and evaluation. (A) Model metric value (accuracy), abscissa: machine learning name; ordinate: model metric value. (B) Dot plot of gene importance. Abscissa: gene importance; ordinate: gene symbol. (C) ROC curve for five hub biomarkers. (D) The expression level of five hub biomarkers in training dataset and test dataset. (E) Artificial neural network (ANN) analysis for five hub biomarkers in training dataset and test dataset. I1–I5 (nodes in the input layer): components of the input vector, here were the five biomarkers; H1–H4 (nodes in the hidden layer): weight; O1–O2 (nodes in the output layer): result to determine whether the sample belongs to the control or treated group; B1: bias from input layer to hidden layer, B2: bias from hidden layer to output layer (bias was an important parameter that served to help adjust the output of the model to better fit the data). Line: weight value, red was positive, grey was negative, thicker line was more weight. (F) ROC curve for ANN analysis.

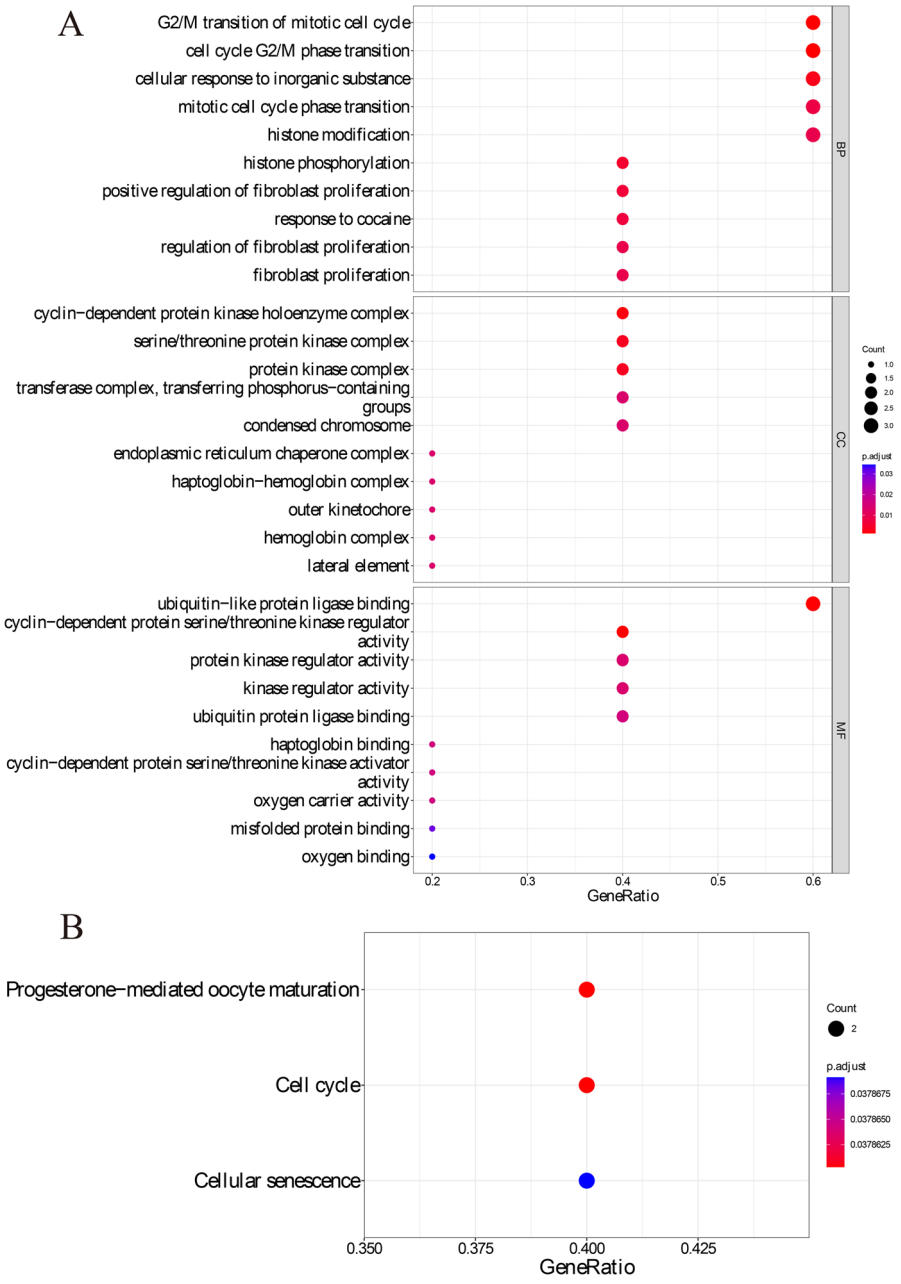


Figure 5. Enrichment analysis of five COVID-19 biomarkers. The GO (A) and KEGG (B) enrichment analysis for these five biomarkers.

acquired five types of cells in the heart: T_cells, tissue_stem_cells, monocytes, endothelial cells and cardiomyocytes. Eight cell types were annotated in the lung: monocytes, B cells, smooth muscle cells, macrophages and T cells. Endothelial cells, erythroblasts, T cells, macrophages and hepatocytes were authenticated in the liver. The largest number of cell types was identified in the kidney tissue, with a total of 10 types, consisting of macrophages, B cells, fibroblasts and descending thin limb cells. It is noteworthy that there were four different types of immune cells, namely B cells, T cells, monocytes and macrophages, and their abundances among different organs were compared and are shown in Figure 7(B). For instance, T cells were the most abundant in the lung tissue, which was significantly higher than that in the heart. The proportion of macrophages in the liver and lung was higher than that in the kidney.

In quick succession, we assessed the expression levels of the five biomarkers in different cell types and then mapped them to the UMAP results. In the heart, the expression of BRCA1, HBB and HSPA5 was relatively higher in cardiomyocytes and endothelial cells, whereas the expression of CCNB1 and CCNA2 was low in all five cells (Figure 7(C,G)). In the lungs, BRCA1, HBB and HSPA5 were expressed at relatively higher levels

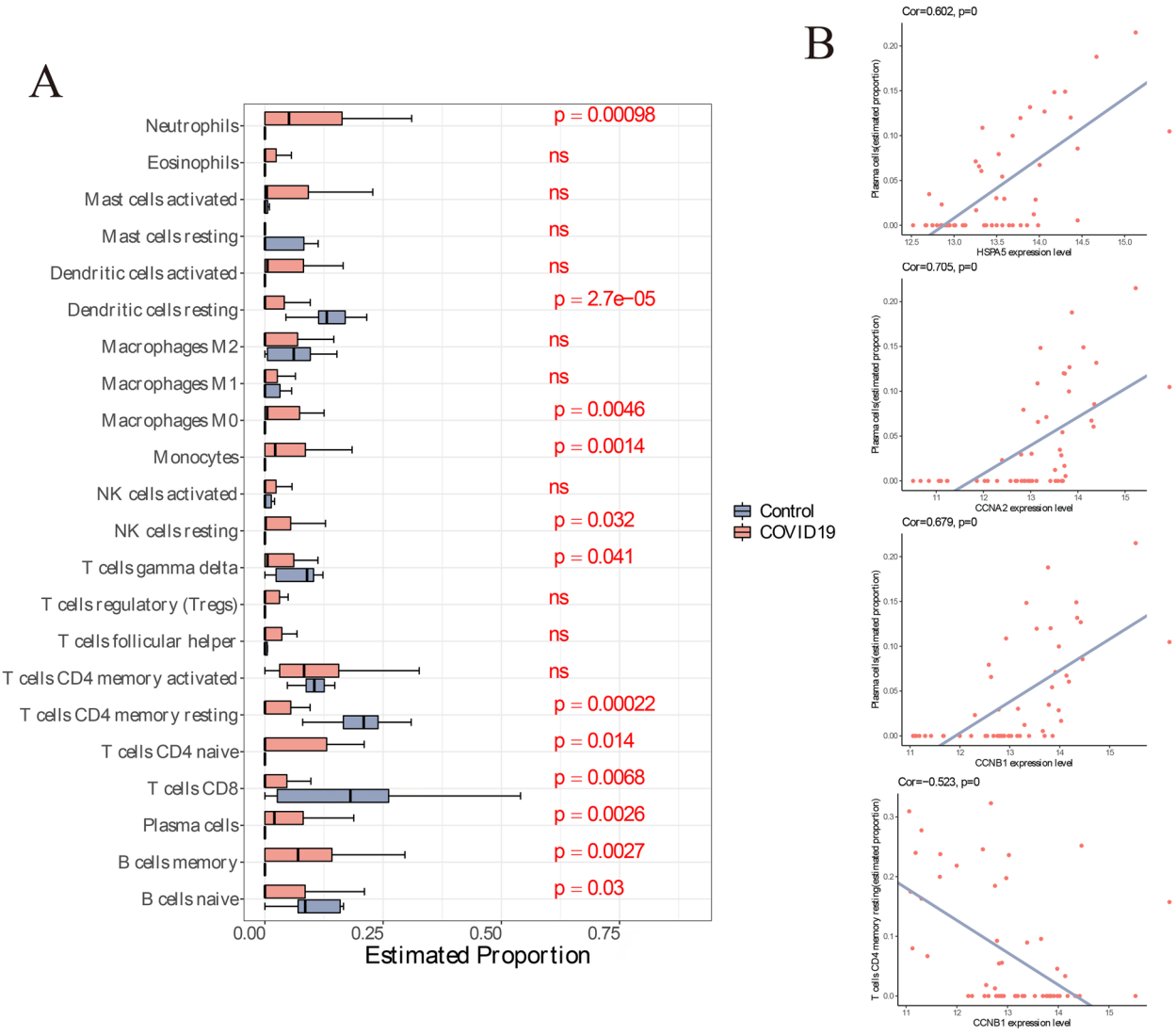


Figure 6. Immune cell infiltration in COVID-19 patients. (A) The expression of multiple immune cells between COVID-19 patients and normal people by differential analysis of immune infiltration. (B) The correlation between biomarker and immune cell infiltration. Abscissa: gene expression; ordinate: abundance value of immune cell infiltration. Point: sample. The blue line was the trend line.

than in monocytes, whereas the expression of CCNB1 and CCNA2 was relatively low in all eight cells (Figure 7(D,H)). In the liver, the expression of HSPA5 and BRCA1 in the five cell lines was the highest, followed by HBB (Figure 7(E,I)). In the kidney, the expression of HSPA5 in 10 cells was the highest, followed by that of BRCA1 and HBB (Figure 7(F,J)).

3.8. The level of biomarkers in serum samples from patients with COVID-19

Finally, we confirmed these five biomarkers in the serum samples of COVID-19 patients and healthy individuals by ELISA. The expression of CCNB1, CCNA2, BRCA1, HBB and HSPA5 was significantly increased in COVID-19 patients compared than in normal individuals (Figure 8(A–E)).

4. Discussion

As a serious epidemic in recent years, COVID-19 has placed a heavy burden on people's health worldwide [34]. COVID-19 can cause a series of pathophysiological effects leading to human death, especially immune abnormalities, cytokines and inflammatory storm [35]. Butyrate exerts anticancer and anti-inflammatory effects by regulating the activity of specific immune cells [36].

In our study, we found that the expression of BMRGs was dysregulated in different tissues such as the brain, liver and pancreas. BMRGs were enriched in multiple molecular functions during COVID19 infection, such as mitotic cell cycle phase transition, regulation of mitotic cell cycle, and cell cycle G1/S phase transition. In a previous study, APOA1 directly drove

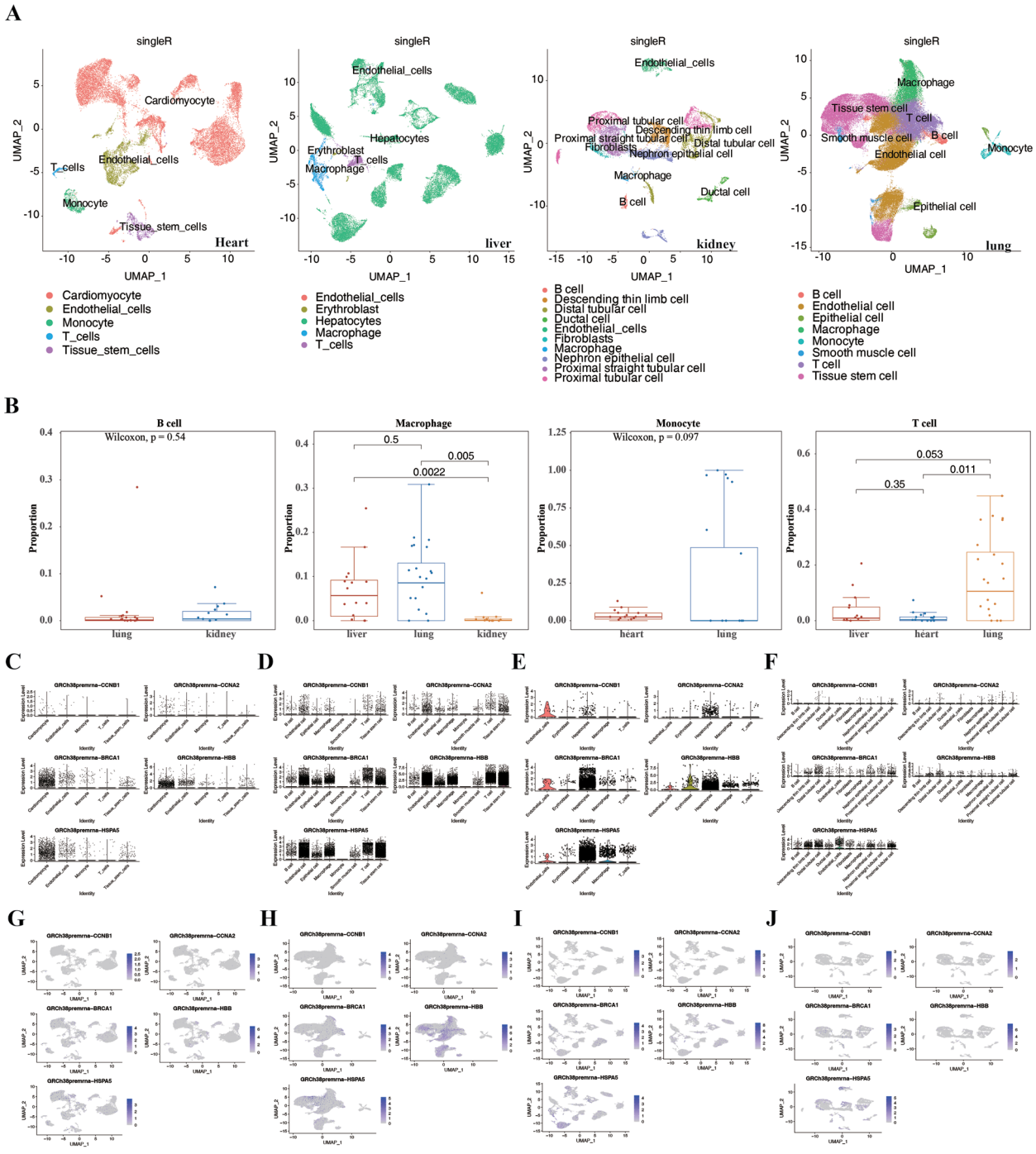


Figure 7. Cellular expression pattern of five biomarkers in COVID-19 patients by single-cell sequencing analysis. (A) UMAP cluster map (annotated) for heart, liver, kidney and lung in COVID-19 patients. Dot: cells, different colours represented different cells. (B) Differential analysis of immune cell components. Five biomarker expression in heart (C), liver (D), kidney (E) and lung (F) of COVID-19 patients. UMAP cluster map (biomarker expression in cells) for five biomarkers in heart (G), liver (H), kidney (I) and lung (J) of COVID-19 patients. The closer the colour to purple, the higher the expression.

cholesterol efflux to maintain the stability of cerebrovascular function [37].

By modulating microglia/macrophage polarization through FPR2/ALX-dependent AMPK-mTOR signalling, ANXA1 prevents cerebral ischemia–reperfusion injury [38]. SHBG attenuates endoplasmic reticulum stress in liver cells by regulating hormone homeostasis [39]. PDK4 is a significant tumour suppressor gene in pancreatic cancer [40]. However, the pathophysiological role of these BMRGs in COVID-19 remains unclear.

Next, we confirmed five biomarkers in COVID-19: CCNB1, CCNA2, BRCA1, HBB and HSPA5. CCNB1 is a regulated mitochondrial gene that drives tumour progression during the cell cycle [41]. It affects cell

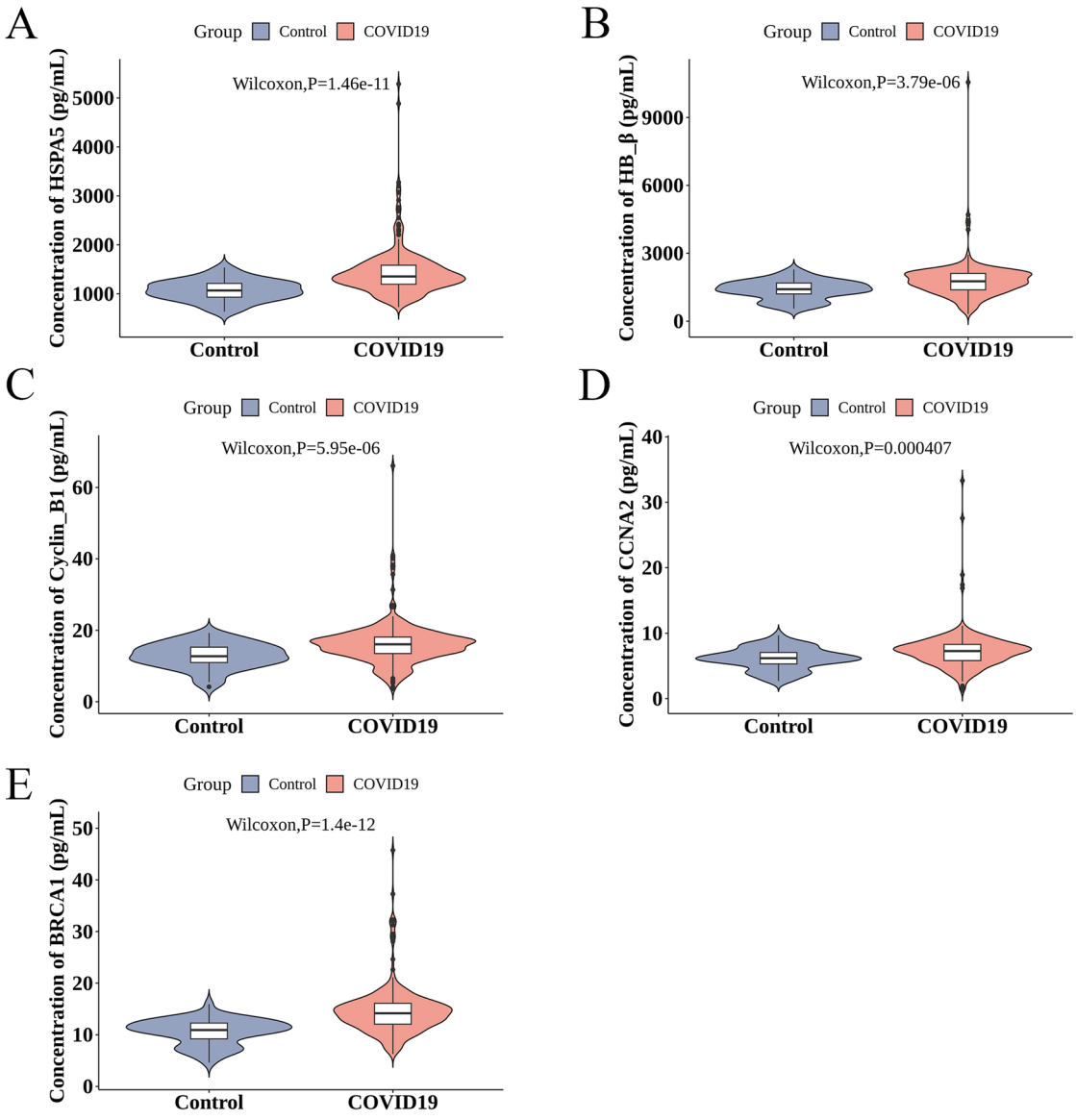


Figure 8. The expression of five biomarkers in serum sample of COVID-19 patients. (A) The expression of multiple HSPA5 (A), HBB (B), CCNB1 (C), CCNA2 (D) and BRCA1 (E) in serum sample of COVID-19 patients by ELISA analysis.

proliferation and division by regulating the entry of cells from the G2 phase into the M phase [42]. In addition, CCNB1 may affect anti-tumour immunity by regulating the distribution of immune cells in the immune microenvironment of liver cancer [43]. In Xu and Gong's study, the risk model associated with CCNB1 (i.e. the high-risk group) showed a higher abundance of CD8 T cells, activated CD4T memory cells, dormant NK cells and macrophages. This suggests that CCNB1 may play a role in the regulation of the tumour immune microenvironment [44]. In addition, CCNB1 may also play a key role in the immune regulation and course of dengue fever [45]. CCNA2 is widely expressed across various human cell types, and its primary molecular biological role involves facilitating cell cycle progression at the G1/S and G2/M transitions through its interaction with CDK2 kinase [46]. Studies have shown that CCNA2 regulates the cell cycle and p53 signalling pathway by interacting with miR-29c-3p, thereby affecting the proliferation, migration, invasion and cell cycle progression of tumour cells [47]. In addition, CCNA2 is closely related to the tumour immune microenvironment, especially to immune cell infiltration and immune checkpoint inhibitory genes (such as PD-1, PD-L1, etc.), which suggests that CCNA2 may play a role in the immune escape of tumours (important role) [48]. BRCA1 functions in the preservation of genome stability and serves as a tumour suppressor. Genetic alterations in this specific gene are responsible for approximately 40% of hereditary breast cancers and over 80% of hereditary breast and ovarian cancers [49]. In addition, BRCA1 regulates cell cycle, tumour

immunity and cell metabolism through various mechanisms [50]. Studies have also shown that BRCA1 plays an important role in transcriptional regulation and can regulate multiple genes related to DNA repair, cell cycle and apoptosis [51]. In addition, Ma et al. found that the loss or functional impairment of BRCA1 activates the cGAS–STING pathway, thereby promoting immune escape and T lymphocyte infiltration in the immune microenvironment of hepatocellular carcinoma [52]. HBB represents one of the two polypeptide chains responsible for encoding haemoglobin subunit beta, and alterations in its gene have the potential to induce sickle cell disease [53]. In addition, HBB is also a key gene related to the cell cycle [52]. Studies have found that increased oxidative stress can lead to a higher apoptosis rate in HBB-KO cells and slow down the cell cycle process and proliferation rate [52]. Studies have also found that HBB plays an important role in the occurrence and progression of the disease by affecting the infiltration of immune cells, especially eosinophils, and participating in inflammatory responses [54]. HSPA5, a member of the HSP70 family, interacts with multiple endoplasmic reticulum proteins to regulate the unfolded protein response [55]. Studies have found that HSPA5/BiP/GRP78 has multiple functions on the cell surface, including regulating cell survival, proliferation, apoptosis, adhesion and innate and adaptive immune responses [56]. In addition, HSPA5 not only plays a role in tumour cells, but also affects the formation of the tumour immune microenvironment by regulating the polarization and immune phenotype of tumour-associated macrophages (TAMs) [57].

In related research on the COVID-19 virus, CCNB1 and CCNA2, as hub genes, might be therapeutic targets for the treatment of HIV patients [58] and digestive system cancer patients infected with COVID-19 [59]. Auwul et al. found that CCNB1 might be a therapeutic target for COVID-19 [60]. Studies have found that CCNB1 plays a key role in the pathogenesis of COVID-19, especially in cell cycle regulation and immune response [61]. In addition, CCNA2 expression in COVID-19 patients is associated with plasma cells, suggesting that it may play a role in immune response, cell proliferation and disease progression [62]. Liu et al. found that BRCA1 was obviously enhanced in COVID-19 patients, which might be closely related to the severity of COVID-19 infection [63]. These findings suggest that BRCA1, CCNB1 and CCNA2 play an important role in the pathogenesis of COVID-19 [64]. Zhang et al. found that HBB is an overlapping protein in COVID-19 progression, which is correlated with complement and coagulation cascades, platelet activation, iron metabolism and anaemia-related pathways [65]. HSPA5 is a potential host cell receptor for COVID-19 [66]. Specifically, HBB plays an important role in COVID-19 infection as part of heme synthesis [64]. HSPA5, a small molecule inhibitor, attenuated the infection of COVID-19 by inhibiting viral protein production [67]. Collectively, butyrate may alleviate the development of COVID-19 through these five key genes. In addition, HSPA5 may be involved in the course of COVID-19 by affecting the mechanism of immune response and viral invasion [56]. In conclusion, butyrate may help alleviate the progression of COVID-19 by regulating these five key genes.

In the current COVID-19 treatment strategy, studies have shown that inhibiting the ORF8 protein of SARS-CoV-2, especially using natural products alone or in combination, can reduce the impact of ORF8 on the immune system, thereby alleviating COVID-19 symptoms [68]. In addition, ACE2-hFcLALA fusion protein, as a decoy receptor, can prevent the virus from entering cells and neutralize different variants, and has the potential to treat and prevent COVID-19 and its variants [69]. Combined with the biomarkers and their functional roles discussed above, they may participate in the treatment of COVID-19 by inhibiting specific signalling pathways or regulating the activity and function of immune cells. Considering the individual differences among COVID-19 patients, detecting the levels of specific biomarkers in patients may provide a basis for formulating more precise personalized treatment plans.

In our study, GO enrichment results showed that five biomarkers were mainly related to the cell cycle and cell proliferation, and KEGG enrichment results showed that five biomarkers were related to the cell cycle and cell senescence. Schmitt et al. found that COVID-19 can promote senescence and exacerbates the senescence-associated secretory phenotype, which can induce 'cytokine storm,' tissue-destructive immune cell infiltration, endothelialitis, fibrosis and microthrombosis [70]. Further studies have shown that the SARS-CoV-2 virus may cause long-term cell damage and dysfunction by disrupting the cell cycle, inducing cell aging and destroying genome stability, and may also have an important impact on long-term COVID symptoms and chronic sequelae [71]. Other studies have found that cells face more environmental stress and accumulation of inorganic substances during aging, and by regulating the cell's response to these substances, it may slow down aging, maintain cell function and improve tolerance [72]. The nucleocapsid protein of COVID-19 has the ability to bind GSDMD, thereby inhibiting pyroptosis while simultaneously inducing cell death through the Smad3-dependent G1 cell cycle arrest mechanism [73]. The interaction between COVID-19 N protein and Smad3

leads to the amplification of TGF-β/Smad3 signalling, resulting in tubular epithelial cell death and the onset of acute kidney injury through the G1 cell cycle arrest mechanism [74]. The study also found that protein kinase D (PKD) plays a key role in cell proliferation, apoptosis and coronavirus replication, and inhibiting PKD and phosphatidylinositol-4 kinase IIIβ (PI4KIIIβ) can effectively inhibit viral replication [75]. In addition, ritonavir has shown therapeutic potential to inhibit COVID-19 virus replication and cancer cell proliferation by interfering with cell cycle and proliferation [76]. Combining our findings with those of previous studies, COVID-19 may induce organ damage by promoting cell senescence and cell cycle disorders.

In our immune infiltration analysis, we found that COVID-19 patients had different degrees of immune cell infiltration compared to healthy people, especially in plasma cells and CD4 memory resting T cells. Bernardes et al. found that severe COVID-19 was distinguished by an increase in proliferative plasma cells that exhibited heightened metabolic activity [77]. In certain instances of severe COVID-19, dysregulated CD4+ T cell signatures may play a role in the increased production of proinflammatory cytokines, thereby contributing to pathogenic inflammation [78]. Mukherjee also hypothesized that CD4 memory resting T cells in patients who have recovered from COVID-19 could be engineered by immunotherapy with selective convalescent blood engineering to stifle COVID-19 [79]. We also found that the five biomarkers were both increased in the serum sample from COVID-19 patients by ELISA. Therefore, these results indicate that COVID-19 may drive these five genes to induce cell cycle disorder and senescence of plasma cells and CD4 memory resting T cells, resulting in immune dysfunction.

Using BMRGs as a starting point, this study investigated the potential function and role of BMRGs in COVID-19, offering a preliminary theoretical foundation and molecular mechanism for the adjuvant treatment of COVID-19 using butyrate, particularly its potential mechanism of action through immune cell interference. Nevertheless, this study still has some limitations. First, due to the use of the public GEO database, the current data set on butyrate metabolism in COVID-19 patients is still very limited, and the data set used is all the data currently available, so the sample size is limited to a certain extent. Second, the findings of this study still need to be verified by more experiments. In order to reduce the impact of insufficient sample size, we use strict data preprocessing and quality control to ensure that the data meets the analysis requirements, and use multiple cross-validation and model parameter adjustment to avoid overfitting as much as possible. Although the sample size limitation may affect the wide applicability of the results, this study still provides preliminary evidence for the potential role of BMRGs in COVID-19. We recognize that expanding the sample size is crucial to verify the performance of the model, and new data sets will be further collected and analysed in the future to verify the research findings and improve the reliability of the results. In addition, we also plan to further study the potential contribution of BMRGs in COVID-19 through molecular biology techniques, and use these five markers to conduct animal or cell experiments to further verify the pharmacological effects of butyrate on COVID-19.

In conclusion, CCNB1, CCNA2, BRCA1, HBB and HSPA5 may be potential biomarkers of butyrate metabolism in COVID-19. These findings provide important directions for further studying the molecular mechanisms of COVID-19. Despite the limitation of small sample size, this study still provides new insights and potential targets for the diagnosis and treatment of COVID-19. Future studies can focus on expanding the sample size, performing functional validation and further exploring the correlation between these biomarkers and other diseases.

Author contributions

WC Zhou, J Zhang, H Li and CS Liu analysed the data. D Liu, XP Chen and J Ouyang used online tools. YK Li, H Li and CS Liu designed the study. T Zeng and S Peng selected the analysed results. YK Li wrote the paper. F Ouyang, YZ Long and YK Li revised the manuscript and designed experiments. All authors have contributed to the manuscript and approved the submitted version.

Ethics statement

The Ethics Committee of Xiangya Hospital Zhuzhou Central South University approved the study (ZZCHEC2022042-01). The privacy and personal identity information of the subjects are protected. All patients involved in this study have signed informed consent and obtained approval from the Ethics Approval Committee. The study was conducted in accordance with the ethical standards of the declaration of Helsinki.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the Natural Science Foundation of HuNan Province (2023JJ41066) and the National Natural Scientific Foundation of China (82303246).

Data availability statement

The data of the article can be obtained from the corresponding author of the manuscript upon reasonable request.

References

[1] Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727–733. doi: 10.1056/NEJMoa2001017.

[2] van Doremalen N, Bushmaker T, Morris DH, et al. Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1. medRxiv. 2020. doi: 10.1101/2020.03.09.20033217.

[3] Wang C, Liu Z, Chen Z, et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol. 2020;92(6):667–674. doi: 10.1002/jmv.25762.

[4] Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. Gene Rep. 2020;19:100682. doi: 10.1016/j.genrep.2020.100682.

[5] Hui DS, Azhar EI, Madani TA, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health – the latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis. 2020;91:264–266. doi: 10.1016/j.ijid.2020.01.009.

[6] Abdar M, Salari S, Qahremani S, et al. UncertaintyFuseNet: robust uncertainty-aware hierarchical feature fusion model with Ensemble Monte Carlo Dropout for COVID-19 detection. Inf Fusion. 2023;90:364–381. doi: 10.1016/j.inffus.2022.09.023.

[7] Lenihan D, Carver J, Porter C, et al. Cardio-oncology care in the era of the coronavirus disease 2019 (COVID-19) pandemic: an International Cardio-Oncology Society (ICOS) statement. CA Cancer J Clin. 2020;70(6):480–504. doi: 10.3322/caac.21635.

[8] Bhat RA, Maqbool S, Rathi A, et al. The effects of the SARS-CoV-2 virus on the cardiovascular system and coagulation state leading to cardiovascular diseases: a narrative review. Inquiry. 2022;59:469580221093442. doi: 10.1177/00469580221093442.

[9] Calligari P, Bobone S, Ricci G, et al. Molecular investigation of SARS-CoV-2 proteins and their interactions with antiviral drugs. Viruses. 2020;12(4):445. doi: 10.3390/v12040445.

[10] Prajapat M, Sarma P, Shekhar N, et al. Drug targets for corona virus: a systematic review. Indian J Pharmacol. 2020;52(1):56–65.

[11] Chen Y, Guo Y, Pan Y, et al. Structure analysis of the receptor binding of 2019-nCoV. Biochem Biophys Res Commun. 2020;525(1):135–140. doi: 10.1016/j.bbrc.2020.02.071.

[12] Zhao S, Lin Q, Ran J, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: a data-driven analysis in the early phase of the outbreak. Int J Infect Dis. 2020;92:214–217. doi: 10.1016/j.ijid.2020.01.050.

[13] LaBouyer M, Holtrop G, Horgan G, et al. Higher total faecal short-chain fatty acid concentrations correlate with increasing proportions of butyrate and decreasing proportions of branched-chain fatty acids across multiple human studies. Gut Microbiome. 2022;3:e2. doi: 10.1017/gmb.2022.1.

[14] Mann ER, Lam YK, Uhlig HH. Short-chain fatty acids: linking diet, the microbiome and immunity. Nat Rev Immunol. 2024;24(8):577–595. doi: 10.1038/s41577-024-01014-8.

[15] Kullberg RFJ, Wikki I, Haak BW, et al. Association between butyrate-producing gut bacteria and the risk of infectious disease hospitalisation: results from two observational, population-based microbiome studies. Lancet Microbe. 2024;5(9):100864. doi: 10.1016/S2666-5247(24)00079-X.

[16] Zhong Z, Zhang Y, Zhao X, et al. Butyrate induces higher host transcriptional changes to inhibit porcine epidemic diarrhea virus strain CV777 infection in porcine intestine epithelial cells. Virol J. 2024;21(1):157. doi: 10.1186/s12985-024-02428-5.

[17] Yang Q, Zaongo SD, Zhu L, et al. The potential of *Clostridium butyricum* to preserve gut health, and to mitigate non-AIDS comorbidities in people living with HIV. Probiotics Antimicrob Proteins. 2024;16(4):1465–1482. doi: 10.1007/s12602-024-10227-1.

[18] Xie Q, Li Q, Fang H, et al. Gut-derived short-chain fatty acids and macrophage modulation: exploring therapeutic potentials in pulmonary fungal infections. Clin Rev Allergy Immunol. 2024;66(3):316–327. doi: 10.1007/s12016-024-08999-z.

[19] Majumdar A, Siva Venkatesh IP, Swarup V, et al. Short-chain fatty acids abrogate Japanese encephalitis virus-induced inflammation in microglial cells via miR-200a-3p/ZBTB20/IKβα axis. mBio. 2024;15(7):e0132124. doi: 10.1128/mbio.01321-24.

[20] Lin Q, Zhou J, Yang F, et al. Sodium butyrate impedes the lymphoma caused by Marek's disease virus via regulating the mitochondrial apoptosis pathway. Front Vet Sci. 2024;11:1360878. doi: 10.3389/fvets.2024.1360878.

[21] Zhang Z, Huang J, Li C, et al. The gut microbiota contributes to the infection of bovine viral diarrhea virus in mice. J Virol. 2024;98(2):e0203523. doi: 10.1128/jvi.02035-23.

[22] Trompette A, Gollwitzer ES, Pattaroni C, et al. Dietary fiber confers protection against flu by shaping Ly6c(–) patrolling monocyte hematopoiesis and CD8(+) T cell metabolism. Immunity. 2018;48(5):992–1005.e8. doi: 10.1016/j.immuni.2018.04.022.

[23] Vieira RS, Castoldi A, Basso PJ, et al. Butyrate attenuates lung inflammation by negatively modulating Th9 cells. Front Immunol. 2019;10:67. doi: 10.3389/fimmu.2019.00067.

[24] Lévy Y, Wiedemann A, Hejblum BP, et al. CD177, a specific marker of neutrophil activation, is associated with coronavirus disease 2019 severity and death. iScience. 2021;24(7):102711. doi: 10.1016/j.isci.2021.102711.

[25] Bost P, De Sanctis F, Canè S, et al. Deciphering the state of immune silence in fatal COVID-19 patients. Nat Commun. 2021;12(1):1428. doi: 10.1038/s41467-021-21702-6.

[26] Jacob F, Pather SR, Huang WK, et al. Human pluripotent stem cell-derived neural cells and brain organoids reveal SARS-CoV-2 neurotropism predominates in choroid plexus epithelium. Cell Stem Cell. 2020;27(6):937–950.e9. doi: 10.1016/j.stem.2020.09.016.

[27] Delorey TM, Ziegler CGK, Heimberg G, et al. COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular

targets. Nature. 2021;595(7865):107–113. doi: 10.1038/s41586-021-03570-8.
[28] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci. 1997;55(1):119–139. doi: 10.1006/jcss.1997.1504.
[29] Zhang Z, Li N, Qian Y, et al. Establishment of an MRI-based radiomics model for distinguishing between intramedullary spinal cord tumor and tumefactive demyelinating lesion. BMC Med Imaging. 2024;24(1):317. doi: 10.1186/s12880-024-01499-8.
[30] Ma Y, Zhan Z, Chen Y, et al. Machine learning-assisted construction of COPD Self-Evaluation Questionnaire (COPD-EQ): a national multicentre study in China. J Glob Health. 2025;15:04052. doi: 10.7189/jogh.15.04052.
[31] Khan S, Abbasi RA, Sindhu MA, et al. Predicting the victims of hate speech on microblogging platforms. Heliyon. 2024;10(23):e40611. doi: 10.1016/j.heliyon.2024.e40611.
[32] Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. doi: 10.1023/A:1010933404324.
[33] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA. Association for Computing Machinery, 2016 Aug 13; Knowledge Discovery and Data Mining (KDD '16); 2016. p. 785–794.
[34] Kevadiya BD, Machhi J, Herskovitz J, et al. Diagnostics for SARS-CoV-2 infections. Nat Mater. 2021;20(5):593–605. doi: 10.1038/s41563-020-00906-z.
[35] Attiq A, Yao LJ, Afzal S, et al. The triumvirate of NF-κB, inflammation and cytokine storm in COVID-19. Int Immunopharmacol. 2021;101(Pt B):108255. doi: 10.1016/j.intimp.2021.108255.
[36] Stoeva MK, Garcia-So J, Justice N, et al. Butyrate-producing human gut symbiont, *Clostridium butyricum*, and its role in health and disease. Gut Microbes. 2021;13(1):1–28. doi: 10.1080/19490976.2021.1907272.
[37] Bahrami A, Barreto GE, Lombardi G, et al. Emerging roles for high-density lipoproteins in neurodegenerative disorders. Biofactors. 2019;45(5):725–739. doi: 10.1002/biof.1541.
[38] Xu X, Gao W, Li L, et al. Annexin A1 protects against cerebral ischemia–reperfusion injury by modulating microglia/macrophage polarization via FPR2/ALX-dependent AMPK-mTOR pathway. J Neuroinflammation. 2021;18(1):119. doi: 10.1186/s12974-021-02174-3.
[39] Kornicka-Garbowska K, Bourebaba L, Röcken M, et al. Sex hormone binding globulin (SHBG) mitigates ER stress in hepatocytes in vitro and ex vivo. Cells. 2021;10(4):755. doi: 10.3390/cells10040755.
[40] Yamashita M, Kumazoe M, Onda H, et al. PPAR/PDK4 pathway is involved in the anticancer effects of cGMP in pancreatic cancer. Biochem Biophys Res Commun. 2023;672:154–160. doi: 10.1016/j.bbrc.2023.06.043.
[41] Xie B, Wang S, Jiang N, et al. Cyclin B1/CDK1-regulated mitochondrial bioenergetics in cell cycle progression and tumor resistance. Cancer Lett. 2019;443:56–66. doi: 10.1016/j.canlet.2018.11.019.
[42] Chen H, Chai W, Li B, et al. Effects of β-catenin on differentially expressed genes in multiple myeloma. J Huazhong Univ Sci Technol Med Sci. 2015;35(4):546–552. doi: 10.1007/s11596-015-1468-4.
[43] Si T, Huang Z, Jiang Y, et al. Expression levels of three key genes CCNB1, CDC20, and CENPF in HCC are associated with antitumor immunity. Front Oncol. 2021;11:738841. doi: 10.3389/fonc.2021.738841.
[44] Xu M, Gong J. Prognostic signature, immune features, and therapeutic responses of a novel ubiquitination-related gene signature in lung adenocarcinoma. J Oncol. 2022;2022:2524649. doi: 10.1155/2022/2524649.
[45] Josyula JVN, Talari P, Pillai AKB, et al. Analysis of gene expression profile for identification of novel gene signatures during dengue infection. Infect Med. 2023;2(1):19–30. doi: 10.1016/j.imj.2023.02.002.
[46] Chotiner JY, Wolgemuth DJ, Wang PJ. Functions of cyclins and CDKs in mammalian gametogenesis. Biol Reprod. 2019;101(3):591–601. doi: 10.1093/biolre/ioz070.
[47] Wang H, Fu L, Wei D, et al. MiR-29c-3p suppresses the migration, invasion and cell cycle in esophageal carcinoma via CCNA2/p53 axis. Front Bioeng Biotechnol. 2020;8:75. doi: 10.3389/fbioe.2020.00075.
[48] Jiang A, Zhou Y, Gong W, et al. CCNA2 as an immunological biomarker encompassing tumor microenvironment and therapeutic response in multiple cancer types. Oxid Med Cell Longev. 2022;2022:5910575. doi: 10.1155/2022/5910575.
[49] Paul A, Paul S. The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers. Front Biosci. 2014;19(4):605–618. doi: 10.2741/4230.
[50] Zheng G, Shi J, Li Q, et al. BAP1 inactivation promotes lactate production by leveraging the subcellular localization of LDHA in melanoma. Cell Death Discov. 2024;10(1):483. doi: 10.1038/s41420-024-02250-6.
[51] Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci. 2004;95(11):866–871. doi: 10.1111/j.1349-7006.2004.tb02195.x.
[52] Ma H, Kang Z, Foo TK, et al. Disrupted BRCA1–PALB2 interaction induces tumor immunosuppression and T-lymphocyte infiltration in HCC through cGAS–STING pathway. Hepatology. 2023;77(1):33–47. doi: 10.1002/hep.32335.
[53] Kato GJ, Piel FB, Reid CD, et al. Sickle cell disease. Nat Rev Dis Primers. 2018;4(1):18010. doi: 10.1038/nrdp.2018.10.
[54] He B, Zhan Y, Cai C, et al. Common molecular mechanism and immune infiltration patterns of thoracic and abdominal aortic aneurysms. Front Immunol. 2022;13:1030976. doi: 10.3389/fimmu.2022.1030976.
[55] Ibrahim IM, Abdelmalek DH, Elfiky AA. GRP78: a cell's response to stress. Life Sci. 2019;226:156–163. doi: 10.1016/j.lfs.2019.04.022.
[56] Li T, Fu J, Cheng J, et al. New progresses on cell surface protein HSPA5/BiP/GRP78 in cancers and COVID-19. Front Immunol. 2023;14:1166680. doi: 10.3389/fimmu.2023.1166680.
[57] Yang H, Xue Y, Jiang Q, et al. HSPA5-mediated glioma hypoxia tolerance promotes M2 macrophage polarization under hypoxic microenvironment. Int Immunopharmacol. 2025;147:113856. doi: 10.1016/j.intimp.2024.113856.
[58] Yan C, Niu Y, Wang X. Blood transcriptome analysis revealed the crosstalk between COVID-19 and HIV. Front

Immunol. 2022;13:1008653. doi: 10.3389/fimmu.2022.1008653.

[59] Xiong Z, Yang Y, Li W, et al. Exploring key biomarkers and common pathogenesis of seven digestive system cancers and their correlation with COVID-19. Curr Issues Mol Biol. 2023;45(7):5515–5533. doi: 10.3390/cimb45070349.

[60] Auwul MR, Rahman MR, Gov E, et al. Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19. Brief Bioinform. 2021;22(5):bbab120. doi: 10.1093/bib/bbab120.

[61] Jin Q, Li W, Yu W, et al. Analysis and identification of potential type II helper T cell (Th2)-related key genes and therapeutic agents for COVID-19. Comput Biol Med. 2022;150:106134. doi: 10.1016/j.compbiomed.2022.106134.

[62] Xu Y, Zhang M, Wang G, et al. Identification of six genes associated with COVID-19-related circadian rhythm dysfunction by integrated bioinformatic analysis. Funct Integr Genomics. 2023;23(3):282. doi: 10.1007/s10142-023-01198-7.

[63] Liu S, Long J, Liang T, et al. Bioinformatics analysis based on high-throughput sequencing data to identify hub genes related to different clinical types of COVID-19. Funct Integr Genomics. 2023;23(1):71. doi: 10.1007/s10142-023-00998-1.

[64] Zhao Z, Zhou C, Zhang M, et al. Analysis of the potential relationship between COVID-19 and Behcet's disease using transcriptome data. Medicine. 2023;102(20):e33821. doi: 10.1097/MD.0000000000033821.

[65] Zhang Z, Lin F, Liu F, et al. Proteomic profiling reveals a distinctive molecular signature for critically ill COVID-19 patients compared with asthma and chronic obstructive pulmonary disease. Int J Infect Dis. 2022;116:258–267. doi: 10.1016/j.ijid.2022.01.008.

[66] Elfiky AA. SARS-CoV-2 spike-heat shock protein A5 (GRP78) recognition may be related to the immersed human coronaviruses. Front Pharmacol. 2020;11:577467. doi: 10.3389/fphar.2020.577467.

[67] Ha DP, Shin WJ, Hernandez JC, et al. GRP78 inhibitor YUM70 suppresses SARS-CoV-2 viral entry, spike protein production and ameliorates lung damage. Viruses. 2023;15(5):1118. doi: 10.3390/v15051118.

[68] Bagheri-Far M, Assadizadeh M, Azimzadeh-Irani M, et al. Non-spike protein inhibition of SARS-CoV-2 by natural products through the key mediator protein ORF8. Mol Biol Res Commun. 2025;14(1):73–91. doi: 10.22099/mbrc.2024.50245.2001.

[69] Bermúdez-Abreut E, Fundora-Barrios T, Hernández Fernández DR, et al. Antiviral activity of an ACE2-Fc fusion protein against SARS-CoV-2 and its variants. PLOS One. 2025;20(1):e0312402. doi: 10.1371/journal.pone.0312402.

[70] Schmitt CA, Tchkonia T, Niedernhofer LJ, et al. COVID-19 and cellular senescence. Nat Rev Immunol. 2023;23(4):251–263. doi: 10.1038/s41577-022-00785-2.

[71] Grand RJ. SARS-CoV-2 and the DNA damage response. J Gen Virol. 2023;104(11):001918. doi: 10.1099/jgv.0.001918.

[72] Sun K, Zhang Y, Li Y, et al. Biochemical targets and molecular mechanism of matrine against aging. Int J Mol Sci. 2023;24(12):10098. doi: 10.3390/ijms241210098.

[73] Wang W, Chen J, Yu X, et al. Signaling mechanisms of SARS-CoV-2 Nucleocapsid protein in viral infection, cell death and inflammation. Int J Biol Sci. 2022;18(12):4704–4713. doi: 10.7150/ijbs.72663.

[74] Wang W, Chen J, Hu D, et al. SARS-CoV-2 N protein induces acute kidney injury via Smad3-dependent G1 cell cycle arrest mechanism. Adv Sci. 2022;9(3):e2103248. doi: 10.1002/advs.202103248.

[75] Han H, Liu H, Steiner R, et al. Inhibition of protein kinase D and its substrate phosphatidylinositol-4 kinase III beta blocks common human coronavirus replication. Microbiol Spectr. 2024;12(12):e0150124. doi: 10.1128/spectrum.01501-24.

[76] Pereira M, Vale N. Ritonavir's evolving role: a journey from antiretroviral therapy to broader medical applications. Curr Oncol. 2024;31(10):6032–6049. doi: 10.3390/curroncol31100450.

[77] Bernardes JP, Mishra N, Tran F, et al. Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. Immunity. 2020;53(6):1296–1314.e9. doi: 10.1016/j.immuni.2020.11.017.

[78] Moga E, Lynton-Pons E, Domingo P. The robustness of cellular immunity determines the fate of SARS-CoV-2 infection. Front Immunol. 2022;13:904686. doi: 10.3389/fimmu.2022.904686.

[79] Mukherjee AP. Hypothesis: immunotherapy by selective convalescent blood engineering to stifle diseases like COVID-19. Int J Med Sci. 2021;18(15):3389–3394. doi: 10.7150/ijms.46363.