Received: 15 February 2024 | Revised: 17 April 2024 | Accepted: 19 June 2024 | Published online: 26 June 2024

RESEARCH ARTICLE

Medinformatics
2024, Vol. 00(00) 1–11
DOI: 10.47852/bonviewMEDIN42022621

BON VIEW PUBLISHING

Unraveling the Shared Genetic Elements Among COVID-19 and Epilepsy: A Computational and Bioinformatics Analysis

Pavithra Nagendran[1], Gowtham Murugesan[1] and Jeyakumar Natarajan[1,*]

[1]*Department of Bioinformatics, Bharathiar University, India*

Abstract: The outbreak of COVID-19 linked to the SARS-CoV-2 virus has raised worries about its potential to worsen pre-existing health issues such as autoimmune conditions. One common autoimmune condition seen across all individuals is epilepsy. Our detailed computational study examined the complex connection between COVID-19 and epilepsy, with a focus on their possible impact on neurodegenerative disorders. Leveraging transcriptome analysis, we aimed to interpret shared pathways and molecular biomarkers between COVID-19 and epilepsy using RNA-seq datasets. Cutting-edge bioinformatics tools facilitated Network, Enrichment, and route analysis, along with the identification of pivotal gene signatures, notably through comprehensive exploration of machine learning methods and protein-protein interactions. This work identified 1040 common differentially expressed genes, forming the basis for unraveling shared pathways and potential drug targets. By employing statistical methods and diverse network analyses, we meticulously identified genetic elements shared between COVID-19 and epilepsy. The analysis reveals that 10 key genes, including CATIP, CDC25C, GPR132, NTS, PDE8B, PLK1, SLC12A9, SPC25, TUBAIA, and TYMS, are crucial in bridging transcriptomic alterations across various scenarios. A subset of these genes exhibited existence within distinct regulatory networks, signifying their significant role in the shared disease mechanisms of COVID-19 and epilepsy. Notably, CDC25C, PLK1, and TYMS emerged as prominent genes within the COVID-19 and epilepsy networks, strongly suggesting their vital roles in connecting regulatory mechanisms across these conditions. Further validation through molecular docking studies will confirm the significance of CDC25C, PLK1, and TYMS, potentially shedding light on new opportunities for targeted therapeutic interventions aimed at reducing the risk of seizures in individuals affected by COVID-19 and epilepsy.

Keywords: COVID-19, epilepsy, DEG analysis, network analysis, biomarkers

1. Introduction

The COVID-19 pandemic has brought to light various complexities beyond its primary respiratory impacts. Among these complexities, the virus's potential impact on neurological health has emerged as a significant concern. As the virus primarily affects the respiratory system, there is growing evidence suggesting its ability to infiltrate the central nervous system (CNS), thereby raising concerns about its potential effects on brain function and neurological well-being. Infections of the CNS, such as meningitis, viral encephalitis, malaria, and neurocysticercosis, are common causes of seizures and acquired epilepsy in the developing world [1, 2]. These infections can lead to increased mortality and morbidity, including subsequent Epilepsy. Notably, a considerable proportion of individuals affected by COVID-19 have presented with neurological manifestations. These neurological complications range from mild symptoms to severe conditions, indicating the virus's ability to affect various regions of the brain. The observation of these neurological manifestations has prompted intensive research into understanding the underlying mechanisms and identifying the shared genes implicated in COVID-19 infection, particularly focusing on individuals with compromised immune systems [3]. Epilepsy is a chronic brain disorder in which groups of nerve cells, or neurons, in the brain sometimes send the wrong signals and cause seizures. Seizures, characterized by sudden, abnormal electrical activity in the brain, have been observed in some COVID-19 patients, especially in individuals with weakened immune systems [4]. The onset of seizures can arise from various triggers, such as sudden electrical impulses or disturbances in neural communication, often leading to sudden discomfort or even temporary paralysis in affected patients. Studies found that the case fatality rate for patients with Epilepsy and COVID-19 was 9.8% [5].

Considering these neurological complications associated with COVID-19, there is a growing imperative to explore deeper into the genetic reinforcements of these neurological functions. Understanding the shared genes involved in COVID-19 infection and their potential influence on neurological manifestations, including seizures, holds promise for developing targeted drug therapies. Such drugs could aim to alleviate the neurological impact of the virus and improve associated complications, thereby improving patient outcomes. COVID-19 can cause various

*Corresponding author: Jeyakumar Natarajan, Department of Bioinformatics, Bharathiar University, India. Email: n.jeyakumar@buc.edu.in

neurological dysfunctions, such as loss of smell and taste, severe strokes, headache, dizziness, and encephalitis [6]. The potential neurological impact of SARS-CoV-2 infection on the brain includes symptoms such as headache, dizziness, impaired consciousness, cognitive impairment, and motor disorders [7].

The relationship between COVID-19 and epilepsy investigates the potential impact of the viral infection on individuals with pre-existing epilepsy, as well as the potential for COVID-19 to induce new-onset seizures or exacerbate existing epilepsy. It also addresses the challenges of managing epilepsy during the pandemic and the potential neurological implications of COVID-19 [8]. The SARS-CoV-2 virus affects the brain, leading to cognitive dysfunction and other neurological sequelae in COVID-19 patients [9]. Also, studies have identified that COVID-19 can cause a temporary loss of smell and strokes related to blood clotting and inflammation in brain blood vessels [10]. New-onset epilepsy noted in 0.3% of patients within 6 months of COVID-19 infection, and three cases of post-COVID-19 epilepsy were identified after non-symptomatic to mild disease [11]. COVID-19 infection affects the nervous system in epilepsy patients resulting autonomic nervous system dysfunction and psychological disorders such as depression and anxiety [12]. COVID-19 worsens seizures in elderly patients with epilepsy, including lung damage and hypoxia, which contribute to a poor prognosis. Apart from epilepsy, severe neurological manifestations such as neuroinvasion, endothelial dysfunction, and neuroinflammation are commonly caused in COVID-19 patients [13, 14]. Research is being conducted to understand and treat these symptoms. A case report described a patient with new-onset focal epilepsy and impaired awareness seizures associated with COVID-19 infection, suggesting a possible post-COVID-19 inflammatory syndrome [15]. Patients with epilepsy may have an immune response that predisposes them to developing COVID-19.It is suggested to explore this correlation by conducting studies specifically focused on investigating the immune response of patients with epilepsy in relation to the development of COVID-19.

In the current study, we selected two single-cell RNA-Seq and RNA-Seq datasets from the Gene Expression Omnibus (GEO) database to identify the key candidate genes between COVID-19 and Epilepsy. First, we identified DEGs between COVID-19 and Epilepsy. Then, we applied ML-based algorithms to identify the significant genes between COVID-19 and Epilepsy, we used Common DEGs to discover the functions and obtained Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. A protein-protein interaction (PPI) network was constructed using the Network Analyst database, incorporating a transcription factor (TF)-gene interaction network to identify highly promising modules. The degree of connectivity within these modules was then assessed to pinpoint potential biomarkers. The objective of this study is to investigate the potential impact of COVID-19 on the susceptibility to seizures and the development of epilepsy in individuals with compromised immune systems. Figure 1 summarized the data preparation, processing, analysis, and validation.

In essence, this study aims to explore the shared genetic factors associated with COVID-19 infection, particularly in the context of neurological manifestations, including seizures. By elucidating these genetic components, the goal is to pave the way for the development of novel therapeutic interventions that specifically target neurological dysfunctions arising from COVID-19 infection.

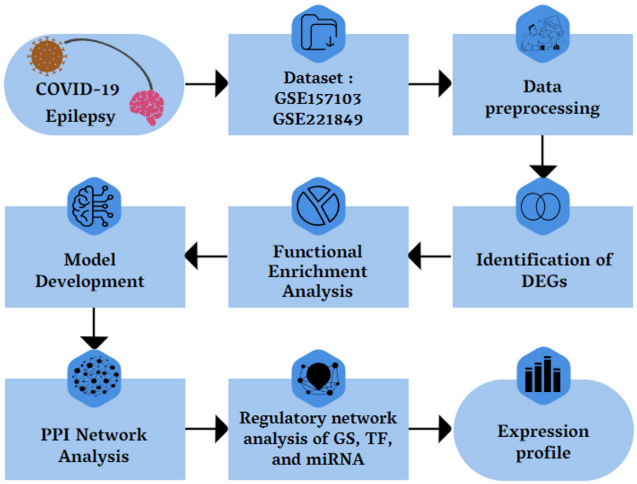

Figure 1. Study design. The study involved data collection, analysis, and validation to identify molecular pathways linking COVID-19 and epilepsy, aiming to understand the connection between these conditions.

2. Methodology

2.1. Datasets summary

Both the RNA sequencing and single-cell RNA sequencing datasets of COVID-19 and Epilepsy were obtained from the GEO database of the National Center for Biotechnology Information [16]. The COVID-19 dataset (GSE157103) and the epilepsy dataset (GSE221849) were selected from Homo sapiens as the ideal pairings to ensure the testing samples were consistent and the sample size was adequate for both disorders.

The GPL24676 Illumina NovaSeq 6000 (Homo sapiens) contained the COVID-19 dataset GSE157103, which was contributed by Overmyer et al. [17]. We classified 100 COVID-19 groups and 26 healthy controls using plasma and leukocyte samples from hospitalized patients with or without COVID-19. The seizure-affected Epilepsy dataset GSE221849 was from the GPL24676 Illumina NovaSeq 6000 (Homo sapiens) provided by Miller et al. [18]. Its samples are also from seizure-affected brain tissue with mosaic copy number gain of chromosome 1q, containing 5 treated groups (with chr1q gain) and 2 healthy controls (without chr1q gain). Both datasets were obtained by high-throughput sequencing. The COVID-19 dataset GSE157103 and the epilepsy dataset GSE221849 are used as validation cohorts. Table 1 shows the basic information for the datasets, with lots of details and numbers.

2.2. Data preprocessing

The data preprocessing is done for the dataset GSE221849, since it belongs to Sc-RNA-Seq, we need to extract count data from the different files we downloaded from the GEO Database. The preprocessing workflow commenced by utilizing the Seurat Package [19] to extract count data from the various files downloaded from the GEO database. This count data extraction process involved several steps within the Seurat framework to preprocess the data. Quality check metrics were applied following the methodology outlined by Ilicic et al. [20], ensuring the identification and removal of low-quality cells, empty droplets, or

Table 1. A description of the dataset's GEO parameters

Disease	GSE Id	GEO platform	Tissue type	Experiment type	Samples
COVID-19	GSE157103	GPL24676	Plasma and leukocyte	RNA-seq	126
Epilepsy	GSE221849	GPL24676	Brain tissues	scRNA-seq	7

potential cell multiplets based on criteria such as the number of unique genes detected, total molecule counts, and the percentage of reads mapping to mitochondrial genes. Normalization and Variance Stabilization were executed using the approach proposed by Hafemeister and Satija [21], aiming to standardize the expression measurements across cells and stabilize the variance, thereby enabling reliable downstream analysis. The dataset underwent a clustering procedure, adopting the methodology described by Xu and Su [22], which likely involved a graph-based clustering approach or similar strategies to group cells based on their similarities in gene expression profiles, thereby delineating distinct cell types or populations. Following these preprocessing steps, the finalized Seurat object contained the processed count data, which was extracted for further analysis, particularly for differential gene expression analysis. This extracted count file served as the foundation for subsequent investigations into gene expression patterns across different cell types or conditions within the dataset.

2.3. DEGs and the common DEGs of COVID-19 and epilepsy

Using the R (v4.1.1) Limma package, differential gene expression analysis (DEGs) was performed to investigate shared gene signatures between COVID-19 and epilepsy [23]. Using metrics such as adj. P value < 0.05 and $|\log_2(FC)| > 2$ as tweaks, we standardized screening criteria to find significant DEGs in both datasets. To illustrate the DEGs in COVID-19 and epilepsy, respectively, the ggplot2 software was used to create volcano plots and heatmaps [24]. Using a Venn diagram, common DEGs between the two conditions were identified.

2.4. Functional enrichment analysis

A database termed GO is used to specify the roles that genes and proteins play in different species [25]. On the other hand, the KEGG pathway is particularly important in genome analysis and gene annotation and is known for its function in controlling metabolism [26]. We performed both GO and KEGG analyses using the "ggplot2" and "Cluster Profiler" packages [27, 28]. Additionally, the same method has been applied to the shared DEGs between COVID-19 and epilepsy for GO analysis, which includes biological process (BP), cellular component (CC), and molecular function (MF). We did functional enrichment analysis to investigate the influencing factors of the differentially expressed genes (DEGs) in further detail.

2.5. Identification of the significant gene signatures between COVID-19 and epilepsy

LASSO, a widely used supervised learning method, is employed in gene signature, biomarker selection, and regression tasks [29]. We utilized a LASSO-based regression model, trained on DEGs, to identify discriminative gene signatures shared by COVID-19 and epilepsy from DEGs using the "glmnet" package [30]. Optimal parameters were chosen via a 10-fold cross-validation procedure. The coefficients of particular features are exposed by the lambda parameter. Genes exhibiting non-zero coefficients in the LASSO-based logistic regression model were deemed discriminative, while those with zero coefficients were excluded from further analysis. Also, the prediction accuracy of the model and these important gene signatures was evaluated using receiver operating characteristic (ROC) analysis, where the area under the curve (AUC) value was computed [31].

2.6. PPI network analysis

PPI network analysis was conducted to gain further insight into the interacting proteins linked to the significant gene signatures. The shared gene signatures were loaded into the Network Analyst 3.0 platform [32], and the "STRING Interactome" database was used, with a confidence score cutoff of 900, to build the PPI network. Afterward, Cytoscape Version 3.8.0 was needed for the network's visualization [33].

2.7. Regulatory networks analysis of gene signature, TF, and miRNA

The Network Analyst 3.0 platform (https://www.networkanalyst.ca) was utilized to do an extensive examination of gene-miRNA interaction networks, TF-gene interaction networks, and TF-miRNA coregulatory networks, among other systematic regulatory networks. Certain DNA sequences that control transcription and underpin many facets of human physiology, illness, and variability can be identified thanks in large part to TFs [34]. To analyze the TF-gene interaction network, we specifically used the ENCODE ChIP-seq database. The miRTarBase v8.0 database, which is acknowledged as the principal experimental database for miRNA-gene interactions and has the highest quantity of verified MTIs in comparison to other comparable databases, was utilized to analyze gene-miRNA interaction networks [35]. TF-miRNA coregulatory networks were built with data from the RegNetwork repository. Mature miRNAs control gene expression by binding to complementary sites on mRNAs through base pairing. The construction of the miRNA-gene interaction network using Network Analyst was done to pinpoint miRNAs with important regulatory functions in target genes. MiRTarBase and Tarbase were employed during the analysis procedure. Additionally, we assessed the biological functions and characteristics of the leading miRNAs and TFs using a degree filter.

2.8. Statistical analysis

The R software version 4.1.1 was utilized to execute all R packages mentioned in this research. A statistically significant result was declared when the P value was less than 0.05.

3. Results

3.1. Preprocessing of Sc-RNA-seq data

With dataset GSE221849, we conducted various preprocessing steps to guarantee data quality and make it ready for further analysis. In the first step, Seurat is used for exploring quality control (QC)

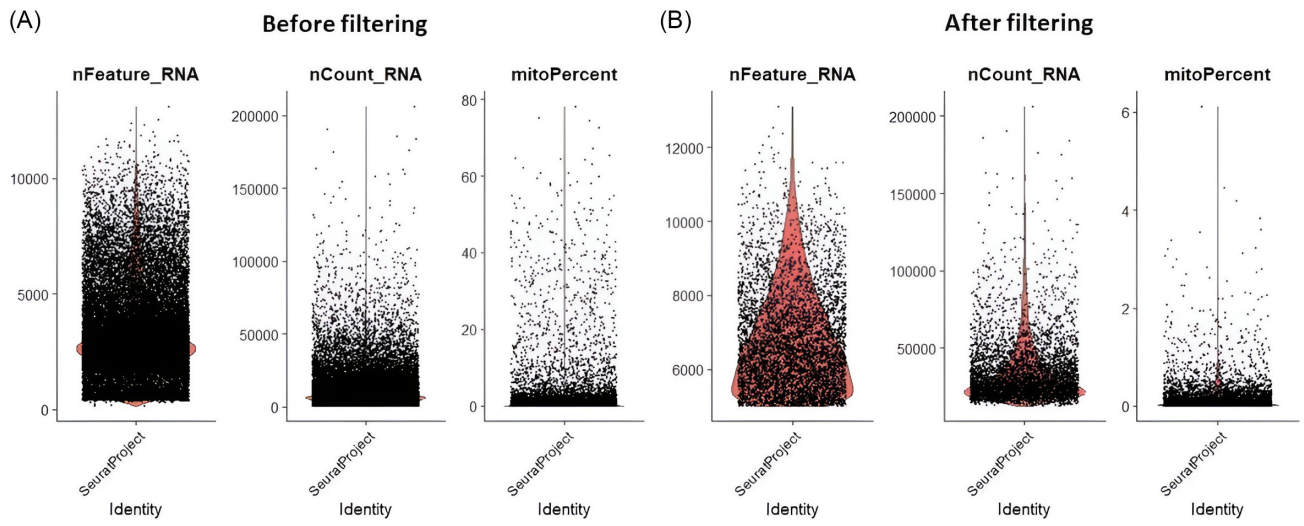

Figure 2. Data preprocessing of Sc-RNA-seq data (Epilepsy dataset). The following charts show the results of filtering count data of nFeature RNA, nCount RNA, and mitopercent to eliminate low count reads. Each dot in the plots represents a cell.

metrics and filtering cells, which is a common tool for analyzing single-cell RNA sequencing data. Evaluation is conducted on specific factors such as the number of distinct genes identified in each cell, overall molecule numbers, and the proportion of reads aligning with the mitochondrial genome. Cells displaying abnormalities like very low gene counts, high molecule counts indicating multiple cells, or increased mitochondrial contamination are detected and removed. The Seurat Object is created with metadata that automatically calculates the unique genes and total molecules. The metadata contains measurements like nFeature_RNA (quantity of genes identified in each cell) and nCount_RNA (overall number of molecules identified in a cell). After quality control, cells that do not meet user-defined criteria, such as having unique feature counts (genes) less than or equal to 5000 or less than 10% mitochondrial counts, are discarded (Figure 2).

Following QC, filtering of low-quality cells across 4750 samples within the Seurat object, the dataset undergoes normalization using the "Log Normalize" method, scaling feature expression measurements for each cell by total expression and log transformation, ensuring a standardized scale for downstream analyses. Highly variable features, as recommended in prior studies, are extracted, resulting in the selection of 2,000 features demonstrating substantial cell-to-cell variation. Subsequent linear scaling of the dataset is conducted, adjusting gene expression to equalize mean expression across cells and standardize variance. Principal component analysis (PCA) is performed to visualize the primary sources of heterogeneity using a heatmap, aiding in the selection of significant principal components for downstream analyses. Cells are clustered using a graph-based approach, constructing a K-nearest neighbor graph based on Euclidean distance in PCA space. The neighbors identify and modularity optimization techniques, such as the Louvain algorithm, iteratively group cells into clusters to identify distinct cell populations. Non-linear dimensional reduction techniques such as UMAP, are applied to visualize the dataset's manifold, positioning cells with similar expression profiles closer together in a lower-dimensional space (**Supplementary_File_4 (A)**).

Finally, the preprocessed dataset is saved, and the extraction of counts was used for differential gene expression analysis.

3.2. Detection of the shared DEGs between COVID-19 and epilepsy

To discern the connection between COVID-19 and epilepsy at the genetic level, we employed RNA-seq data from the GSE157103 dataset for COVID-19 and the GSE221849 dataset for epilepsy, utilizing limma for analysis. Employing a stringent filter criterion of an adjusted p-value < 0.05 and a $|\log_2 FC|$ value > 2.0, we identified 2402 DEGs in COVID-19 (44 upregulated, 2388 downregulated) (Figure 3A) and 27309 upregulated DEGs in epilepsy after preprocessing (Figure 3B). Hierarchical clustering visualizations depicted DEG expression patterns (Figure 3C and 3D). Through Venn analysis, 1040 common DEGs were identified, shedding light on shared molecular pathways between COVID-19 and epilepsy. Further details on DEGs can be found in **Supplementary_file_1**.

3.3. Gene enrichment analysis

The analysis delved into the 1040 common DEGs, conducting GO enrichment analysis to unveil their biological significance. This analysis revealed the top 5 enriched pathways in BP, CC, and MF categories, visualized through a Sunburst graph, illustrating hierarchical relationships among these pathways. Additionally, KEGG pathway enrichment analysis was executed on the same DEGs, identifying the top 10 significantly enriched KEGG pathways (**Supplementary_file_4 (B)**). Together, these analyses offer insights into the functional roles and interactions among genes, shedding light on the biological mechanisms underpinning the observed gene expression changes in the studied conditions.

3.4. Finding the key gene signatures between COVID-19 and epilepsy

Utilizing the LASSO algorithm on the pool of 1040 common DEGs, we identified 10 key gene signatures (CATIP, CDC25C, GPR132, NTS, PDE8B, PLK1, SLC12A9, SPC25, TUBA1A, TYMS) with significant implications for both COVID-19 and epilepsy (Figure 4A, B). The model score, determined by the coefficients of these gene signatures, was calculated by summing

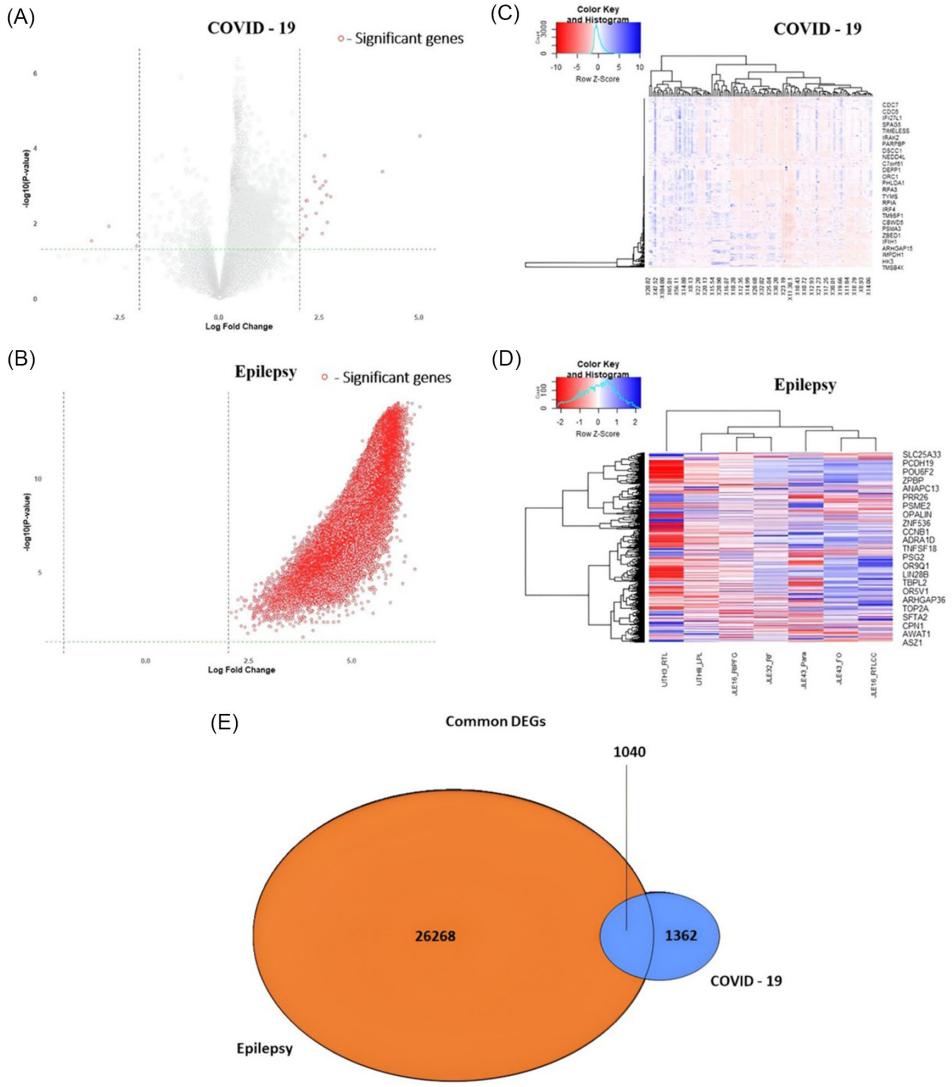

Figure 3. Differential expression analysis using Limma. Volcano plots of DEGs are shown in (A, B). The plot of the heatmap displays the important genes that are present between (C) COVID-19 and (D) epilepsy (C, D). The genes shared by COVID-19 and epilepsy are depicted in the Venn diagram (E).

each gene's value multiplied by its corresponding coefficient. Subsequently, we assessed the diagnostic efficacy of these key gene signatures, revealing their predictive potential for COVID-19 and epilepsy through ROC analysis (Table 2).

Interestingly, the model developed using 10 important gene signatures highlighted the potential of diagnostic efficacy in both epilepsy and COVID-19 (AUC of 0.911, Figure 4E) (AUC of 0.917, Figure 4F).

3.5. Detection of PPI network

Using the STRING interactome model within the Network Analyst 3.0 platform, we constructed PPI networks. **Supplementary_file_4 (C)**, showcases three distinct PPI networks, with one network containing 244 nodes and 253 edges, visualized using the large graph layout and highlighting seed nodes. Within this network, seven key genes were identified. Further details regarding the nodes in the PPI network can be found in **Table (Supplementary_file_2)**

3.6. Transcription factor, miRNA, and gene signature analysis in regulatory networks

In order to investigate the connection between gene signatures and TFs, we started a study using the Network Analyst 3.0 platform (TFs). DisGeNET's regulatory Explorer highlights three TF genes linked to epilepsy in **Supplementary_file_4 (D)**, which shows a network of TF-gene interactions. With 146 nodes and 190 edges with 7 seeds, the regulatory network analysis most notably demonstrated a strong link between TF genes and important gene signatures. Particular connections were seen between TF genes EZHZ, ESRRA, and MYC and PDE8B, GPR132, and SLC12A9. **Supplementary_file_3** contains the detailed information about the regulatory network analysis of gene signatures, TF genes, and miRNA interaction networks.

We also investigated the regulation network of microRNAs and gene signatures (miRNAs). Despite the creation of a network, single gene-miRNA networks were constructed **Supplementary_file_4 (E)**, comprising 120 nodes and 120 edges. Unfortunately, no connections

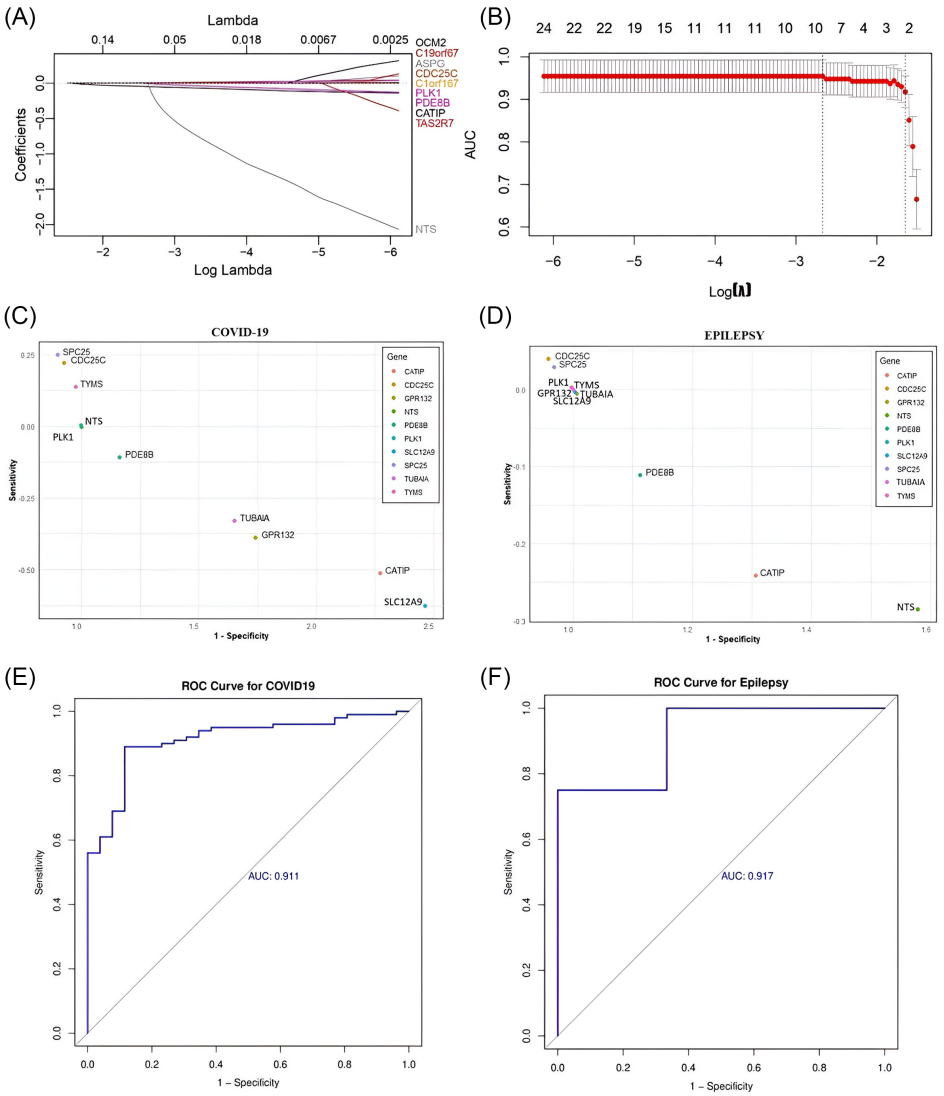

Figure 4. Gene selection. Using LASSO coefficients (A & B), AUC values (C & D), and ROC curves (E & F) to demonstrate the predictive capacity of the 10 gene signatures, the figure depicts the inquiry into the relationship between COVID-19 and epilepsy.

Table 2. AUC values for key gene signatures. This table summarizes the diagnostic efficiency (Area under the curve, AUC) of key gene signatures in COVID-19 and epilepsy. The values represent the AUC for each gene signature in the respective conditions

	Area under the curve (AUC)	
Gene	COVID-19	Epilepsy
CATIP	0.8713	0.6666
CDC25C	0.8876	0.6666
GPR132	0.9480	0.8333
NTS	0.505	1
PDE8B	0.7042	0.5833
PLK1	0.8592	0.9166
SLC12A9	0.9357	0.75
SPC25	0.8588	0.75
TUBAIA	0.8942	0.5
TYMS	0.9111	0.5

between miRNAs and genes were found. This is probably because there haven't been many studies done on epilepsy in relation to COVID-19 infection. To clarify the mechanics and possible therapeutic applications of these interactions, more investigation is essential.

Using the Network Analyst 3.0 software, we then performed TF-miRNA coregulatory network analysis to look at the relationship between miRNAs, TFs, and important gene signatures. The TF-miRNA coregulatory network is illustrated in Figure 5, which also features three TF genes such as PLK1, BRCA2, and TP53, that were discovered via interactions with seed genes. Blue hexagons show miRNAs associated with gene signatures, while yellow hexagons indicate possible gene signatures.

TP53, PLK1, BRCA2 from the TF-miRNA coregulatory network, ESRRA, EZH2, and MYC from the TF-gene interaction network, and CDC25C, PDE8B, SLC12A9, SPC25, YWHAE, SMC3, PLK1, and TYMS from the PPI network were among the significant gene signatures found within the various networks after

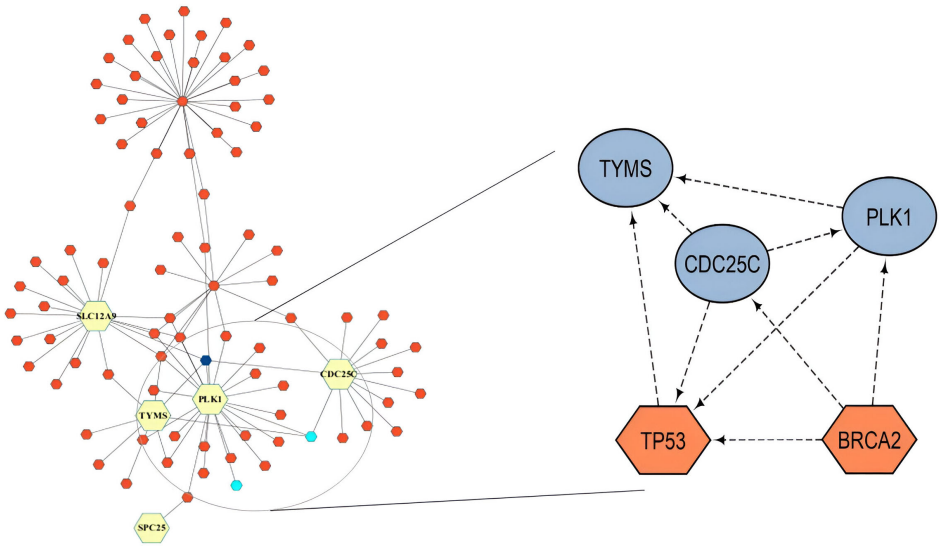

Figure 5. TF-miRNA coregulatory network. This figure shows the formation of a TF-miRNA coregulatory network, with candidate gene signatures represented by yellow hexagons, miRNA signatures by blue hexagons, and novel gene signatures by blue circles.

analysis. Notably, PLK1, TYMS, and CDC25C were among the genes that were discovered in many networks, which may indicate that these genes play important roles in shared pathways between conditions like COVID-19 and epilepsy.

Even though we used well-established databases with strict curation processes, it's necessary to admit that expected interactions might still come with restrictions requiring experimental confirmation. Consequently, while our examination offers useful insights into gene-miRNA and TF-gene connections, care should be taken when interpreting these outcomes as absolute proof of biological interactions. We plan to be transparent about the methodology and data sources employed for regulatory network analysis to aid in understanding and validation by the scientific society, allowing readers to evaluate the dependability of our discoveries in the broader literary context. Furthermore, addressing the possible limitations of expected interactions will put the outcomes into perspective and grant a more detailed grasp of the regulatory networks deduced in our investigation.

In addition to this, we compared the mRNA expression levels of our seed genes CDC25C, PLK1, and TYMS in COVID-19 and Epilepsy. The box plot evident that mRNA expression levels of CDC25C and TYMS were shown as higher in COVID-19 patients also samples with epilepsy patients who got recurrent seizures, PLK1 shows higher expression in COVID-19 patients, whereas expression is similar in both cases of Epilepsy patients with or without seizures (Figure 6 (A–F)).

4. Discussion

In our study, we utilized bioinformatics methods similar to those outlined in [36], which focused on COVID-19 and acute kidney injury based on the shared gene signatures and regulatory network. As we delved into understanding relationship between COVID-19 and epilepsy, there COVID-19 has presented significant challenges for individuals with pre-existing epilepsy, manifesting in altered seizure frequencies independent of COVID-19 infection. The potential exacerbation of seizures in COVID-19 patients, particularly due to fever-induced seizures, underscores the intricate relationship between the virus and neurological conditions like epilepsy [37]. Additionally, severe COVID-19 illness has been linked to complications such as hypoxic encephalopathy and cytokine storms, further elevating the risk of seizures in epilepsy patients. These challenges are compounded by barriers to accessing medications and healthcare services, exacerbating the frequency and severity of seizures in affected individuals.

The study explores the genetic underpinnings and biological mechanisms of COVID-19 and epilepsy, offering insights into potential overlapping pathways and therapeutic targets. It focuses on shared gene signatures and regulatory networks, obtaining shared DEGs and determining representative gene signatures using machine learning. The model has high predictive efficacy in both conditions.

In this study, 1040 DEGs were identified with the criteria of FDR < 0.05 and $|\log2\ FC| > 2$ in both diseases. As a result of the functional annotation obtained by the R package cluster Profiler, the GO enrichment analysis revealed that the 1040 DEGs were mainly associated with "homophilic cell adhesion via plasma membrane adhesion molecules", "cell-cell adhesion via plasma membrane adhesion molecules", "meiotic cell cycle", "meiotic cell cycle'' etc. The KEGG pathway analysis showed "Reductive pentose phosphate cycle (Calvin cycle)" and "Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate" pathways were significantly enriched, which suggests that these genes may be involved in the action process or metabolic reaction of drugs. This indicates COVID-19 infection may exhibit dysregulated neurological changes which affect the brain mechanism. With the LASSO logistic regression model, 10 shared key genes (CATIP, CDC25C, GPR132, NTS, PDE8B, PLK1, SLC12A9, SPC25, TUBAIA, TYMS) were identified and could effectively distinguish between COVID-19 and Epilepsy.

Particularly, three of the gene signatures such as CDC25C, PLK1, and TYMS are significant genes in both the COVID-19 and Epilepsy networks, indicating their critical roles in tying together regulatory mechanisms. The results suggested that these candidate gene signatures might help predict a person's risk of having seizures. By understanding the neurological manifestations, especially in the case of Epilepsy, and guiding the appropriate treatment for COVID-19 patients.

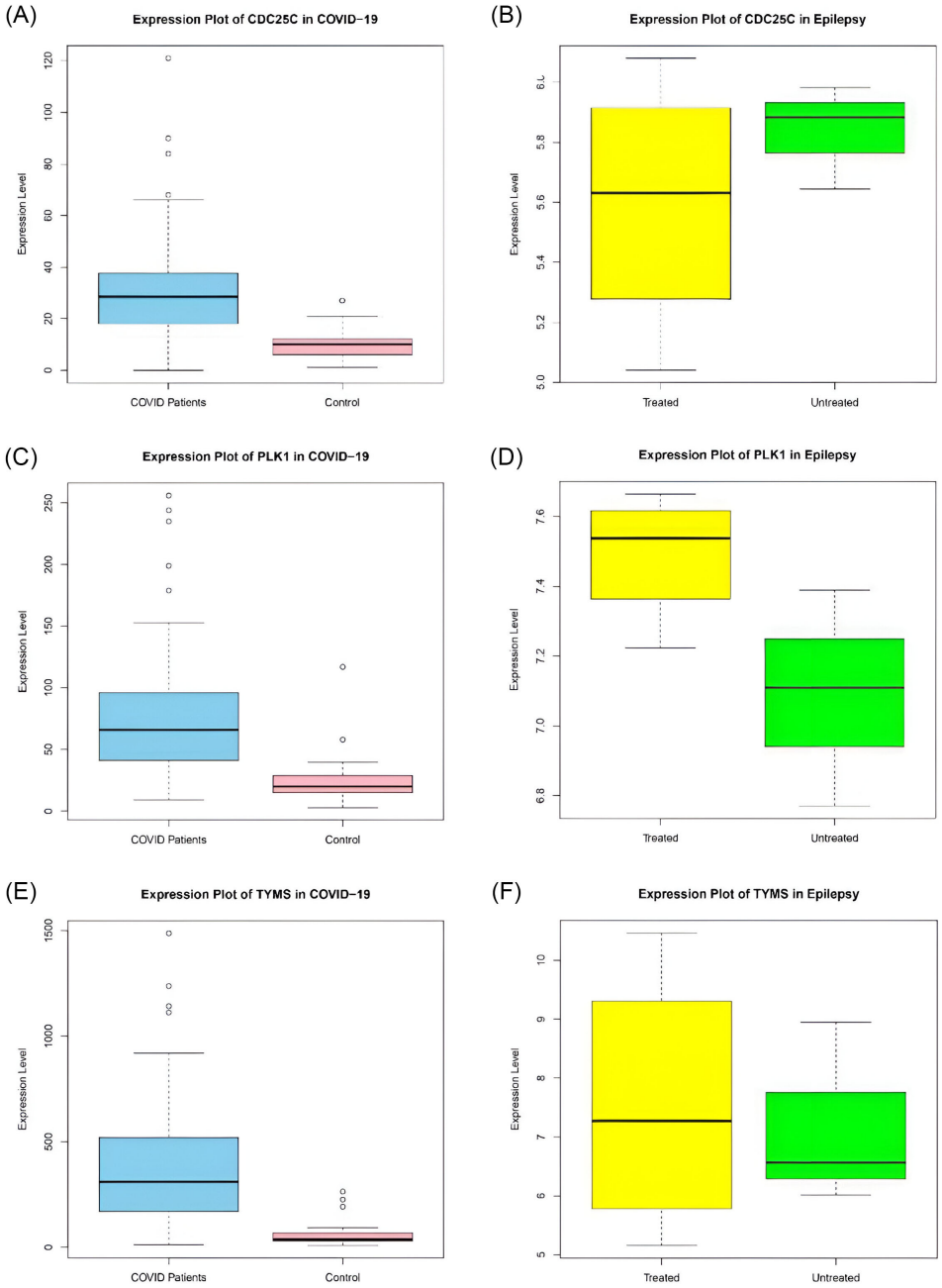

Figure 6. Expression boxplot. A boxplot depicting the expression of COVID-19 and epilepsy, along with shared gene signatures (A, B) CDC25C, (C, D) PLK1, and (E, F) TYMS expression in brain samples affected with seizures (treated, yellow box), brain samples not affected with seizures (untreated, green box), and COVID patients (blue box) and controls (pink box), respectively.

The dysregulation of cell cycle proteins, such as PLK1, CDC25C, and TYMS, has been implicated in epilepsy. PLK1 has been found to be increased in the brains of scrapie-infected hamsters [38]. CDC25C, a cell cycle regulatory protein, plays a role in G2/M progression and DNA damage repair, and changes in its expression have been linked to tumorigenesis and tumor development [39]. Additionally, aberrant cell cycle activity and DNA damage have been observed in neurons in neurodegenerative diseases, including epilepsy [40]. The dysregulation of cell cycle and DNA repair processes may contribute to neuronal vulnerability in epilepsy, potentially leading to cell cycle arrest and apoptosis [41].

While [42] elucidated molecular connections between COVID-19 and epilepsy by identifying 373 common genes primarily involved in immune response processes, our study differs in approach and findings. The previous study focused on identifying diagnostic candidates and immune cell correlations, whereas our work extends beyond to encompass broader aspects of genetic interactions, regulatory networks, and pathway analyses. While we also identified common genes between COVID-19 and epilepsy, our emphasis lies on exploring regulatory networks and understanding the underlying biological mechanisms driving these connections. Additionally, our study delves into gene-miRNA and TF-gene interaction networks, shedding light on regulatory

processes underlying the observed gene expression changes in both conditions.

Furthermore, this study explores the potential neurological manifestations of COVID-19, including the possibility of the virus causing direct or indirect effects on the CNS, which could potentially lead to seizures or other neurological complications in affected individuals. It also addresses the need for further research to understand the long-term neurological impact of COVID-19, particularly in individuals with epilepsy.

5. Conclusion

In the investigation, gene signatures that imply global gene regulatory networks were determined in COVID-19 patients with epilepsy as a comorbidity, providing a possible avenue for intervention. Although the study captured data from both conditions, limitations remain due to insufficient sample sizes. The suggested further research should verify the sample sizes and assess clinical importance when considering predictive aspects or therapeutic opportunities connected to COVID-19 disease.

The disruptions of CDC25C, PLK1, and TYMS gene expressions may relate to epilepsy from COVID-19, but the exact pathogenic role and clinical consequences need further exploration. This work extends our comprehension of the interaction between COVID-19 and epilepsy that goes beyond neuroinflammation at the molecular level, contributing new insights into therapeutic strategies by identifying specific molecular targets. However, the extent of this impact needs further studies into the management of seizures and epilepsy during the COVID-19 pandemic. An in-depth study calls for future research on how COVID-19 is capable of influencing epilepsy.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in GEO database of the National Center for Biotechnology Information at COVID-19 dataset: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157103 and Epilepsy dataset: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE221849

Author Contribution Statement

Pavithra Nagendran: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Gowtham Murugesan:** Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Jeyakumar Natarajan:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration.

Supplementary Information

The supplementary files are available at https://doi.org/10.47852/bonviewMEDIN42022621

References

[1] Yacubian, E. M. T., Kakooza-Mwesige, A., Singh, G., Carpio, A., de Figueiredo, N. V., Lutzky Saute, R., & Marques de Haes, T. (2022). Common infectious and parasitic diseases as a cause of seizures: Geographic distribution and contribution to the burden of epilepsy. *Epileptic Disorders*, *24*(6), 994–1019. https://doi.org/10.1684/epd.2022.1491

[2] Bustos, J., Gonzales, I., Saavedra, H., Handali, S., Garcia, H. H., & Cysticercosis Working Group in Peru. (2021). Neurocysticercosis. A frequent cause of seizures, epilepsy, and other neurological morbidity in most of the world. *Journal of the Neurological Sciences*, *427*, 117527. https://doi.org/10.1016/j.jns.2021.117527

[3] Skvortsov, V. V., Aleksandrovna, M. E., Ivanovich, M. G., Ponomareva, A. R., Shamilyevna, T. R., Valeryevna, K. A., & Viktorovna, L. A. (2023). Neurological manifestations of long COVID: Clinical aspects and impact on quality of life (in Russian). *Bulletin of Neurology, Psychiatry and Neurosurgery*. 264–269. https://doi.org/10.33920/med-01-2304-03

[4] Sonneville, R., Dangayach, N. S., & Newcombe, V. (2023). Neurological complications of critically ill COVID-19 patients. *Current Opinion in Critical Care*, *29*(2), 61–67. https://doi.org/10.1097/MCC.0000000000001029

[5] Siahaan, Y. M. T., Ketaren, R. J., Hartoyo, V., & Hariyanto, T. I. (2021). Epilepsy and the risk of severe coronavirus disease 2019 outcomes: A systematic review, meta-analysis, and meta-regression. *Epilepsy & Behavior*, *125*, 108437. https://doi.org/10.1016/j.yebeh.2021.108437

[6] Hou, J., Bhat, A. M., Ahmad, S., Raza, K., & Qazi, S. (2022). *In silico* analysis of ACE2 receptor to find potential herbal drugs in COVID-19 associated neurological dysfunctions. *Natural Product Communications*, *17*(8), 1934578X221118549. https://doi.org/10.1177/1934578X221118549

[7] Mesmoudi, S., Lapina, C., Rodic, M., & Peschanski, D. (2022). Multi-data integration towards a global understanding of the neurological impact of human brain severe acute respiratory syndrome coronavirus 2 infection. *Frontiers in Integrative Neuroscience*, *16*, 756604. https://doi.org/10.3389/fnint.2022.756604

[8] Khatoon, F., Prasad, K., & Kumar, V. (2022). COVID-19 associated nervous system manifestations. *Sleep Medicine*, *91*, 231–236. https://doi.org/10.1016/j.sleep.2021.07.005

[9] Vlaicu, S. I., Tatomir, A., Cuevas, J., Rus, V., & Rus, H. (2023). COVID, complement, and the brain. *Frontiers in Immunology*, *14*, 1216457. https://doi.org/10.3389/fimmu.2023.1216457

[10] Goldenberg, D., & Dichter, M. (2022). COVID-19 direct effects on the central nervous system. In D. Goldenberg & M. Dichter (Eds.), *Unravelling long COVID* (pp. 39–54). John Wiley & Sons Ltd. https://doi.org/10.1002/9781119891338.ch3

[11] Dunn, E., Ibrahim, F., Neme-Mercante, S., Nair, D., Morris, S., Wang, I., & Punia, V. (2023). Possible post-COVID epilepsy: A review of epilepsy monitoring unit admissions during the two years of COVID-19 pandemic. *Epilepsy & Behavior Reports*, *21*, 100584. https://doi.org/10.1016/j.ebr.2023.100584

[12] Sorokina, N. D., Pertsov, S. S., Selitsky, G. V., Zherdeva, A. S., & Tsagashek, A. V. (2023). Impact of COVID-19 on some neurological and psychological features in epilepsy. *Zhurnal Nevrologii i Psikhiatrii Imeni SS Korsakova*, *123*(3), 7–11. https://doi.org/10.17116/jnevro20231230317

[13] Musoni, L., Ezzouine, H., Etouki, O., Habibou, R., Nour, M., ElKhaouri, I., & Charra, B. (2021). Covid-19 decompensating epilepsy in the elderly: A case report. *Annals of Medicine and Surgery*, *68*, 102642. https://doi.org/10.1016/j.amsu.2021.102642

[14] Dewanjee, S., Vallamkondu, J., Kalra, R. S., Puvvada, N., Kandimalla, R., & Reddy, P. H. (2021). Emerging COVID-19 neurological manifestations: Present outlook and potential neurological challenges in COVID-19 pandemic. *Molecular Neurobiology*, *58*(9), 4694–4715. https://doi.org/10.1007/s12035-021-02450-6

[15] Park, S., Majoka, H., & Ajaz Sheikh, I. A. (2021). A case report of delayed manifestation of new-onset focal seizures presenting after COVID infection. *Epilepsy & Behavior Reports*, *16*, 100447 https://doi.org/10.1016/j.ebr.2021.100447

[16] Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., . . . , & Soboleva, A. (2012). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Research*, *41*(D1), D991–D995. https://doi.org/10.1093/nar/gks1193

[17] Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., . . . , & Jaitovich, A. (2021). Large-scale multi-omic analysis of COVID-19 severity. *Cell Systems*, *12*(1), 23–40. https://doi.org/10.1016/j.cels.2020.10.003

[18] Miller, K. E., Rivaldi, A. C., Shinagawa, N., Sran, S., Navarro, J. B., Westfall, J. J., . . . , & Bedrosian, T. A. (2023). Post-zygotic rescue of meiotic errors causes brain mosaicism and focal epilepsy. *Nature Genetics*, *55*(11), 1920–1928. https://doi.org/10.1038/s41588-023-01547-z

[19] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., . . . , & Satija, R. (2019). Comprehensive integration of single-cell data. *Cell*, *177*(7), 1888–1902. https://doi.org/10.1016/j.cell.2019.05.031

[20] Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., & Teichmann, S. A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biology*, *17*, 1–15. https://doi.org/10.1186/s13059-016-0888-1

[21] Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*, *20*(1), 296. https://doi.org/10.1186/s13059-019-1874-1

[22] Xu, C., & Su, Z. (2015). Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics*, *31*(12), 1974–1980. https://doi.org/10.1093/bioinformatics/btv088

[23] Ritchie, M. E., Phipson, B., Wu, D. I., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

[24] Chen, T., Zhang, H., Liu, Y., Liu, Y. X., & Huang, L. (2021). EVenn: Easy to create repeatable and editable Venn diagrams and Venn networks online. *Journal of genetics and genomics= Yi chuan xue bao*, *48*(9), 863–866. https://doi.org/10.1016/j.jgg.2021.07.007

[25] Gene Ontology Consortium. (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*(suppl_1), D258–D261. https://doi.org/10.1093/nar/gkh036

[26] Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

[27] Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. USA: Springer International Publishing. ISBN: 978-3-319-24275-0. https://doi.org/10.1007/978-3-319-24277-4

[28] Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

[29] Swain, S. M., Tang, G., Brauer, H. A., Goerlitz, D. S., Lucas, P. C., Robidoux, A., . . . , & Wolmark, N. (2020). NSABP B-41, a randomized neoadjuvant trial: Genes and signatures associated with pathologic complete response. *Clinical Cancer Research*, *26*(16), 4233–4241. https://doi.org/10.1158/1078-0432.CCR-20-0152

[30] Zhang, Y., Wilson, R., Heiss, J., Breitling, L. P., Saum, K. U., Schöttker, B., . . . , & Brenner, H. (2017). DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nature Communications*, *8*(1), 14617. https://doi.org/10.1038/ncomms14617

[31] Craig, J. E., Han, X., Qassim, A., Hassall, M., Cooke Bailey, J. N., Kinzy, T. G., . . . , & MacGregor, S. (2020). Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. *Nature Genetics*, *52*(2), 160–166. https://doi.org/10.1038/s41588-019-0556-y

[32] Zhou, G., Soufan, O., Ewald, J., Hancock, R. E., Basu, N., & Xia, J. (2019). NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Research*, *47*(W1), W234–W241. https://doi.org/10.1093/nar/gkz240

[33] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . , & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303

[34] Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., . . . , & Weirauch, M. T. (2018). The human transcription factors. *Cell*, *172*(4), 650–665. https://doi.org/10.1016/j.cell.2018.01.029

[35] Hsu, S. D., Lin, F. M., Wu, W. Y., Liang, C., Huang, W. C., Chan, W. L., . . . , & Huang, H. D. (2011). miRTarBase: A database curates experimentally validated microRNA–target interactions. *Nucleic Acids Research*, *39*(suppl_1), D163–D169. https://doi.org/10.1093/nar/gkq1107

[36] Zhou, X., Wang, N., Liu, W., Chen, R., Yang, G., & Yu, H. (2023). Identification of the potential association between SARS-CoV-2 infection and acute kidney injury based on the shared gene signatures and regulatory network. *BMC Infectious Diseases*, *23*(1), 655. https://doi.org/10.1186/s12879-023-08638-6

[37] Wang, G., Bai, L., Zhao, M., & Wang, S. (2022). Global landscape of COVID-19 and epilepsy research: A bibliometric analysis. *Frontiers in Neurology*, *13*, 1034070. https://doi.org/10.3389/fneur.2022.1034070

[38] Liu, K., Zheng, M., Lu, R., Du, J., Zhao, Q., Li, Z., . . . , & Zhang, S. (2020). The role of CDC25C in cell cycle regulation and clinical cancer therapy: A systematic review. *Cancer Cell International*, *20*, 1–16. https://doi.org/10.1186/s12935-020-01304-w

[39] Wang, H., Tian, C., Xu, Y., Xie, W. L., Zhang, J., Zhang, B. Y., . . . , & Dong, X. P. (2013). Abortive cell cycle events in the brains of scrapie-infected hamsters with remarkable decreases of PLK3/Cdc25C and increases of PLK1/cyclin

B1. *Molecular Neurobiology*, *48*, 655–668. https://doi.org/10.1007/s12035-013-8455-1

[40] Zhang, X., Song, S., & Peng, W. (2021). Cell cycle deregulation in neurodegenerative diseases. *International Journal of Neuroscience*, *133*(4), 408–416. https://doi.org/10.1080/00207454.2021.1924705

[41] Kim, D., & Tsai, L. H. (2009). Linking cell cycle reentry and DNA damage in neurodegeneration. *Annals of the New York Academy of Sciences*, *1170*(1), 674–679. https://doi.org/10.1111/j.1749-6632.2009.04105.x

[42] Liu, Y. Z., Jiang, T., & Jiang, C. L. (2023). Bioinformatic analysis identified common pathogenetic processes between epilepsy and COVID-19. *European Review for Medical & Pharmacological Sciences*, *27*(24), 11673–11690. https://doi.org/10.26355/eurrev_202312_34765

How to Cite: Nagendran, P., Murugesan, G., & Natarajan, J. (2024). Unraveling the Shared Genetic Elements Among COVID-19 and Epilepsy: A Computational and Bioinformatics Analysis. *Medinformatics*. https://doi.org/10.47852/bonviewMEDIN42022621