



## OPEN ACCESS

## EDITED BY

Lilianne Ganges,  
Institute of Agrifood Research and  
Technology (IRTA), Spain

## REVIEWED BY

Shetty Ravi Dyavar,  
Adicet Bio, Inc., United States  
Zaki A. Sherif,  
Howard University, United States

## \*CORRESPONDENCE

Zhiyong Zhang

✉ zzyzhang@ustc.edu.cn

RECEIVED 20 May 2024

ACCEPTED 27 August 2024

PUBLISHED 26 September 2024

## CITATION

Kanwal A and Zhang Z (2024) Exploring  
common pathogenic association between  
Epstein Barr virus infection and long-COVID  
by integrating RNA-Seq and molecular  
dynamics simulations.  
*Front. Immunol.* 15:1435170.  
doi: 10.3389/fimmu.2024.1435170

## COPYRIGHT

© 2024 Kanwal and Zhang. This is an open-  
access article distributed under the terms of  
the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Exploring common pathogenic association between Epstein Barr virus infection and long-COVID by integrating RNA-Seq and molecular dynamics simulations

Ayesha Kanwal<sup>1</sup> and Zhiyong Zhang<sup>1,2\*</sup>

<sup>1</sup>MOE Key Laboratory for Cellular Dynamics and Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China, <sup>2</sup>Department of Physics, University of Science and Technology of China, Hefei, Anhui, China

The term "Long-COVID" (LC) is characterized by the aftereffects of COVID-19 infection. Various studies have suggested that Epstein–Barr virus (EBV) reactivation is among the significant reported causes of LC. However, there is a lack of in-depth research that could largely explore the pathogenic mechanism and pinpoint the key genes in the EBV and LC context. This study mainly aimed to predict the potential disease-associated common genes between EBV reactivation and LC condition using next-generation sequencing (NGS) data and reported naturally occurring biomolecules as inhibitors. We applied the bulk RNA-Seq from LC and EBV-infected peripheral blood mononuclear cells (PBMCs), identified the differentially expressed genes (DEGs) and the Protein–Protein interaction (PPI) network using the STRING database, identified hub genes using the cytoscape plugins CytoHubba and MCODE, and performed enrichment analysis using ClueGO. The interaction analysis of a hub gene was performed against naturally occurring bioflavonoid molecules using molecular docking and the molecular dynamics (MD) simulation method. Out of 357 common genes, 22 genes (CCL2, CCL20, CDCA2, CEP55, CHI3L1, CKAP2L, DEPDC1, DIAPH3, DLGAP5, E2F8, FGF1, NEK2, PBK, TOP2A, CCL3, CXCL8, DEPDC1, IL6, RETN, MMP2, LCN2, and OLR1) were classified as hub genes, and the remaining ones were classified as neighboring genes. Enrichment analysis showed the role of hub genes in various pathways such as immune-signaling pathways, including JAK-STAT signaling, interleukin signaling, protein kinase signaling, and toll-like receptor pathways associated with the symptoms reported in the LC condition. ZNF and MYBL TF-family were predicted as abundant TFs controlling hub genes' transcriptional machinery. Furthermore, OLR1 (PDB: 7XMP) showed stable interactions with the five shortlisted refined naturally occurring bioflavonoids, i.e., apigenin, amentoflavone, ilexgenin A, myricetin, and orientin compounds. The total binding energy pattern was observed, with amentoflavone being the top docked molecule (with a binding affinity of  $-8.3$  kcal/mol) with the lowest total binding energy of  $-18.48$  kcal/mol. In conclusion, our research has predicted the hub genes, their molecular

pathways, and the potential inhibitors between EBV and LC potential pathogenic association. The *in vivo* or *in vitro* experimental methods could be utilized to functionally validate our findings, which would be helpful to cure LC or to prevent EBV reactivation.

#### KEYWORDS

RNA-Seq, long-COVID, hub-genes, EBV-reactivation, bioflavonoids, molecular docking, molecular dynamics simulation

## 1 Introduction

After the declaration of the coronavirus disease 2019 (COVID-19) pandemic, the world continues to face its aftereffects (1). There is considerable knowledge available on COVID-19 that focuses on the acute illnesses linked to it (2–4); overall, it is also evident that there are tremendous long-term consequences (5, 6) associated with it. Even after recovering from COVID-19, there has been a population of patients who are continuously experiencing symptoms such as breathlessness, cough, fatigue, and neurological symptoms (7); this condition was termed “post-COVID syndrome”. Symptoms may last from a few weeks to several months, severely impacting everyday life activities. In recent times, knowledge on post-COVID syndrome has increased dramatically, which explored clinical manifestations such as pulmonary, neurological, and thromboembolic complications (5, 8–11).

An alternative term called Long COVID (also referred to as “post-acute sequelae of COVID-19”) has recently evolved, which is characterized by a multisystemic condition ranging from severe to mild symptoms like severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. Several reports show that at least 65 million individuals globally have long COVID (LC) (12, 13).

LC cases have been reported in all ages, and the severity level varies in different age groups; the age range 30–50 years has the highest percentage of diagnosis (14). It is characterized by multiple vital organs malfunction, including cardiovascular, thrombotic, and cerebrovascular disease (15), type 2 diabetes (16), and chronic forms of fatigue including myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and dysautonomia, especially postural orthostatic tachycardia syndrome (POTS) (17–20). The duration of these chronic symptoms can last from months to years, and in chronic conditions. The difference between COVID-19 and LC is based on the duration as well. The duration of ongoing COVID-19 is approximately 4–12 weeks, and if the COVID-19 symptoms last for more than 12 weeks, it is referred to as post-acute sequelae of COVID-19 (PASC) or LC (21, 22).

Several studies have formulated various hypotheses regarding the pathogenesis of LC such as persisting reservoirs of SARS-CoV-2 in tissues (23) after the COVID infection and immune dysregulation; pathogen reactivation, including herpesviruses such as Epstein–Barr virus (EBV) and human herpesvirus 6 (HHV-6), is

another potential cause of LC (24). Other causes include the impact of SARS-CoV-2 on the set of naturally occurring microbiomes within the cells, blood clotting in micro vessels with endothelial dysfunction (25, 26), and dysfunctional signaling in the brainstem.

Multiple studies have reported that during the LC condition, the reactivation of EBV infection has been among the most reported phenomena. This reactivation process has also been associated with the severity of the COVID-19 illness (27, 28). Various longitudinal multi-omics studies have suggested that other risk factors such as type 2 diabetes, SARS-CoV-2 RNAemia, specific autoantibodies, and EBV viremia (29) associated with LC development. EBV infection causes various central nervous system illness conditions such as viral meningitis, encephalitis, sleep disorders, psychosis, and multiple sclerosis (30, 31). Previously, a study from Wuhan University (China) was carried out in which the population sample was co-infected with SARS-CoV-2 and EBV. The risk of severe symptoms was approximately threefold increased compared to the samples infected with SARS-CoV-2 only (32), indicating that EBV infection reactivation might directly contribute to increase in the severity of clinical symptoms.

Furthermore, Gold et al. (33) have reported that general symptoms of LC, such as fatigue, insomnia, headaches, myalgia, and confusion, might be because of the EBV reactivation by SARS-2CoV-2 infection. Another study that used COVID-19 patients ( $n = 309$ ) has demonstrated that the EBV virus is greatly associated with LC symptoms (33). Taken together, the relationship between EBV reactivation and LC symptoms has been potentially established (33, 34).

Most studies have used just one-dimensional data, for example, either the observational data or the clinical/serological data, to prove their co-infection biological hypothesis, which is insufficient to provide a solid explanation for the pathogenic association mechanism in EBV-LC scenario and potentially associated genes (35–37). In the recent past, various scientists have applied next-generation sequencing (NGS) data especially the bulk RNA-Seq to decipher the association between two co-infections (38–41) and proposed the hub genes that could be used as biomarkers.

In this study, we have attempted to identify the common hub-genic profile between LC and EBV infection. We applied next-generation sequencing data, mainly RNA-Seq from the EBV-

infected and LC-positive PBMCs combined with the protein structural bioinformatics data, and protein three-dimensional data using cutting-edge techniques such as molecular docking and dynamic simulations to explore the potential pathogenic correlation between EBV infection and LC at the genomic and protein structural level. We investigated the interaction of small molecules, specifically the bioflavonoids, with the shortlisted key hub genes to assess their potential as a drug candidate in mitigating the effects associated with LC and EBV infection scenarios. The results of this study will help us to understand the potential pathogenic connection and could serve as a tool to open a new therapeutics avenue against LC disease-associated manifestations.

## 2 Materials and methods

### 2.1 Data source

The bulk RNA-Seq from human peripheral blood mononuclear cells (PBMCs) of LC (42) 8<sup>th</sup> month post-SARS-CoV-2 infection, non-LC, and the bulk transcriptome of 4<sup>th</sup> day of post-EBV-infected PBMCs (43). The criteria of data selection were based on the patients with no treatment/wild samples. The high-resolution crystal protein structures of shortlisted drug target gene were downloaded from the RCSB Protein databank, and 3D structures of bioflavonoid small molecules were retrieved from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). The complete details of input data are given in [Supplementary Table S1](#).

### 2.2 RNA-sequence data analysis

Sequence Reads Archives (SRA) files of RNA-Seq data from LC and EBV+ infection samples were converted into fastq raw format using fastq-dump and subjected to quality control analysis. The adapters, linkers, and overrepresented and poor-quality reads were evaluated and removed using TrimGalore (44). Refined raw reads were subjected to mapping against Human Genome 38 assembly (Hg38). STAR align (45) was used for the mapping at default parameters, and post-mapped files were formatted and sorted using SAMtools (46). The percentage of mapped reads are shown in [Supplementary Table S2](#). It is crucial to count the number of mapped reads against the reference genome; we have used featureCounts (47), a part of the subreads platform, to quantify the mapped reads. FeatureCounts uses hg39\_annotation.gtf bamfiles from normal and diseased samples as an input to quantify the reads at the gene level (47).

### 2.3 Identification of differentially expressed genes using DESeq2

After the mapped reads quantification, the next step is identifying the DEGs between conditions such as normal and diseased samples. The DESeq2 R-Package in R-studio package uses the shrinkage estimation for the read's dispersion and fold

changes (FCs) among the sample; besides that, it also estimates the *p*-values and expression mean (48). First, the samples were normalized in a way that only the genes with greater than 10 reads counts were used for the DEGs analysis. We have selected two parameters, i.e., Log2FoldChange > 1 and *p*-values < 0.05, to define the gene expression. The common DEGs (cDEGs) between LC and EBV+ infection samples were extracted using the "Venn" R-package. The final output files were exported in the excel format.

### 2.4 Gene ontology analyses

Significant cDEGs were subjected to identify functional categories such as gene ontology (GO), including biological process (BP), cellular component (CC), and molecular functions (MF), and underlying pathways like Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichments in the web-based platform ShinyGo 0.80 version (49) (<http://bioinformatics.sdstate.edu/go/>) that uses the list of DEGs as an input. To ensure the quality of results, we used false discovery rate (FDR) values > 0.05 as a threshold and the top 20 enrichment terms were visualized.

### 2.5 Protein–protein interaction network analysis

Cells perform various complex functions mediated by the regulatory interactions among several proteins. Therefore, to get an insight into the involvement of common genes in PPIs, we used the STRING database (<https://string-db.org/>). This database collects data from different sources via text-mining from published literature, computational predictions using co-expression information, and conserved genomic landscape of genes (50). We have used official symbols of common genes as an input to predict their PPI network. PPIs with an interaction score > 0.4 were refined and used for downstream steps.

### 2.6 Hub gene identification and transcription factor enrichment analysis

In the genome, some genes are involved in the formation of frequent interactions with other genes; such genes are called hub genes. These genes are essentially involved in the formation of a regulatory network and play a crucial role in gene regulation and biological processes (51). Hub genes were detected using the CytoHubba plug-in followed by conformation through different modules (degree) (52). CytoHubba takes PPI network as an input; therefore, the output of STRING was exported and visualized in cytoscape's plug-in CytoHubba (53). Molecular Complex Detection (MCODE) (degree cutoff = 2, max depth = 100, node score cutoff = 0.2 and K-core = 2) was applied for the clustering analysis (54) followed by the ClueGo to carry out KEGG and GO analysis (55). Since gene expression is mainly controlled by transcription factors (TFs), it is thus worthy to perform a TF analysis for the hub genes to

get an insight into the correlation between hub gene expression and TF enrichment. Therefore, we used the EnrichR (56) platform to identify the TF-lof enrichment and ENCODE TF ChIP-Seq enrichment.

## 2.7 Molecular docking

Molecular docking is an efficient method to identify the interactions between proteins and small molecules. The three-dimensional protein structure's selected hub gene OLR1 (PDB ID: 7XMP) was retrieved from the RCSB protein data bank (<https://www.rcsb.org/>) in pdb format. The attached ions, solvent, and unsaturated molecules were removed using the Chimera visualization tool (57); protein was prepared using Chimera's Dock Prep function, and hydrogen and other missing atoms were added and subjected to molecular docking using the PyRx software (58). The small bioflavonoids molecules library of bioflavonoids was downloaded from PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and converted into the pdb format. The AutoDock suite was applied, the small molecules were imported as a library, their energies were minimized using *uff force field*, and conjugate gradients were used as the optimization algorithm saved in *pdbqt* format. The maximize grid was selected and blind docking was performed. The best docking pose was selected based on the lowest binding energy and root mean square deviation (RMSD) value. The two-dimensional view of interacting residues was visualized using Discovery Studio (59).

## 2.8 Molecular dynamics simulations of receptor–ligand complexes

MD simulations were conducted using AMBER (60) and GROMACS software suites (61) with the Amber force field (62). Initial coordinates and topology files of both receptor and ligands were generated by using Gaff force field 2 and ff99SB force field, respectively. Both ligands and proteins were prepared via the Antechamber and the Leap program of Amber tools. The particle mesh Ewald method was applied to compute the long-range electrostatic interactions while the short-range interactions, such as van der Waals interactions, were calculated with a default cutoff value (1.0 nm). The system was solvated using the OPC water model with a size of 12.0 Å. The GROMACS.top and.gro files were generated for the MD simulations run using the amb2gro\_top\_gro.py tool, which were based on the AMBER prmtop and inpcrd files (60). Energy minimization was performed on the receptor–ligand complexes and unbound proteins using the GROMACS v. 2021.4-plumed-2.8.0 software package. The initial 2,000 steps involved minimizing steepest descents, while the subsequent 2,000 steps were for minimizing the conjugate gradients. The temperature equilibration was performed at 300 K temperatures for 50,000 steps. After the equilibration step, the prepared systems of both apo and ligand-bound complexes were subjected MD simulations for 100 ns at 300 K temperature and 1 atm pressure. Several downstream analyses such as RMSDs, radius of gyration (Rg), root mean square fluctuations

(RMSFs), hydrogen bonding analysis, and energy decomposition were carried out. The Bio3D (63) r-package was used to calculate the principal component analysis (PCA) and the dynamic cross-correlation map (DCCM). The analysis output graphs were plotted using the r-package ggplot2. Molecular mechanics Poisson–Boltzmann surface area (MMPBSA) is the remarkable method used to calculate the binding free energy (in kcal/mol) of a complex between protein and small molecules (64). The determination of the binding free energy denoted as  $\Delta G_{\text{bind}}$  between a ligand and a protein in the formation of a complex in aqueous solution is based on the disparity stated as follows:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{Protein}} + G_{\text{Ligand}}) \quad (1)$$

$$\Delta G_{\text{bind}} = \Delta G_{\text{gas}} + \Delta G_{\text{sol}} + T\Delta S \quad (2)$$

$$\Delta G_{\text{gas}} = \text{Bond} + \text{Angle} + \text{Dihed} + \text{EEL} + \text{VDWAAL} \quad (3)$$

$$\Delta G_{\text{sol}} = \Delta \text{EGB} + \Delta \text{ESURF} \quad (4)$$

## 3 Results

### 3.1 Mapping and analysis of differentially expressed genes

The processed clean raw reads were mapped against the human genome reference; mapping reads percentage statistics are presented in [Supplementary Table S2](#). The mapped output files are saved in bam files format, the low-quality mapped reads were filtered out with the “-bq1” parameter, and the resulting files were sorted and saved again in new bam files. Filtration and sorting were performed using SAMtools (46). In LC data, nearly ~22,500 genes were found to be expressed. In order to differentiate between the up- and downregulated genes, we used the following criteria:  $\log_2\text{FC} > 1$  and  $\log_2\text{FC} < -1$ , and  $p\text{-value} < 0.05$ , respectively. Out of ~22,500 expressed genes, 866 genes were found to be upregulated and 330 genes were found to be downregulated while rest of the genes did not meet the criteria of significance and considered as outliers. In EBV samples, a total of ~19,700 genes were detected as expressed genes, and similar to LC samples, up- and downregulated genes were detected in EBV data with the following criteria:  $\log_2\text{FC} > 1$ ,  $\log_2\text{FC} < -1$ , and  $p\text{-value} < 0.05$ , respectively. A total of 3,075 genes fulfilled the upregulation criteria of DEGs, where 2,806 genes were found to be downregulated. The rest of the genes did not fulfill the criteria specified and were considered as non-significant. The genome-wide upregulated and downregulated genes of both LC and EBV-infected samples are shown in [Figure 1](#). Collectively, we found that in LC and EBV-infected samples, several genes were dysregulated. Detection of gene expression within the samples is important because it enables us to get more explicit details about the expression pattern of each replicate. Therefore, we selected the top genes from the genome-wide mRNA samples and plotted them in an expression heatmap ([Supplementary Figure S1](#)). The selection of genes was based on the  $\log_2\text{FC}$  and  $p$ -values, together. Among the



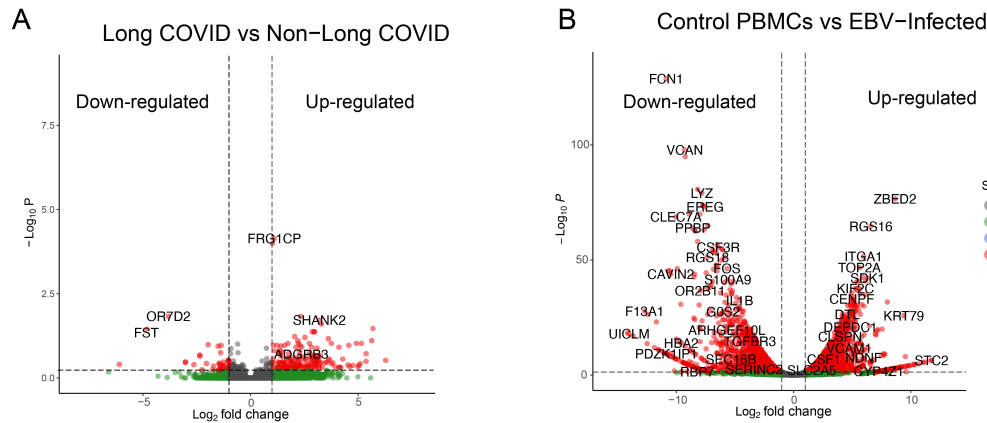


FIGURE 1

Identification of differentially expressed genes (DEGs) in LC and EBV-infected RNA-Seq data. Up and down regulated genes are represented by a single dot. (A) DEGs from 8<sup>th</sup> month post-SARS-CoV-2-infected PBMCs (42). (B) DEGs from 4<sup>th</sup> day post-EBV-infected PBMC samples (43). Log<sub>2</sub>FC values are plotted on the x-axis and  $-\text{Log}_{10} p$ -values are shown on the y-axis. Green dots represent the genes that are greater than  $\text{log}_2\text{FC} > 1$ , gray dots are non-significant genes, red dots show the genes that fulfill the criteria of  $p$ -values  $< 0.05$  and  $\text{log}_2\text{FC}$  collectively, and blue dots represent genes with  $p$ -value  $< 0.05$  only.

top 30 genes, 25 genes were found to be upregulated in non-LC samples while 5 genes [FST (65), OR7D2 (66), SERPINE1 (67), TMEM176A (68), and TMEM176B (69)] were dysregulated in the LC sample cluster. The association of these five genes with LC/COVID-19 has been previously reported by various scientists via bulk RNA-Seq and genomic data analysis. A clinical genomic data study has reported that follistatin (FST), which is an inhibitor of the follicle-stimulating hormone (FSH), has been upregulated (36).

In EBV samples, the top 20 genes were plotted, and 18 genes were found to be downregulated on the 4th day of post-EBV infection samples, while two genes ZBED2 and RGS16 were among the upregulated ones in EBV-infected samples.

### 3.2 Protein–protein interaction analysis of common genes and identification of hub genes

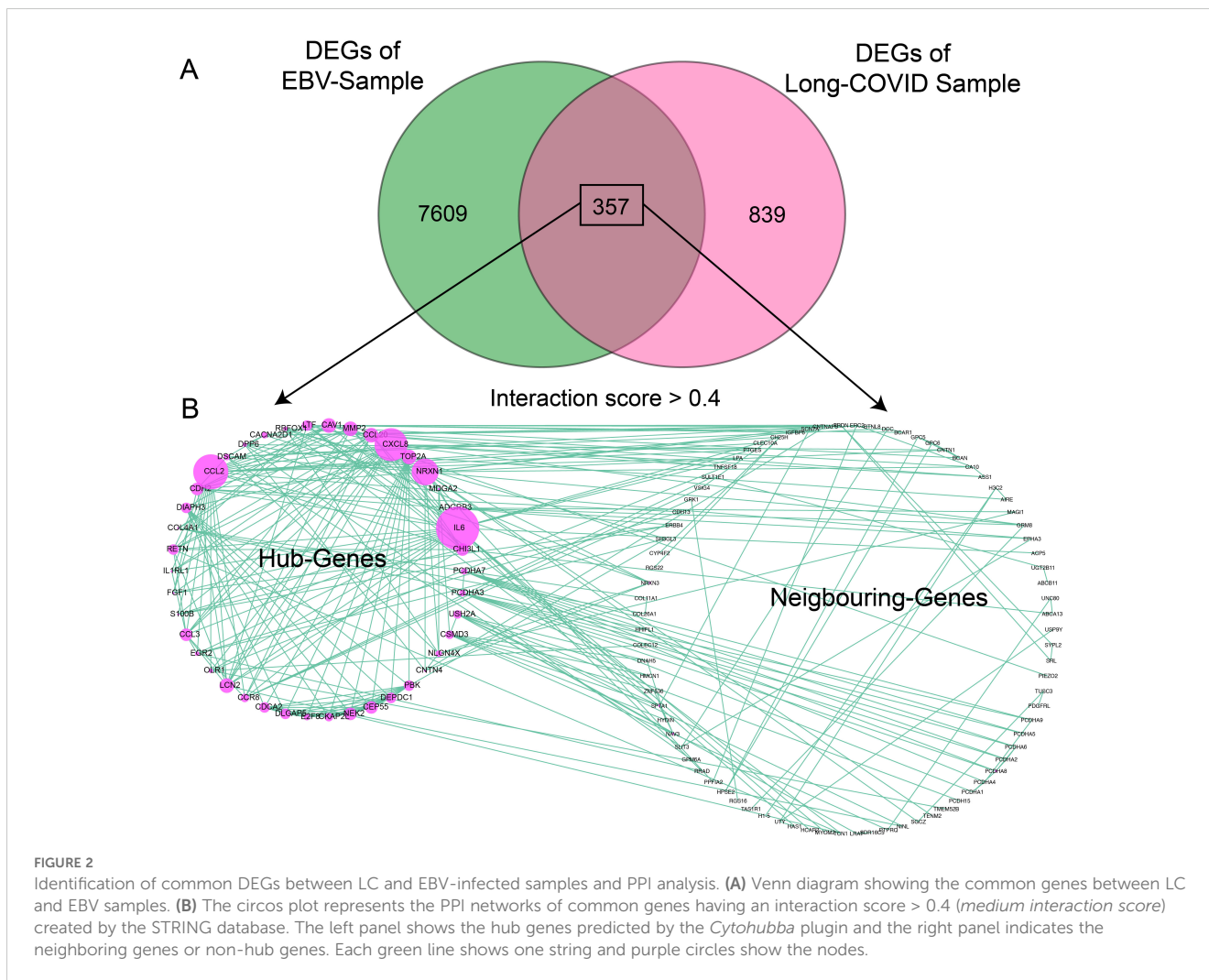
Significant DEGs from both samples (7,609 genes from EBV samples and 839 genes from LC samples) were intersected; 357 common/overlapped DEGs between two samples were extracted and considered as the common signature genes between both diseases. The intersection criteria were based on the gene names and were plotted in a Venn diagram as shown in Figure 2A. The enrichment analysis of the cDEGs showed that only five genes were functionally enriched in glutathione transferase activity with an FDR of 4.1E-03 and 24 genes were enriched in calcium ions binding with an FDR of 1.5E-03. The common genes were subjected to PPI analysis using the STRING website. The network with the interaction score  $> 0.4$  from the STRING server was exported to cytoscape's plugin CytoHubba to identify hub and neighboring genes. A total of 125 genes were involved in the formation of PPIs (Supplementary Figure S2). Out of these 125 genes, 43 were classified as hub genes [with more frequent edges (degree  $> 5$ ) and strings] predicted by CytoHubba using multiple built-in

modules and are shown in Supplementary Table S4, while 82 genes were classified as neighboring genes as shown in Figure 2B. The hub genes were further broken down into clusters via the MCODE plugin. Out of the 43 hub genes, 28 were clustered together in cluster 1, while the rest of the genes were clustered together in cluster 2 as shown in Figure 3. In cluster 1, CCL2, MMP2, IL6, CXCL8, CCL20, NRCN1, and CDH2, had the highest number of edges. In cluster 2, NEK2, TOP2A, CDCA2, CEP55, DEPDC1, DIAPH3, PBK, and DLGAP5 had the highest number of edges. These genes were highlighted with a purple bubble as presented in Figure 3A. Since these genes have an ability to interact frequently, it is worthy to study in detail their role in generating any pathogenic association between LC and EBV reactivation.

We were interested to determine their role in biological pathways. Most of the genes exhibit their biological involvement in the pathways shared by LC and EBV infection, for example, the behavioral maintenance-related pathways. Hub genes were subjected to GO analysis; functional analysis, including biological process, cellular components, and molecular functions, and KEGG pathway analysis were performed as plotted in Figure 3B. It was found that the hub genes were involved in the following GO terms: (1) growth factor receptor binding, (2) maintain external encapsulating structure, (3) protein kinase B signaling, (4) social behavior associated, (5) regulation of chemotaxis, (6) regulation of biological processes involved in symbiotic interactions, and (7) cellular response to fibroblast growth factor stimulus.

### 3.3 Differential expression of hub genes in LC and EBV and analysis of transcription factors

Since hub genes have gained a great functional importance, we have identified their individual gene expression in both LC and EBV samples. We chose the list of genes with frequent edges and



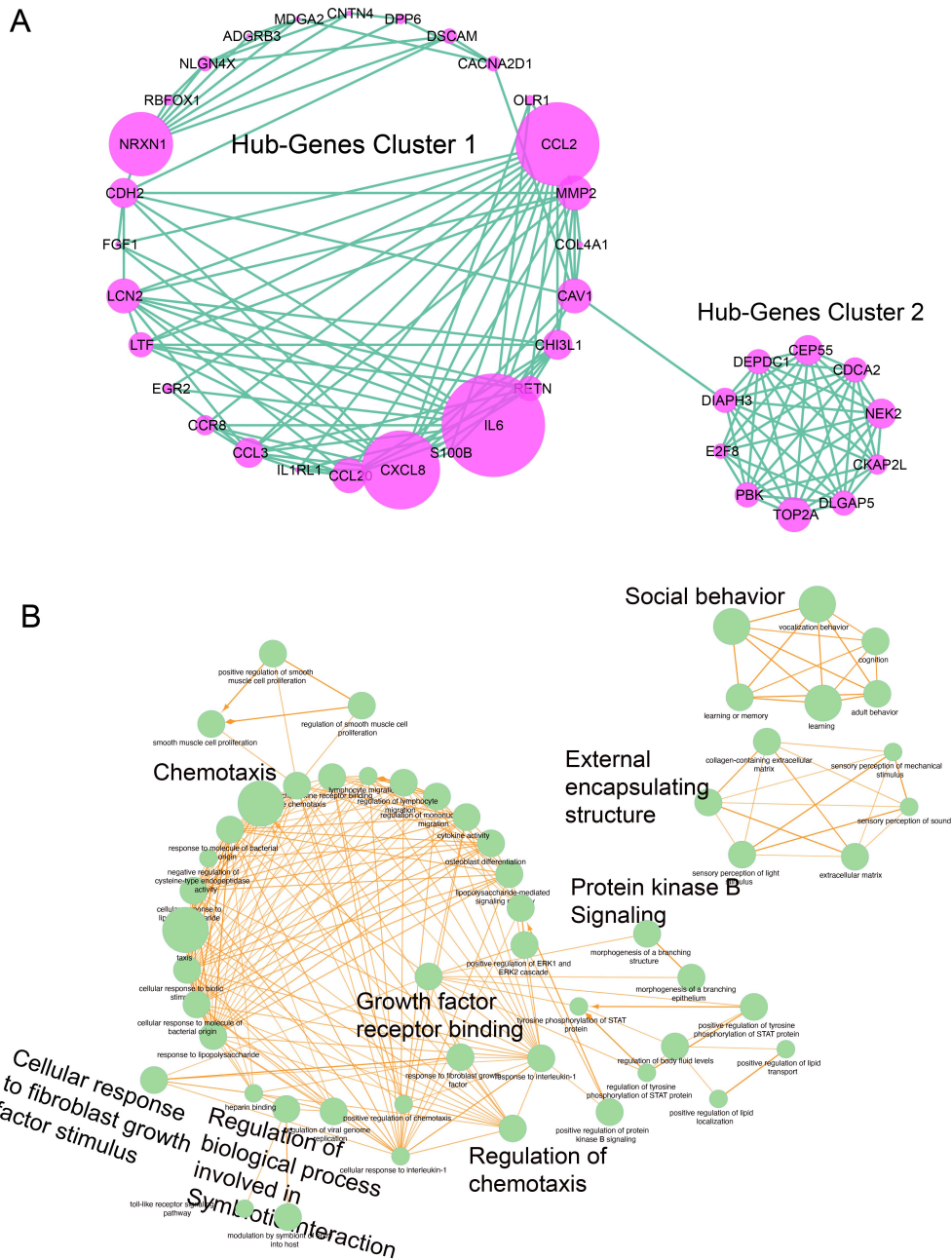
investigated their individual gene expression in both LC and EBV samples. We selected the genes with dense edges (degree > 5) and identified their gene expression in both conditions. We have found that CDCA2, CEP55, CKAP2L, DEPDC1, DIAPH3, DLGAP5, E2F8, FGF1, NEK2, PBK, and TOP2A were upregulated in both LC and EBV infection. Only CCL2, CHI3L1, and EGR2 showed the opposite expression, i.e., higher gene expression in EBV sample and lower gene expression in LC. The expression of MMP2 and RETN was upregulated in LC- and downregulated in EBV+ infection samples.

On the other hand, CCL20, CCL3, CXCL8, DEPDC1, DIAPH3, DLGAP5, E2F8, IL6, LCN2, and OLR1 were depleted in both EBV+ infection samples and LC data. The expression was calculated on the basis of FC value. The range that we observed was from -6 to +6. In EBV infection, CDCA2, CEP55, CKAP2L, DEPDC1, DIAPH3, DLGAP5, FGF1, PBK, TOP2A, CCL3, LCN2, MMP2, OLR1, IL6, and LCN2 were the genes exhibiting a significant change of expression (based on log2FC) as compared to others. However, in LC samples, only the FGF1 gene was observed to produce a higher log2FC as shown in Figure 4A. The gene expression is controlled by the TFs.

Next, we performed the TF enrichment analysis in selected hub genes. It was found that out of 22 genes, only 17 were enriched with multiple TFs and most of the genes shared a common TF. For

example, the TF ARNTL2 controlled multiple genes such as EGR2, CCL2, DLGAP5, CEP55, E2F8, PBK, TOP2A, DEPDC1, CDCA2, NEK2, CKAP2L, DIAPH3, CXCL8, and IL8. Additionally, we have found that TFs belonging to the Zinc-finger protein family and MYB transcription family were among those associated with the 10 hub genes. Previous studies have also shown the association of these family protein with COVID-19's manifestations and EBV infection as shown in Figure 4B. DLGAP5, CEP55, E2F8, PBK, TOP2A, DEPDC1, CDCA2, NEK2, CKAP2L, and DIAPH3 were reported as the genes associated with the MYB and ZNF-TF families. Three members of the ZNF-TF family, i.e., ZNF-215, ZNF-2670, and ZNF-219, were predicted as the main TFs controlling the hub genes. In the MYB family, MYBL1 and MYBL2 were the members identified as the TFs showing the regulatory association with the hub genes. Since hub genes are the key genes that are involved in controlling various biological functions, it is essential to monitor their gene expression. We identify the expression of hub genes and found that their expression changes depending on EBV and LC conditions.

Furthermore, we identify the TFs involved in controlling these hub genes; common TFs are shared by the genes that co-regulate in EBV or LC, suggesting that these genes shared a common transcriptional machinery. For example, the TF ARNTL2 is



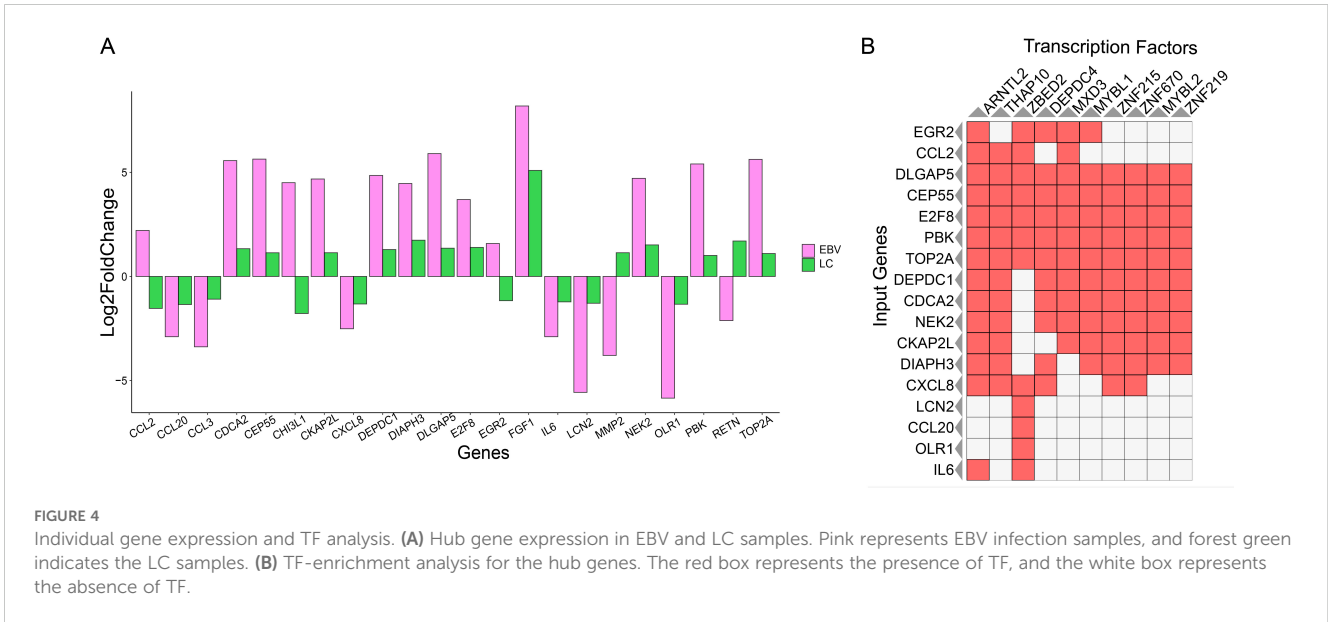
**FIGURE 3** Hub genes clustering based on the number of nodes and enrichment analysis. **(A)** Clustering analysis was performed by applying different modules in MCODE, (degree of connection) was applied, and two clusters were detected. The size of a ball indicates the number of associated nodes (smallest degree = 5); a higher number has a bigger size (highest degree = 25). Each green line show one string and purple circles show the nodes. **(B)** Enrichment analysis of hub genes, in which each green ball represents the hub gene and its involvement in various biological functions; significant terms were assessed using adj. *p*-value<0.05. The size of the node indicates the significance of GO terms. The yellow line shows the nodes interacting with each other.

associated with 14 out of 17 hub genes in the list (Figure 4B), and it was observed that these genes were either upregulated or downregulated together in EBV and LC.

### 3.4 Molecular docking with bioflavonoids

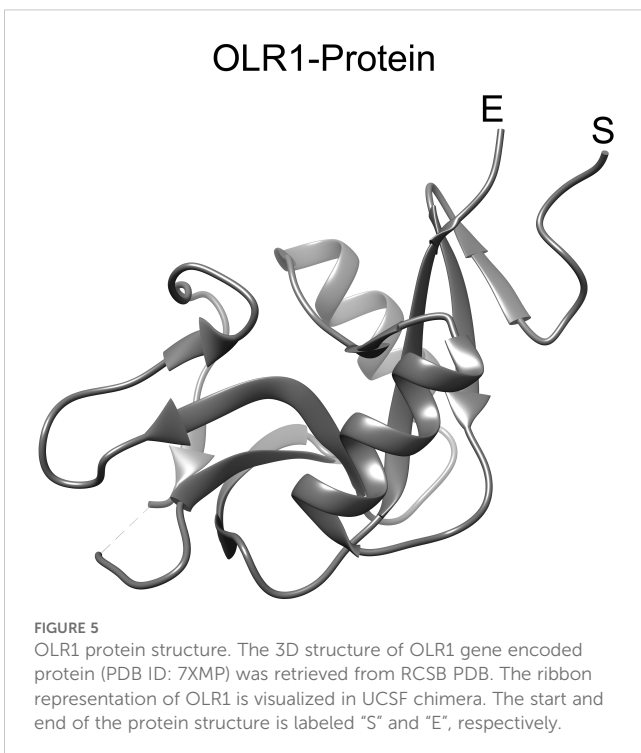
In this study, we have shortlisted the hub genes that have a direct association with SARS-CoV-2 infection-related diseases and EBV

infection. The selection criteria were solely based on the literature survey. We have used Lectin-like Ox-LDL receptor 1 (OLR1) and searched their corresponding proteins (PDB ID: 7XMP) from PDB with 1.27 Å resolution (Figure 5). The protein was docked against biological active compounds mainly in the bioflavonoid library. It was previously reported that bioflavonoids have a strong antiviral and antitoxic ability (70). The top five drug/compounds were used for further analysis (Supplementary Figure S3). In the OLR1–apigenin



complex, the binding energy was  $-7.6$  kcal/mol, and Ser207 and Arg209 mainly formed hydrogen bonds with apigenin (Figure 6A). Amentoflavone docked with OLR1 protein and showed a binding energy of  $-8.3$  kcal/mol. Two amino acids, Phe158 and Phe261, were involved in pi bonds, while Ser159 and Tyr179 formed conventional hydrogen bonds as shown in Figure 6B. In the OLR1–ilexgenin A complex, the binding energy was  $-7.6$  kcal/mol. A total of 13 amino acids were involved in the interaction formation. Out of these 13 bonds, two residues, i.e., Trp217 and Glu218, formed hydrogen bonds as represented in Figure 6C. In the OLR1–myricetin complex, the binding energy was  $-6.5$  kcal/mol and a total of 16 amino acids were involved in the interaction formation. Out of these 16 bonds, three

amino acid residues, i.e., Val230, Leu223, and Pro211, formed hydrogen bonds and the three-dimensional view and bound residues are shown in Figure 6D. In the OLR1–orientin complex, the binding energy between ligand and receptor was  $-7$  kcal/mol. A total of 15 amino acids were involved in the interaction formation. Three residues, i.e., Pro211, Val230 and Leu223, formed hydrogen bonds, and three amino acids, Tyr213, Pro214, and Trp215, formed the pi bonds. Other residues such as Ser212, Arg208, Pro225, Pro222, Leu216, His226, Phe228, and Arg229 were bound with orientin via van der Waals interactions as presented in Figure 6E. Overall, a variety of bonding was observed, such as hydrogen bonding, van der Waals interactions, sulfur interactions, pi bonds, and C–H bonding; a detailed list of bonding is shown in Table 1.

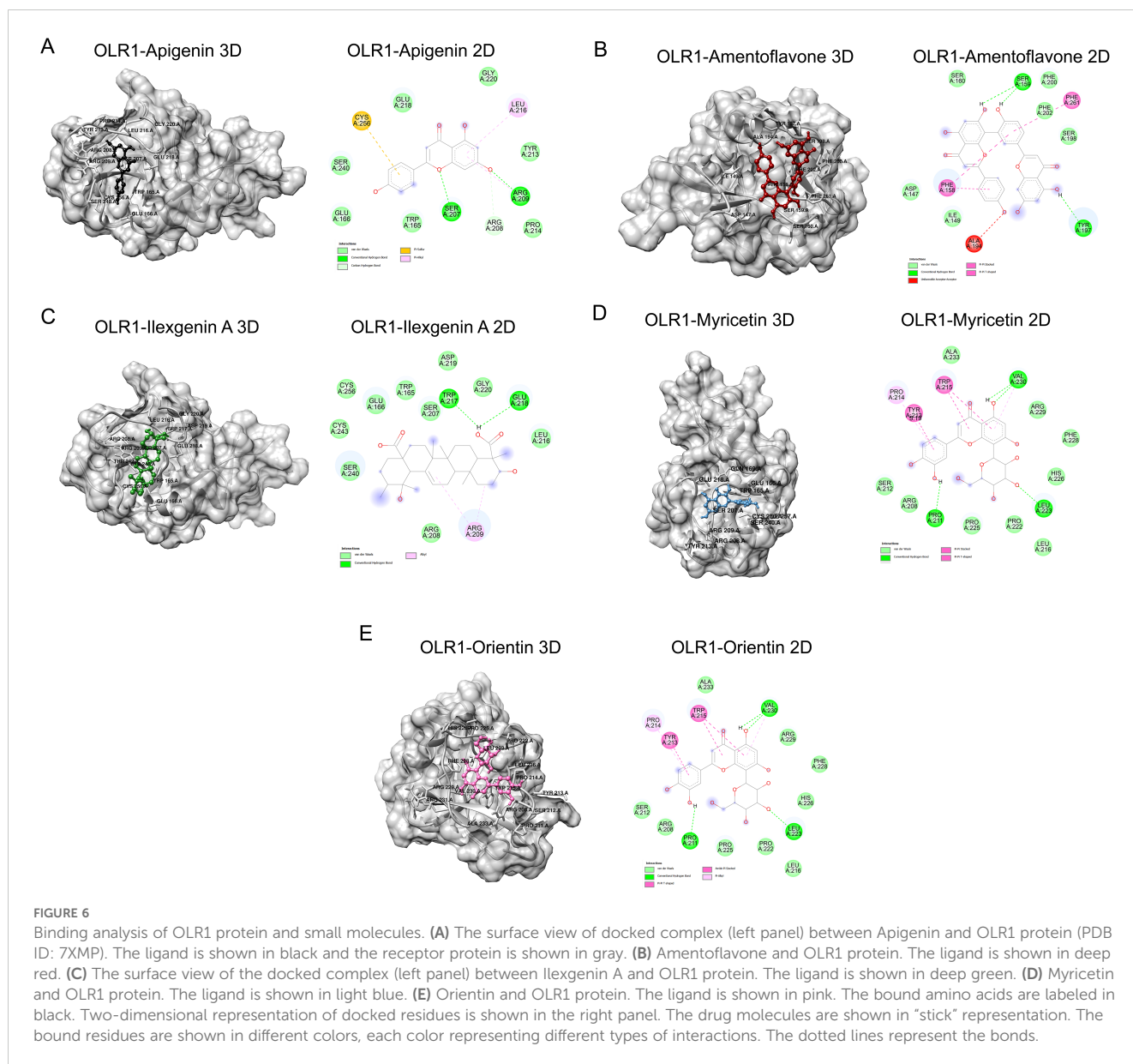


### 3.5 Molecular dynamics simulations analysis of OLR1 protein and complexes

The molecular changes and stability of small-molecule inhibitors–OLR1 complexes were investigated from their corresponding 100-ns MD simulation trajectories. The trajectories obtained from 100-ns simulations for the hub gene protein complexed with apigenin, amentoflavone, ilexgenin A, myricetin, and orientin systems were prepared using the Antechamber of the Amber program (62). The atom-positional RMSD generated by roto-translational least-squares fitting is perhaps the most widely used for structural comparison and stability measure.

We analyzed the structure stability of the backbone of protein in the complexes and in unbound form. The structure remained intact throughout the simulation time of 100 ns (Figure 7). RMSD plots of OLR1 complexed with the five ligands showed the fluctuations ranging between  $\sim 0.1$  and  $0.4$  Å. All the five ligand-bound complexes up to 100 ns remain intact. However, the protein bound with myricetin showed some fluctuation from  $\sim 60$  to 70 ns comparatively. Later, another fluctuation was observed at 75 to 80 ns followed by a stable conformation. After myricetin, orientin

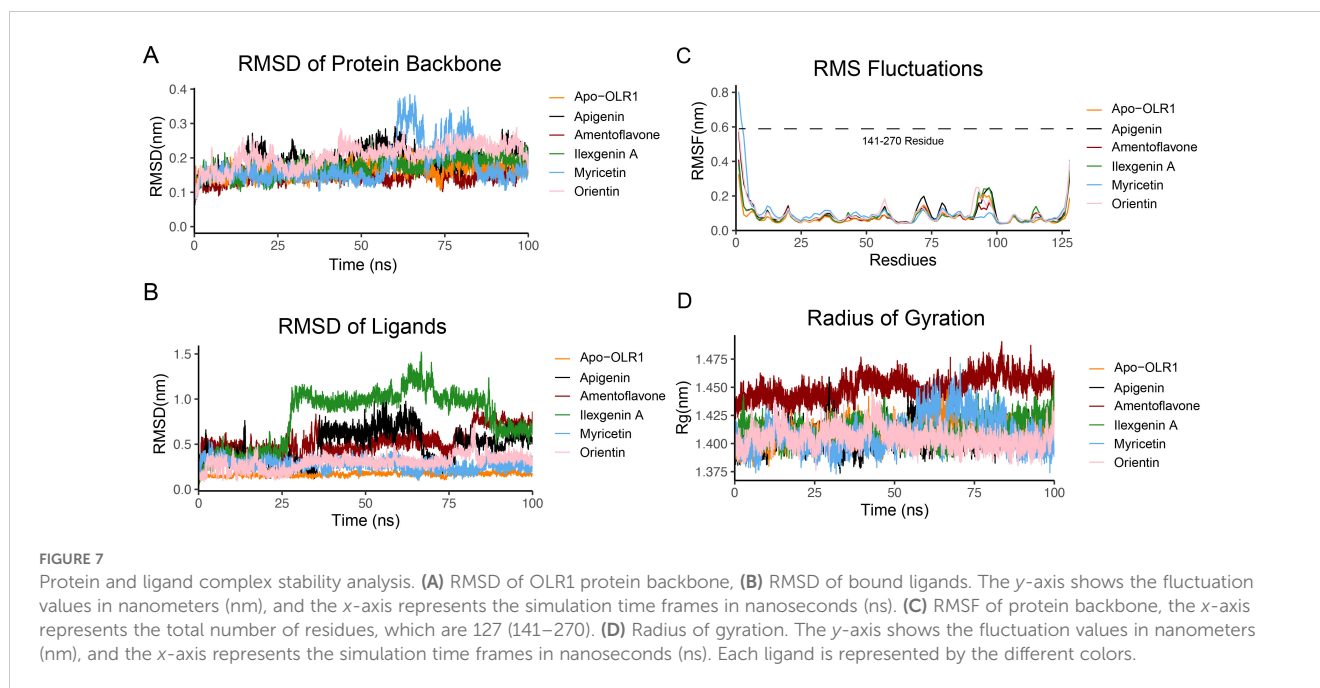




**FIGURE 6** Binding analysis of OLR1 protein and small molecules. **(A)** The surface view of docked complex (left panel) between Apigenin and OLR1 protein (PDB ID: 7XMP). The ligand is shown in black and the receptor protein is shown in gray. **(B)** Amentoflavone and OLR1 protein. The ligand is shown in deep red. **(C)** The surface view of the docked complex (left panel) between Ilexgenin A and OLR1 protein. The ligand is shown in deep green. **(D)** Myricetin and OLR1 protein. The ligand is shown in light blue. **(E)** Orientin and OLR1 protein. The ligand is shown in pink. The bound amino acids are labeled in black. Two-dimensional representation of docked residues is shown in the right panel. The drug molecules are shown in "stick" representation. The bound residues are shown in different colors, each color representing different types of interactions. The dotted lines represent the bonds.

**TABLE 1** Docking results for lectin like Ox-LDL receptor 1 (PDB ID: 7XMP).

Sr. no.	Compound (Pubchem CID)	Binding affinity (kcal/mol)	Interactions at the binding site		
			H-bond interaction	Van der Waals interactions	Other interactions
1	Apigenin (5280443)	-7.6	Ser207, Arg209	Ser240, Glu166, Trp165, Arg208, Pro214, Tyr213, Gly220, Glu218	Cys256, Leu216
2	Amentoflavone (5281600)	-8.3	Ser159, Tyr197	Asp147, Ile149, Ser198, Phe202, Phe200, Ser160	Phe261, Phe158, Ala194
3	Ilexgenin A (21672638)	-7.6	Trp217, Glu218	Arg208, Leu216, Glu218, Gly220, Asp219, Trp165, Glu166, Cys256, Cys243, Ser240	Arg209
4	Myricetin (5281672)	-6.5	Val230, Leu223, Pro211	Ala233, Arg229, Phe228, His226, Leu216, Pro222, Pro225, Arg208, Ser212	Trp215, Tyr212, Pro214
5	Orientin (5281675)	-7	Pro211, Leu223, Val230	Ser212, Arg208, Pro225, Pro222, Leu216, His226, Phe228, Arg229, Ala233	Trp215, Tyr212, Pro214



showed slight fluctuations; overall, the complexes exhibit the compactness of a complex structure. From the RMSD analysis of ligands, all ligands remained intact up to 25 ns, and no motion was observed, except for the ilexgenin A, which showed larger fluctuations in RMSD values, from 0.5 to 1 nm. It remained steady at the same value until 80 ns. It was also observed that the ligands' moment comes down to nearly  $\sim 0.6$  nm and then remained steady until 100 ns. After ilexgenin A, the second ligand apigenin showed changes in ligand moment from  $\sim 30$  to 60 ns. It remained steady until the end of simulation. The third ligand amentoflavone showed changes in ligand moment from 45 to 65 ns. Overall, the fluctuation rate was from 0.0 to 1.5 nm, while the other ligands did not exhibit any significant changes in moment (Figures 7A, B).

The RMSF analysis, as depicted by the RMSF graph (Figure 7C), elucidates the dynamic characteristics of specific regions within a protein structure. The RMSF graph shows that amino acid residues of the protein exhibit a higher degree of movement or flexibility. Quantitatively, these fluctuations were observed to extend up to 0.8 nm. Overall, the fluctuations reached 0.2 nm. Further inspection of the RMSF data reveals that the residues constituting the active site, as well as those in its vicinity, demonstrate considerable stability, with an average RMSF value of approximately 0.17 nm. Additionally, the analysis highlights a specific loop region, spanning from His226 to Cys243, exhibiting fluctuations up to 0.3 nm.

All five complexes showed a similar pattern in terms of Rg value. A noteworthy observation from the analysis is that all five complexes exhibited a consistent pattern regarding their Rg values (Figure 7D). This consistency in Rg values across the different complexes suggests a level of long-term stability and compactness within their structures, indicating that, despite the dynamic environment of the MD simulations, these complexes maintain their structural integrity over time. Specifically, the myricetin

complex was identified to exhibit notable fluctuations from 60 to 80 ns within the MD simulation trajectory.

The number of hydrogen bonds formed between ligands and protein target is shown in Figure 8. Myricetin and orientin formed high average numbers of hydrogen bonds compared to the other ligands. Myricetin formed five hydrogen bonds and orientin formed six hydrogen bonds with OLR1, whereas apigenin, amentoflavone, and ilexgenin displayed three, four, and four average hydrogen bonds, respectively. The average amount of hydrogen bonds produced over the whole simulation time may be helpful in determining which molecule is continually producing the greatest number of hydrogen bonds. Furthermore, we have identified the individual residues involved in the hydrogen bond formation across 100 frames. The residues mainly involved in two kinds of hydrogen bonding, such as bonding between backbone and side chain or bonding between side chain and side chain, were observed and are shown in Supplementary Figure S4. Among all the complexes, the number of residues involved in the hydrogen formation varies based on the attached ligands. Additionally, the lifetime of these hydrogen bonds was also examined, with specific criteria for analysis, including a cutoff radius of 0.35 nm and an angle cutoff of  $120^\circ$  (71).

The MMPBSA net binding energy order from the highest to lowest stability was as follows: amentoflavone  $-18.48$  kcal/mol, myricetin  $-14.13$  kcal/mol, orientin  $-13.75$  kcal/mol, ilexgenin A  $-11.07$  kcal/mol, and apigenin  $-9.55$  kcal/mol (Figure 9). Among all five complexes, amentoflavone, myricetin, and orientin had minimum binding energy, indicating these compounds as effective inhibitors (72). The lower binding energies for these complexes suggest that they have a higher potential to effectively inhibit the target (73), making them significant candidates for further investigation in drug discovery efforts against the OLR1. The complete details of binding energies are given in Supplementary Table S3.

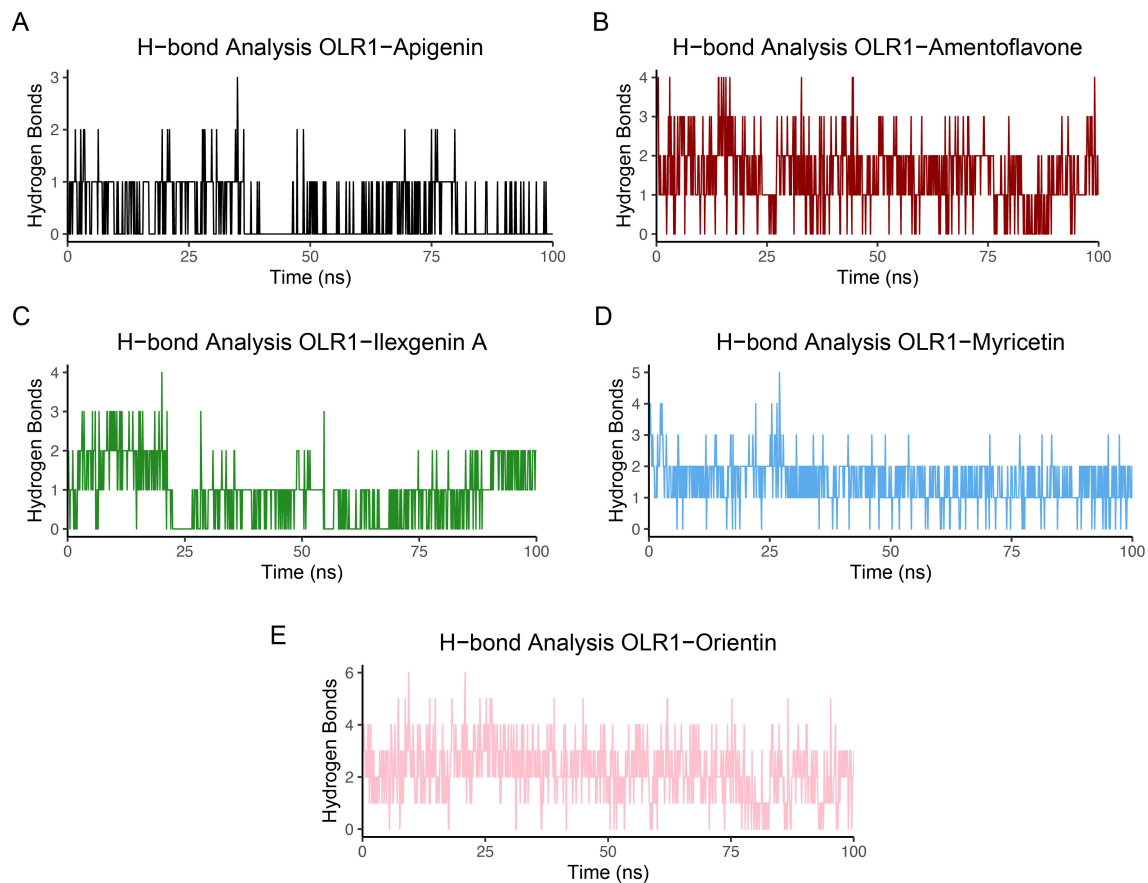


FIGURE 8

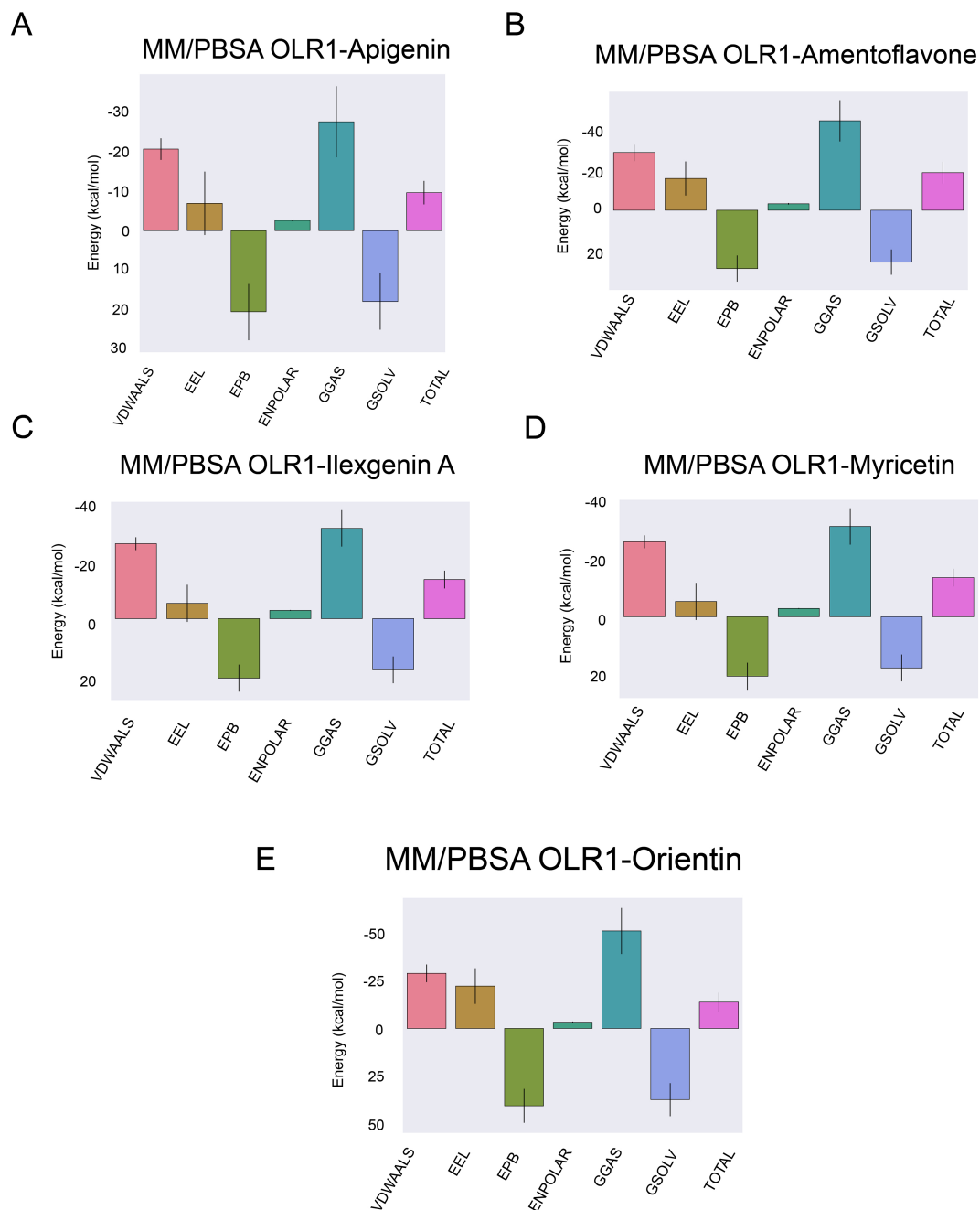
Hydrogen bond analysis between OLR1-receptor and bioflavonoid molecules. (A) OLR1–Apigenin, (B) OLR1–Amentoflavone, (C) OLR1–Ilexgenin A, (D) OLR1–Myricetin, and (E) OLR1–Orientin. The time frames of MD are plotted on the x-axis, and the number of hydrogen bonds is shown on the y-axis.

PCA was applied to calculate the motion of OLR1 protein in the bound and unbound state with the ligands (63). Additionally, the correlations between the conserved region of the protein were determined by computing a dynamic cross-correlation matrix (DCCM) (63) (Supplementary Figure S5). The PC1, PC2, PC3, and eigenvalues of apo were plotted against the respective eigenvector index for the first 20 modes of motion. Overall protein movement was controlled by eigenvectors, especially the higher ones, and the top five eigenvectors in our system demonstrated dominant movements with eigenvalues of 16.6%–60.6% while the remaining eigenvectors had lower eigenvalues. According to the PCA plot, the PC1 cluster retained the highest variability of 16.62%, PC2 exhibited 9.75% variability, while PC3 showed minimal variability (8.67%) (Supplementary Figure S6).

## 4 Discussion

The pathogenic association between EBV infection and LC has been previously reported by various studies (28) using clinical findings. It is believed that LC occurs as the result of the SARS-COV-mediated EBV reactivation mainly in the COVID-19 recovered population (74). However, their genomic bases of pathogenic

correlation need to be deeply explored; moreover, it is imperative to identify the key genes and their role in molecular pathways that could possibly act as biomarkers under LC conditions. The significance of association between COVID and EBV has been previously heightened by the studies of hospitalized COVID patients showing that reactivated EBV significantly increased mortality when compared to EBV-negative patients (75, 76). Recently, apart from the molecular techniques, NGS data have provided huge support to predict the genes and drug targets in similar scenarios (39, 77, 78). Hence, in this study, we have attempted to integrate the whole genome transcriptomic data of independent LC- and EBV-infected PBMCs to detect the key DEGs. A total 357 cDEGs were detected between EBV and LC samples. As it is crucial to understand the interaction network of genes within the cell (79), these genes were subjected to PPI analysis using the STRING database (50). Out of 357 cDEGs, only 125 genes were involved in the interaction's formation (interacting score > 0.4). The rationale for considering cDEGs was to exclude sample-specific DEGs and focus on the shared transcriptomic profile, as this shared profile potentially reflects the regulation of common key biological pathways in the cell (80). Furthermore, we identified the 22 hub genes using the cytoscope plugin CytoHubba (52) and MCODE (54), followed by the enrichment analysis with ClueGo. Our enrichment results were



**FIGURE 9** Binding free energy calculations. Energy calculations were performed using the MMPBSA method for each ligand and receptor complex. **(A)** OLR1–Apigenin, **(B)** OLR1–Amentoflavone, **(C)** OLR1–Ilexgenin A, **(D)** OLR1–Myricetin, and **(E)** OLR1–Orientin. The x-axis indicates the energy level in kcal/mol and types of energies are plotted on the y-axis.

consistent with previously reported studies (81). Since, during the disease condition, hub genes (because of being involved in the key functions) undergo substantial expression changes (82), we identified the individual gene expression under LC and EBV conditions. It was found that all 22 genes significantly changed ( $\log_2FC > 1$  and  $p\text{-value} < 0.05$ ) their expression in EBV and LC samples. TF enrichment analysis revealed that most of the hub genes were regulated by ZNF and MYB-family TFs. The involvement of these two TF classes has been previously reported in COVID (83) and EBV (84). As per our knowledge, there was not a single study available that

utilized a similar multi-dimensional strategy to identify the pathogenic connection of LC and EBV infection. However, there is a vast range of COVID-19-related studies available such as Noor et al., which predicted the common pathogenic genes' profile between COVID-19 and HFRS (40). Luo et al. explored the common pathogenic mechanism between COVID-19 and primary Sjogren's syndrome (pSS) (77). In our study, the enrichment analysis revealed the role of hub genes in several immunological, neurological, and pulmonary system-related pathways. Most of the gene ontologies have already been reported in accordance with roughly



COVID and EBV infections (28, 84–87). Because enrichment analysis comes from the biological intersection of LC and EBV DEGs, we expect their role in both LC and EBV infection symptoms simultaneously. These symptoms generally involved, the psychological problems/neurodegenerative disorders, or the terms associated with the brain function and the immunological disruption. Thus, there are existing lines of evidence suggesting that hub genes have a possible correlation with LC and EBV supported by previously published studies (15, 33, 88–90). Hub genes are functionally important and could act as very fine drug targets (91). Therefore, out of 22 hub genes, we choose one representative gene, OLR1 (oxidized low-density lipoprotein receptor 1), which was not studied previously with respect to LC-related manifestations. In COVID-19 patients, OLR1 gene positively expresses and causes cytokine storms and thrombosis (55). In severe COVID conditions, OLR1 is primarily involved in the activation of inflammatory immune responses (92). Considering the important role of hub genes in the cross-connection between LC and EBV reactivation, we used OLR1 protein as an important drug target and was docked against bioflavonoid inhibitors (70). The best five inhibitors, i.e., apigenin, amentoflavone, ilexgenin A, myricetin, and orientin, were screened out by virtual screening (93) based on their binding energies with OLR1. Several studies have reported the inhibitory and antiviral activities of these five compounds (72, 94–97).

Docking results provided great insight into the binding of shortlisted inhibitors with OLR1. Out of all five compounds, amentoflavone was considered as the best docked molecule with  $-8.3$  kcal/mol energy, and the remaining compounds were in  $6$ – $7.7$  kcal/mol range. Molecular docking gives insights into the ligand-receptor binding poses and pinpoints the bound amino acids with the ligands; however, it overlooks the strength and conformational modifications of the complex as well as the individual binding members that occur during the interaction's formation (98). The conformational stability of the complexes was further evaluated by MD simulations analyses, i.e., RMSD, RMSF, Rg, and hydrogen bonding. Comparative RMSD evaluation of ligand-bound and apo-OLR1 suggested the deviations' amplitude and the modest change in the average RMSD value of the C-backbone atoms, which clearly shows that the five OLR1 protein–ligand complex structures have a stable dynamic behavior. The comparative residual fluctuations of OLR1 were observed in both bound and apo form. N-terminus residues (start of protein) remain stable and did not exhibit any of the fluctuations until the 55th amino acid, followed by the fluctuations in between nearly 70 and 100 amino acids. Visual inspection of OLR1 protein suggested that this region mainly contains the loops that connect the alpha helices, and loops are considered as a variable structural component and they show a very dynamic behavior during the simulations (99, 100). Apart from the secondary structure, it was observed that the residues in the binding cavity also remained stable. Rg depicts the information about the compactness, shape, and folding of the four complex structures at various point scales throughout the 100 ns of MD simulations trajectory. The interaction between ligand and protein is influenced by non-covalent interactions including hydrogen bonding, hydrophobic forces, and ionic bonds. The quantity of hydrogen bonds and the duration for which they exist are key indicators of

the stability and binding strength of the complexes (101). All five complexes exhibited an average four to six hydrogen bonds and remained stable throughout the course of 100-ns simulations. Binding free energy is another parameter to estimate the stability of a complex (102, 103). The binding free energy of each complex was calculated, and the overall order was amentoflavone > myricetin > orientin > ilexgenin A > apigenin with the total binding energy of  $-18.48$ ,  $-14.13$ ,  $-13.75$ ,  $-11.07$ , and  $-9.55$  kcal/mol, respectively. The lowest binding energy is proportional to the high stability of the complex (102); among all bound complexes, amentoflavone showed the lowest total binding energy. The highest positive correlation after the apo form was observed in the amentoflavone-bound complex, which indicates that, overall, the global changes between amino acids remained stable. The highest fluctuations (negative correlation) were observed in apigenin, myricetin, orientin, and ilexgenin A. Thus, all the MD analyses provided evidence that five inhibitors' compounds formed stable and dynamic conformations; however, amentoflavone was ranked as potentially the best inhibitor against OLR1 protein.

Although several previous independent studies have been carried out to explore the LC and EBV infection-related manifestations, there is room for studies that could predict the possible pathogenic biomarker genes and their inhibitors in the context of LC using integrative bioinformatics methods. Collectively, in this study, for the first time, we have attempted to explore and detect cDEGs, followed by hub genes and their associated TF identification using bioinformatics enrichment methods. Moreover, the hub gene protein (OLR1) interaction with bioflavonoids was performed using molecular docking and simulations. Among the five bioflavonoids, amentoflavone was predicted as the best inhibitor against OLR1.

However, it is important to acknowledge the potential limitations of this study. First, our findings are primarily based on publicly available transcriptomic datasets with a limited sample size. This may not be sufficient to identify all the key genes; thus, an increased sample size and sequencing depth could enhance the accuracy of predictions. Second, the statistical tools and models used to analyze the data may lead to overinterpretation, particularly when making inferences about causality or functional roles without adequate validation. Therefore, further *in vitro* or *in vivo* experiments would be valuable to functionally validate the hub genes and inhibitors predicted to target the hub protein in the clinical context of LC.

In summary, this study offers a valuable *in silico* framework for predicting the molecular links between EBV-mediated LC pathogenesis and may help identify potential marker genes with clinical significance, as well as inhibitors that could block the active sites of their corresponding proteins in LC. To sum up, our analysis predicted the possible genes and their pathways linking EBV reactivation and LC, elucidating some unknown clues in between. Nonetheless, as this is a thorough computational work, further case reports and follow-up experiments of LC patients can corroborate this association.

## 5 Conclusion

In summary, our results predicted 22 hub genes (CCL2, CCL20, CDCA2, CEP55, CHI3L1, CKAP2L, DEPDC1, DIAPH3, DLGAP5, E2F8, FGF1, NEK2, PBK, TOP2A, CCL3, CXCL8, DEPDC1, IL6,

RETN, MMP2, LCN2, and OLR1) in the EBV infection and LC collective scenario with a significant change in expression. Through a series of comprehensive analyses, we predicted the biological roles of these genes in relation to EBV and LC. Our findings suggest that these genes are primarily involved in various immune-signaling pathways, including JAK-STAT signaling, interleukin signaling, protein kinase signaling, and toll-like receptor pathways, all of which are connected to LC. Furthermore, we propose an *in silico* framework to potentially uncover the clinical significance of hub genes associated with EBV infection in LC. Moreover, our results predicted the potential of hub genes as a drug target against bioflavonoids, which may serve as valuable therapeutic signatures in the clinical field upon further validation.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

## Author contributions

AK: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. ZZ: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is

## References

1. Coronavirus disease (COVID-19) situation Reports (n.). Available online at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (Accessed February 18, 2024).
2. Eythorsson E, Helgason D, Ingvarsson RF, Bjornsson HK, Olafsdottir LB, Bjarnadottir V, et al. Clinical spectrum of coronavirus disease 2019 in Iceland: population based cohort study. *BMJ*. (2020) 371. doi: 10.1136/BMJ.M4529
3. Lagier JC, Million M, Gautret P, Colson P, Cortaredona S, Giraud-Gatineau A, et al. Outcomes of 3,737 COVID-19 patients treated with hydroxychloroquine/azithromycin and other regimens in Marseille, France: A retrospective analysis. *Travel Med Infect Dis*. (2020) 36:101791. doi: 10.1016/j.tmaid.2020.101791
4. Maltezou HC, Raftopoulos V, Vorou R, Papadima K, Mellou K, Spanakis N, et al. Association between upper respiratory tract viral load, comorbidities, disease severity, and outcome of patients with SARS-CoV-2 infection. *J Infect Dis*. (2021) 223:1132–8. doi: 10.1093/infdis/jiaa804
5. Havervall S, Rosell A, Phillipson M, Mangsbo SM, Nilsson P, Hober S, et al. Symptoms and functional impairment assessed 8 months after mild COVID-19 among health care workers. *JAMA*. (2021) 325:2015–6. doi: 10.1001/JAMA.2021.5612
6. Pavli A, Theodoridou M, Maltezou HC. Post-COVID syndrome: incidence, clinical spectrum, and challenges for primary healthcare professionals. *Arch Med Res*. (2021) 52:575. doi: 10.1016/j.arcmed.2021.03.010
7. Jones OY, Yeralan S. Is long COVID a state of systemic pericyte disarray? *J Clin Med*. (2022) 11:572. doi: 10.3390/JCM11030572
8. Alpert O, Begun L, Garren P, Solkhah R. Cytokine storm induced new onset depression in patients with COVID-19. A new look into the association between depression and cytokines -two case reports. *Brain Behav Immun Health*. (2020) 9:100173. doi: 10.1016/j.bbih.2020.100173
9. Fernández-de-las-Peñas C, Gujjarro C, Plaza-Canteli S, Hernández-Barrera V, Torres-Macho J. Prevalence of post-COVID-19 cough one year after SARS-CoV-2 infection: A multicenter study. *Lung*. (2021) 199:249. doi: 10.1007/S00408-021-00450-W
10. McCorkell L, Assaf GS, Davis HE, Wei H, Akrami A. Patient-Led Research Collaborative: embedding patients in the Long COVID narrative. *Pain Rep*. (2021) 6:e913. doi: 10.1097/PR9.0000000000000913
11. Morin L, Savale L, Pham T, Colle R, Figueiredo S, Harrois A, et al. Four-month clinical status of a cohort of patients after hospitalization for COVID-19. *JAMA*. (2021) 325:1525–34. doi: 10.1001/JAMA.2021.3331
12. Bull-Otterson L, Baca S, Saydah S, Boehmer TK, Adjei S, Gray S, et al. Post-COVID conditions among adult COVID-19 survivors aged 18–64 and ≥65 years — United States, March 2020–November 2021. *MMWR Morb Mortal Wkly Rep*. (2022) 71:713–7. doi: 10.15585/MMWR.MM7121E1
13. Ceban F, Ling S, Lui LMW, Lee Y, Gill H, Teopiz KM, et al. Fatigue and cognitive impairment in Post-COVID-19 Syndrome: A systematic review and meta-analysis. *Brain Behav Immun*. (2022) 101:93–135. doi: 10.1016/j.bbi.2021.12.020

supported by the National Key Research and Development Program of China (2021YFA1301504), the Chinese Academy of Sciences Strategic Priority Research Program (XDB37040202), and the National Natural Science Foundation of China (91953101).

## Acknowledgments

The authors are thankful to all lab members for providing technical support and advice.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2024.1435170/full#supplementary-material>

14. Patients diagnosed with post-COVID conditions: an analysis of private healthcare claims using the official ICD-10 diagnostic code - Digital Collections - National Library of Medicine (n). Available online at: <https://collections.nlm.nih.gov/catalog/nlm:nlmuid-9918504887106676-pdf> (Accessed February 18, 2024).
15. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med.* (2022) 28:583–90. doi: 10.1038/S41591-022-01689-3
16. Xie Y, Al-Aly Z. Risks and burdens of incident diabetes in long COVID: a cohort study. *Lancet Diabetes Endocrinol.* (2022) 10:311–21. doi: 10.1016/S2213-8587(22)00044-4
17. Mancini DM, Brunjes DL, Lala A, Trivieri MG, Contreras JP, Natelson BH. Use of cardiopulmonary stress testing for patients with unexplained dyspnea post-coronavirus disease. *JACC Heart Fail.* (2021) 9:927–37. doi: 10.1016/J.JCHF.2021.10.002
18. Demko ZO, Yu T, Mullapudi SK, Heslin MG, Dorsey CA, Payton CB, et al. Post-acute sequelae of SARS-CoV-2 (PASC) impact quality of life at 6, 12 and 18 months post-infection. *medRxiv.* (2022) 2022:8.08.22278543. doi: 10.1101/2022.08.08.22278543
19. Kedor C, Freitag H, Meyer-Arndt L, Wittke K, Hanitsch LG, Zoller T, et al. A prospective observational study of post-COVID-19 chronic fatigue syndrome following the first pandemic wave in Germany and biomarkers associated with symptom severity. *Nat Commun.* (2022) 13:1–13. doi: 10.1038/s41467-022-32507-6
20. Larsen NW, Stiles LE, Shaik R, Schneider L, Muppidi S, Tsui CT, et al. Characterization of autonomic symptom burden in long COVID: A global survey of 2,314 adults. *Front Neurol.* (2022) 13:1012668. doi: 10.3389/FNEUR.2022.1012668
21. Venkatesan P. NICE guideline on long COVID. *Lancet Respir Med.* (2021) 9:129. doi: 10.1016/S2213-2600(21)00031-X
22. Greenhalgh T, Sivan M, Delaney B, Evans R, Milne R. Long covid—an update for primary care. *BMJ.* (2022) 378. doi: 10.1136/BMJ-2022-072117
23. Swank Z, Senussi Y, Manickas-Hill Z, Yu XG, Li JZ, Alter G, et al. Persistent circulating severe acute respiratory syndrome coronavirus 2 spike is associated with post-acute coronavirus disease 2019 sequelae. *Clin Infect Dis.* (2023) 76:E487–90. doi: 10.1093/CID/C1AC722
24. Proal AD, VanElzakker MB. Long COVID or post-acute sequelae of COVID-19 (PASC): an overview of biological factors that may contribute to persistent symptoms. *Front Microbiol.* (2021) 12:698169/BIBTEX. doi: 10.3389/FMICB.2021.698169/BIBTEX
25. Charfeddine S, Ibn Hadj Amor H, Jdidi J, Torjmen S, Kraiem S, Hammami R, et al. Long COVID 19 syndrome: is it related to microcirculation and endothelial dysfunction? Insights from TUN-endCOV study. *Front Cardiovasc Med.* (2021) 8:745758. doi: 10.3389/FCVM.2021.745758
26. Pretorius E, Venter C, Laubscher GJ, Kotze MJ, Oladejo SO, Watson LR, et al. Prevalence of symptoms, comorbidities, fibrin amyloid microclots and platelet pathology in individuals with Long COVID/Post-Acute Sequelae of COVID-19 (PASC). *Cardiovasc Diabetol.* (2022) 21:1–17. doi: 10.1186/S12933-022-01579-5/FIGURES/12
27. Zubchenko S, Kril I, Nadizhko O, Matsyura O, Chopyak V. Herpesvirus infections and post-COVID-19 manifestations: a pilot observational study. *Rheumatol Int.* (2022) 42:1523–30. doi: 10.1007/S00296-022-05146-9
28. Vojdani A, Vojdani E, Saidara E, Maes M. Persistent SARS-coV-2 infection, EBV, HHV-6 and other factors may contribute to inflammation and autoimmunity in long COVID. *Viruses.* (2023) 15:400. doi: 10.3390/V15020400
29. Su Y, Yuan D, Chen DG, Ng RH, Wang K, Choi J, et al. Multiple early factors anticipate post-acute COVID-19 sequelae. *Cell.* (2022) 185:881–895.e20. doi: 10.1016/J.CELL.2022.01.014
30. Aloisi F, Giovannoni G, Salvetti M. Epstein-Barr virus as a cause of multiple sclerosis: opportunities for prevention and therapy. *Lancet Neurol.* (2023) 22:338–49. doi: 10.1016/S1474-4422(22)00471-9
31. Bjornevik K, Münz C, Cohen JL, Ascherio A. Epstein-Barr virus as a leading cause of multiple sclerosis: mechanisms and implications. *Nat Rev Neurol.* (2023) 19:160–71. doi: 10.1038/S41582-023-00775-5
32. Chen T, Song J, Liu H, Zheng H, Chen C. Positive Epstein-Barr virus detection in coronavirus disease 2019 (COVID-19) patients. *Sci Rep.* (2021) 11:10902. doi: 10.1038/S41598-021-90351-Y
33. Gold JE, Okyay RA, Licht WE, Hurley DJ. Investigation of long COVID prevalence and its relationship to Epstein-Barr virus reactivation. *Pathogens.* (2021) 10:763. doi: 10.3390/PATHOGENS10060763
34. Paganelli R. Resurrecting Epstein-Barr virus. *Pathogens.* (2022) 11:772. doi: 10.3390/PATHOGENS11070772
35. von der Thüsen J, van der Eerden M. Histopathology and genetic susceptibility in COVID-19 pneumonia. *Eur J Clin Invest.* (2020) 50:e13259. doi: 10.1111/EJC.13259
36. Evans RA, Leavy OC, Richardson M, Elneima O, McAuley HJC, Shikotra A, et al. Clinical characteristics with inflammation profiling of long COVID and association with 1-year recovery following hospitalisation in the UK: a prospective observational study. *Lancet Respir Med.* (2022) 10:761–75. doi: 10.1016/S2213-2600(22)00127-8
37. Chapola H, de Bastiani MA, Duarte MM, Freitas MB, Schuster JS, de Vargas DM, et al. A comparative study of COVID-19 transcriptional signatures between clinical samples and preclinical cell models in the search for disease master regulators and drug repositioning candidates. *Virus Res.* (2023) 326:199053. doi: 10.1016/J.VIRUSRES.2023.199053
38. Zhou L, Ding L, Gong Y, Zhao J, Xin G, Zhou R, et al. Identification of hub genes associated with the pathogenesis of diffuse large B-cell lymphoma subtype one characterized by host response via integrated bioinformatic analyses. *PeerJ.* (2020) 8:e10269. doi: 10.7717/PEERJ.10269/SUPP-11
39. Chen Q, Xia S, Sui H, Shi X, Huang B, Wang T. Identification of hub genes associated with COVID-19 and idiopathic pulmonary fibrosis by integrated bioinformatics analysis. *PLoS One.* (2022) 17:e0262737. doi: 10.1371/JOURNAL.PONE.0262737
40. Noor F, Ashfaq UA, Bakar A, ul Haq W, Allemailem KS, Alharbi BF, et al. Discovering common pathogenic processes between COVID-19 and HFERS by integrating RNA-seq differential expression analysis with machine learning. *Front Microbiol.* (2023) 14:1175844/BIBTEX. doi: 10.3389/FMICB.2023.1175844/BIBTEX
41. Zhao P, Zhen H, Zhao H, Huang Y, Cao B. Identification of hub genes and potential molecular mechanisms related to radiotherapy sensitivity in rectal cancer based on multiple datasets. *J Transl Med.* (2023) 21:1–16. doi: 10.1186/S12967-023-04029-2/FIGURES/10
42. Yin K, Peluso MJ, Luo X, Thomas R, Shin MG, Neideman J, et al. Long COVID manifests with T cell dysregulation, inflammation and an uncoordinated adaptive immune response to SARS-CoV-2. *Nat Immunol.* (2024) 25:218–25. doi: 10.1038/s41590-023-01724-6
43. Ma N, Lu J, Pei Y, Robertson ES. Transcriptome reprogramming of Epstein-Barr virus infected epithelial and B cells reveals distinct host-virus interaction profiles. *Cell Death Dis.* (2022) 13:894. doi: 10.1038/S41419-022-05327-1
44. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes.* (2016) 9:1–7. doi: 10.1186/S13104-016-1900-2
45. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* (2013) 29:15. doi: 10.1093/BIOINFORMATICS/BTS635
46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* (2009) 25:2078–9. doi: 10.1093/BIOINFORMATICS/BTP352
47. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* (2014) 30:923–30. doi: 10.1093/BIOINFORMATICS/BTT656
48. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) 15:1–21. doi: 10.1186/S13059-014-0550-8/FIGURES/9
49. Ge SX, Jung D, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics.* (2020) 36:2628–9. doi: 10.1093/BIOINFORMATICS/BTZ931
50. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* (2023) 51:D638–46. doi: 10.1093/NAR/GKAC1000
51. Yu D, Lim J, Wang X, Liang F, Xiao G. Enhanced construction of gene regulatory networks using hub gene information. *BMC Bioinf.* (2017) 18:1–20. doi: 10.1186/S12859-017-1576-1/FIGURES/7
52. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* (2014) 8:1–7. doi: 10.1186/1752-0509-8-S4-S11/TABLES/4
53. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/GR.1239303
54. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf.* (2003) 4:1–27. doi: 10.1186/1471-2105-4-2/FIGURES/12
55. Combadière B, Adam L, Guillou N, Quentrec P, Rosenbaum P, Dorgham K, et al. LOX-1-expressing immature neutrophils identify critically-ill COVID-19 patients at risk of thrombotic complications. *Front Immunol.* (2021) 12:752612/BIBTEX. doi: 10.3389/FIMMU.2021.752612/BIBTEX
56. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* (2016) 44:W90. doi: 10.1093/NAR/GKW377
57. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* (2004) 25:1605–12. doi: 10.1002/JCC.20084
58. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. *Methods Mol Biol.* (2015) 1263:243–50. doi: 10.1007/978-1-4939-2269-7\_19
59. Yue-Dong G, Jing-Fei H, Yue-Dong G, Jing-Fei H. An extension strategy of Discovery Studio 2.0 for non-bonded interaction energy automatic calculation at the residue level. *Zoological Res.* (2011) 32:262–6. doi: 10.3724/SP.J.1141.2011.03262
60. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The amber biomolecular simulation programs. *J Comput Chem.* (2005) 26:1668. doi: 10.1002/JCC.20290



61. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. (2015) 1–2:19–25. doi: 10.1016/j.softx.2015.06.001
62. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins*. (2006) 65:712. doi: 10.1002/PROT.21123
63. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Cavas LSD. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*. (2006) 22:2695–6. doi: 10.1093/BIOINFORMATICS/BTL461
64. Zhang X, Perez-Sanchez H, XXXC, Lightstone F. A comprehensive docking and MM/GBSA rescoring study of ligand recognition upon binding antithrombin. *Curr Top Med Chem*. (2017) 17:1631–9. doi: 10.2174/1568026616666161117112604
65. Espín E, Yang C, Shannon CP, Assadian S, He D, Tebbutt SJ. Cellular and molecular biomarkers of long COVID: a scoping review. *EBioMedicine*. (2023) 91:104552. doi: 10.1016/j.ebiom.2023.104552
66. Nyamundanda G, Poudel P, Patil Y, Sadanandam A. A novel statistical method to diagnose, quantify and correct batch effects in genomic studies. *Sci Rep* 2017 7:1. (2017) 7:1–10. doi: 10.1038/s41598-017-11110-6
67. Fricke-Galindo I, Buendia-Roldan I, Chavez-Galan L, Pérez-Rubio G, Hernández-Zenteno R, de J, et al. SERPINE1 rs6092 variant is related to plasma coagulation proteins in patients with severe COVID-19 from a tertiary care hospital. *Biol (Basel)*. (2022) 11:595. doi: 10.3390/BIOLOGY11040595
68. Nikitopoulou I, Fanidis D, Ntatsoulis K, Moulos P, Mpekoulis G, Evangelidou M, et al. Increased autotaxin levels in severe covid-19, correlating with il-6 levels, endothelial dysfunction biomarkers, and impaired functions of dendritic cells. *Int J Mol Sci*. (2021) 22:10006. doi: 10.3390/IJMS221810006/S1
69. Hill M, Russo S, Olivera D, Malcuori M, Galliussi G, Segovia M. The intracellular cation channel TMEM176B as a dual immunoregulator. *Front Cell Dev Biol*. (2022) 10:1038429. doi: 10.3389/FCELL.2022.1038429
70. Patil R, Chikhale R, Khanal P, Gurav N, Ayyanar M, Sinha S, et al. Computational and network pharmacology analysis of bioflavonoids as possible natural antiviral compounds in COVID-19. *Inform Med Unlocked*. (2021) 22:100504. doi: 10.1016/J.IMU.2020.100504
71. Scheurer M, Rodenkirch P, Siggel M, Bernardi RC, Schulten K, Tajkhorshid E, et al. PyContact: rapid, customizable, and visual analysis of noncovalent interactions in MD simulations. *Biophys J*. (2018) 114:577–83. doi: 10.1016/J.BJP.2017.12.003
72. Dey D, Hossain R, Biswas P, Paul P, Islam MA, Ema TI, et al. Amentoflavone derivatives significantly act towards the main protease (3CLPRO/MPRO) of SARS-CoV-2: in silico admet profiling, molecular docking, molecular dynamics simulation, network pharmacology. *Mol Divers*. (2023) 27:857–71. doi: 10.1007/S11030-022-10459-9
73. Wu C, Lui Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharmaceutica Sinica B*. (2020) 10(5):766–88.
74. Meng M, Zhang S, Dong X, Sun W, Deng Y, Li W, et al. COVID-19 associated EBV reactivation and effects of ganciclovir treatment. *Immun Inflammation Dis*. (2022) 10. doi: 10.1002/IID3.597
75. Xie Y, Cao S, Dong H, Lv H, Teng X, Zhang J, et al. Clinical characteristics and outcomes of critically ill patients with acute COVID-19 with Epstein-Barr virus reactivation. *BMC Infect Dis*. (2021) 21:1–8. doi: 10.1186/S12879-021-06638-Y
76. Manoharan S, Ying LY. Epstein Barr Virus Reactivation during COVID-19 Hospitalization Significantly Increased Mortality/Death in SARS-CoV-2(+)/EBV(+) than SARS-CoV-2(-)/EBV(-) Patients: A Comparative Meta-Analysis. *Int J Clin Pract*. (2023) 2023. doi: 10.1155/2023/1068000
77. Luo H, Zhou X. Bioinformatics analysis of potential common pathogenic mechanisms for COVID-19 infection and primary Sjogren's syndrome. *Front Immunol*. (2022) 13:938837/BIBTEX. doi: 10.3389/FIMMU.2022.938837/BIBTEX
78. Xiong Z, Yang Y, Li W, Lin Y, Huang W, Zhang S. Exploring key biomarkers and common pathogenesis of seven digestive system cancers and their correlation with COVID-19. *Curr Issues Mol Biol*. (2023) 45:5515. doi: 10.3390/CIMB45070349
79. Jordin F, Nguyen TP, Liu WC. Studying protein–protein interaction networks: a systems view on diseases. *Brief Funct Genomics*. (2012) 11:497–504. doi: 10.1093/BFGP/ELS035
80. Dugourd A, Saez-Rodriguez J. Footprint-based functional analysis of multiomic data. *Curr Opin Syst Biol*. (2019) 15:82–90. doi: 10.1016/J.COISB.2019.04.002
81. Hashimoto K. Detrimental effects of COVID-19 in the brain and therapeutic options for long COVID: The role of Epstein–Barr virus and the gut–brain axis. *Mol Psychiatry* 2023. (2023) 28:1–9. doi: 10.1038/s41380-023-02161-5
82. Mengyan X, Kun D, Xinming J, Yutian W, Yongqian S. Identification and verification of hub genes associated with the progression of non-small cell lung cancer by integrated analysis. *Front Pharmacol*. (2022) 13:997842/FULL. doi: 10.3389/FPHAR.2022.997842/FULL
83. Esposito S, D'abrosca G, Antolak A, Pedone PV, Isernia C, Malgieri G. Host and viral zinc-finger proteins in COVID-19. *Int J Mol Sci*. (2022) 23, 3711. doi: 10.3390/IJMS23073711
84. Tune CE, Pilon M, Saiki Y, Dosch H-M. Sustained expression of the novel EBV-induced zinc finger gene, ZNFEB, is critical for the transition of B lymphocyte activation to oncogenic growth transformation. *J Immunol*. (2002) 168:680–8. doi: 10.4049/JIMMUNOL.168.2.680
85. Thorley-Lawson DA. EBV persistence—Introducing the virus. *Curr Top Microbiol Immunol*. (2015) 390:151. doi: 10.1007/978-3-319-22822-8\_8
86. Lin W, Yip YL, Jia L, Deng W, Zheng H, Dai W, et al. Establishment and characterization of new tumor xenografts and cancer cell lines from EBV-positive nasopharyngeal carcinoma. *Nat Commun*. (2018) 9:1–17. doi: 10.1038/s41467-018-06889-5
87. Shannon-Lowe C, Rickinson A. The global landscape of EBV-associated tumors. *Front Oncol*. (2019) 9:713. doi: 10.3389/FONC.2019.00713
88. Klein J, Wood J, Jaycox J, Lu P, Dhodapkar RM, Gehlhausen JR, et al. Distinguishing features of Long COVID identified through immune profiling. *Nature*. (2023) 623.7985:139–48. doi: 10.1101/2022.08.09.22278592
89. Monje M, Iwasaki A. The neurobiology of long COVID. *Neuron*. (2022) 110, 3484. doi: 10.1016/J.NEURON.2022.10.006
90. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nat Rev Microbiol*. (2023) 21:133. doi: 10.1038/S41579-022-00846-2
91. Chen X, Xia Z, Wan Y, Huang P. Identification of hub genes and candidate drugs in hepatocellular carcinoma by integrated bioinformatics analysis. *Medicine*. (2021) 100:E27117. doi: 10.1097/MD.00000000000027117
92. Schimke LF, Marques AHC, Baiocchi GC, Prado CA, de S, Fonseca DLM, et al. Severe COVID-19 shares a common neutrophil activation signature with other acute inflammatory states. *Cells*. (2022) 11:847. doi: 10.3390/CELLS11050847/S1
93. Li Q, Shah S. Structure-based virtual screening. *Methods Mol Biol*. (2017) 1558:111–24. doi: 10.1007/978-1-4939-6783-4\_5
94. Yang H, Wang J, Fan JH, Zhang YQ, Zhao JX, Dai XJ, et al. Ilexgenin A exerts anti-inflammation and anti-angiogenesis effects through inhibition of STAT3 and PI3K pathways and exhibits synergistic effects with Sorafenib on hepatoma growth. *Toxicol Appl Pharmacol*. (2017) 315:90–101. doi: 10.1016/J.TAAP.2016.12.008
95. Bhowmik A, Biswas S, Hajra S, Saha P. In silico validation of potent phytochemical orientin as inhibitor of SARS-CoV-2 spike and host cell receptor GRP78 binding. *Heliyon*. (2021) 7. doi: 10.1016/J.HELIYON.2021.E05923
96. Lin L, Chen DY, Scartelli C, Xie H, Merrill-Skoloff G, Yang M, et al. Plant flavonoid inhibition of SARS-CoV-2 main protease and viral replication. *iScience*. (2023) 26:107602. doi: 10.1016/J.ISCL.2023.107602
97. Pan H, He J, Yang Z, Yao X, Zhang H, Li R, et al. Myricetin possesses the potency against SARS-CoV-2 infection through blocking viral-entry facilitators and suppressing inflammation in rats and mice. *Phytomedicine*. (2023) 116:154858. doi: 10.1016/J.PHYMED.2023.154858
98. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. *Molecules*. (2015) 20:13384–421. doi: 10.3390/MOLECULES200713384
99. Mitusińska K, Skalski T, Góra A. Simple selection procedure to distinguish between static and flexible loops. *Int J Mol Sci*. (2020) 21. doi: 10.3390/IJMS21072293
100. Galdadas I, Qu S, Oliveira ASF, Olehnovics E, Mack AR, Mojica MF, et al. Allosteric communication in class A  $\beta$ -lactamases occurs via cooperative coupling of loop dynamics. *Elife*. (2021) 10. doi: 10.7554/ELIFE.66567
101. Varma AK, Patil R, Das S, Stanley A, Yadav L, Sudhakar A. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PLoS One*. (2010) 5. doi: 10.1371/JOURNAL.PONE.0012029
102. Majewski M, Ruiz-Carmona S, Barril X. An investigation of structural stability in protein–ligand complexes reveals the balance between order and disorder. *Commun Chem*. (2019) 2:1–8. doi: 10.1038/s42004-019-0205-5
103. Hata H, Tran DP, Sobeh MM, Kitao A, Sobeh MM. Binding free energy of protein/ligand complexes calculated using dissociation Parallel Cascade Selection Molecular Dynamics and Markov state model. *Biophys Physicobiol*. (2021) 18:305. doi: 10.2142/BIOPHYSICO.BPPB-V18.037