Heliyon 10 (2024) e39039

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identification of potential biomarkers and drug of ischemic stroke in patients with COVID-19 through machine learning

Sixian Wang [a,c], Yuxing Tai [a,c], Xiaoqian Yang [a,b], Peizhe Li [a,c], Han Wang [a,c], Yi Tan [a,c], Tianjiao Gao [a,c], Mingrui Chu [a,c], Mingjun Liu [a,c,*]

[a] Changchun University of Chinese Medicine, Changchun 130117, China
[b] Jilin Ginseng Academy, Changchun University of Chinese Medicine, Changchun 130117, China
[c] College of Acupuncture and Massage, Changchun University of Chinese Medicine, Changchun 130117, China

ARTICLE INFO

Keywords:
Machine learning
COVID-19
Ischemic stroke
Biomarkers
Gene expression
Bioinformatics

ABSTRACT

The relationship between COVID-19 and ischemic stroke (IS) has attracted significant attention, yet the precise mechanism at the gene level remains unclear. This study aims to reveal potential biomarkers and drugs for COVID-19-related IS through bioinformatics methods. We collected two gene expression profiling datasets, GSE16561 and GSE213313, and selected GSE179879 and GSE196822 as validation sets for analysis. Through analysis, we identified 77 differentially expressed genes (DEGs) shared between COVID-19 and IS. Further gene enrichment analysis revealed that these genes are primarily involved in immune regulation. By constructing a protein-protein interaction network, we screened out nine hub genes, including FCGR3A, KLRB1, IL2RB, CD2, IL7R, CCR7, CD3D, GZMK, and ITK. In LASSO regression analysis, we evaluated the ROC curve's area under the curve (AUC) scores of key genes to assess their diagnostic accuracy. Subsequently, we performed random forest (RF), Support Vector Machine Recursive Feature Elimination (SVM-RFE), and neural network construction on hub genes to ensure accurate diagnosis of IS. Finally, by intersecting the results of three algorithms (LASSO regression, random forest, and SVM), CD3D and ITK were identified as the ultimate key genes. Based on this, we predicted potential targeted drug Blinatumomab. These research findings provide clues for a deeper understanding of the biological mechanisms of COVID-19-related IS and offer new insights for exploring novel treatment approaches.

1. Background

COVID-19 is an emerging infectious disease due to SARS-CoV-2 infection. Since 2019, it has become a global pandemic, affecting more than 700 million people globally. The probability of sequelae after infection with SARS-CoV-2 ranges between 9 % and 63 %, six times higher than similar viral infections. The sequelae involve fatigue, difficulty breathing, cognitive impairment/brain fog, discomfort post-exercise, memory problems, cough, sleep disturbances, tachycardia/palpitations, smell/taste changes, headache, chest pain, and muscle pain/cramps [1,2]. Relevant studies have indicated that human blood is hypercoagulable after SARS-CoV-2 infection [3].Therefore, the thrombosis probability in COVID-19 patients increases, elevating the risk of ischemic stroke (IS) in such patients. IS can cause various neurological sequelae, such as involuntary motor and cognitive dysfunction, with associated

* Corresponding author. Department of Acupuncture and Tuina, Changchun University of Chinese Medicine, Changchun 130117, China.
E-mail address: LiuMJ_1@163.com (M. Liu).

https://doi.org/10.1016/j.heliyon.2024.e39039
Received 6 June 2024; Received in revised form 19 September 2024; Accepted 5 October 2024
Available online 11 October 2024

disabilities enhancing the social, economic, and occupational burden [4]. Epidemiological statistics indicate that stroke, characterized by high disability and high fatality rates, is the second leading mortality cause worldwide and one of the most important reasons for death in China [5–9].Among stroke patients, the prevalence of IS was as high as 62 % in 2019 [10]. Acute ischemic stroke (AIS) mortality ranges from less than 5 % at one month to 10 % at three months and 15 % at one year [10–18]. IS occurs due to blood flow interruption in the cerebral arteries from various factors. Reversible tissue function loss occurs during insufficient blood supply to the brain tissue. Then, the loss of neurons and supporting structures occurs after a long enough period of infarction. The leading causes of IS include atherosclerosis, thrombus, and embolus [19]. Atherosclerosis results from complex cellular interactions between smooth muscle, endothelial and immune cells, and plaque buildup caused by atherosclerosis can restrict blood flow, which may lead to stroke [20]. Thrombosis results from activating the clotting cascade, causing platelets to aggregate and the prothrombin to become a fibrin clot. The embolus differs from thrombus in that it is thrombus caused in another site, eventually affecting the blood vessel through movement. Studies have indicated that SARS-CoV-2 is related to hypercoagulability [3]. Therefore, COVID-19 patients had increased blood vessel embolism and stroke risk. Early definitive diagnosis and treatment are the primary means to decrease the disability rate and mortality of IS. In the context of the COVID-19 epidemic, many patients have been diagnosed with thrombotic complications and IS. Stroke incidence in hospitalized patients ranges from 0.1 % to 6.9 %, and 79 % of COVID-19 patients have IS [21]. These data are limited by imaging means and insufficient diagnosis in sedation and mechanical ventilation patients [22]. However, during the historical period without COVID-19, the incidence of IS was about 62.4 %, and the number increased significantly. Around 51 % of these hospitalized patients with COVID-19 with a stroke had severe disabilities after hospital discharge. Moreover, the mortality rate was significantly higher than the control group without COVID-19. Currently, stroke is diagnosed primarily through neuroimaging diagnoses, such as CT and MRI. Neuroimaging is incapable of quickly determining the diagnosis of IS during COVID-19. It is increasingly utilized in routine physical examinations across large urban hospitals. However, the increase in testing of asymptomatic patients could lead to stroke overdiagnosis and make doctors uncertain about prescribing secondary prevention drugs[7]. Therefore, a diagnostic biomarker for quickly and accurately diagnosing stroke can improve the diagnostic accuracy of IS in COVID-19 patients and reduce the associated disability and mortality.

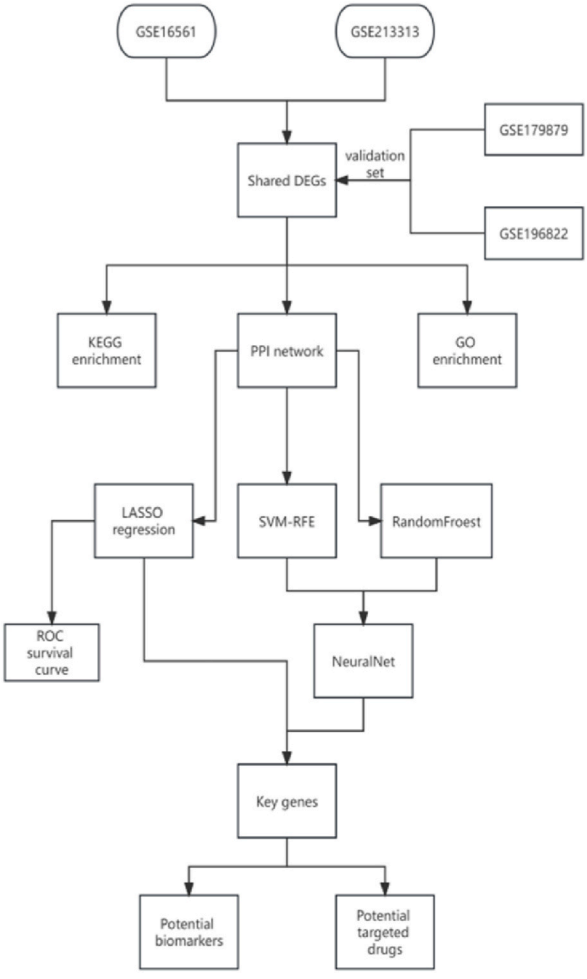

Fig. 1. The analysis processes.

With microarray technology advancement, bioinformatics is an interdisciplinary approach enabling researchers to uncover disease pathogenesis at the genetic and transcriptional levels. Based on the molecular biology analysis, the main IS mechanisms include excitotoxicity and calcium overload, oxidative stress (OS), and inflammation [19]. Among them, excitotoxicity, calcium overload, and OS are associated with glutamate receptors NMDA and AMPA. Inflammation can destroy the blood-brain barrier in the early stage and may affect nerve recovery later, exacerbating nerve damage. Additionally, Na/Ca2+ exchangers and acid-sensitive ion channels can behave as potential targets for IS injury [23,24]. The autopsy of COVID-19 patients has validated the presence of SARS-CoV-2 in the brain [25]. Thus, COVID-19 leads to systemic infectivity and can attack the nervous system, leading to brain damage and neurological dysfunction. Since the concept of Artificial Intelligence (AI) was introduced in 1965, machine learning has been widely used in the medical field through continuous research and development [26]. Our study applied integrated bioinformatics and machine learning methods to determine the thrombosis mechanism, reveal molecular regulatory networks, and identify shared critical genes associated with the pathogenesis of IS and COVID-19. This could provide a new perspective on the biological mechanism of IS associated with COVID-19. Then, the shared key genes selected by the machine learning method formed the basis to identify corresponding targeted drugs to prevent and treat patients with increased stroke risk symptoms related to COVID-19.

2. Data and methods

2.1. Dataset preparation

The flow design analysis of this study is represented in Fig. 1. Related datasets GSE16561 [27] and GSE213313 [28] were obtained from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). GSE16561 is a whole-blood mRNA from 34 patients with IS and 24 healthy individuals. GSE213313 is a whole-blood mRNA-related dataset of COVID-19 patients, with 19 critically and 15 non-critically sick patients and 11 healthy people. Then the log2 transform was performed, with the "normalize Between Arrays" function from the R package "limma"to normalize the raw count expression data. Then, the batch effect of GSE16561 and GSE213313 was eliminated by the R package "sva" [29] feature "ComBat."

Simultaneously, the datasets GSE179879 [30] and GSE196822 [31] from GEO were chosen as validation sets to confirm the differential expression of genes identified in GSE16561 and GSE213313. This verification aims to ensure that the observed differences in gene expression persist across other datasets. We utilized publicly available data from the Gene Expression Omnibus (GEO) database, and all data sources are appropriately cited in accordance with the GEO data usage policy.

2.1.1. Identifying shared differentially expressed genes (DEGs) between COVID-19 and stroke

R-packet "limma" helped screen DEGs between COVID-19 and IS samples and control groups. The cut-off standard for DEGs was $|\log2\text{Fold-Change (log2FC)}|>0.5$, and $p < 0.05$ after adjustment. This threshold was chosen to allow for a comprehensive exploration of potential DEGs in this study. By using a relatively lower threshold, a greater number of genes with notable changes can be included, thereby enriching the dataset for subsequent analyses.

2.1.2. Gene enrichment analysis

GO enrichment analysis annotates the genetic information based on three aspects: MF, BP, and CC. It is widely used in mining biological function information and the corresponding biological mechanism in microarray results [32]. KEGG (https://www.genome.jp/kegg/) is a comprehensive database involving pathways, genes, compounds, drugs, and diseases for annotating the biological functions of genes and genomes at the molecular level [33]. KEGG enrichment analyses were performed on the screened hub genes using Omicshare tools, an online platform for data analysis and gene annotation of gene DENOVO A (https://www.omicshare.com/tools/). We investigated the potential biological functions and upregulated pathways of DEGs by GO analyses using the R packages "cluster Profiler" [34] and "DOSE." Enrichment of GO and KEGG pathway with $p < 0.05$ was considered significant.

2.1.3. Protein-protein interaction (PPI) network analysis

The STRING tools (version 11.5; https://cn.string-db.org/) are a database to customize protein networks and seek functional signatures of gene sets. The database helps explore PPI networks sharing DEGs and their visualization. In addition, the module analysis of the PPI network was performed using MCODE.

2.1.4. Identification of key genes

To further screen for diagnostic biomarkers associated with IS, we employed machine learning algorithms for LASSO regression. We utilized the "glmnet" R package to integrate clinical features and gene expression data for LASSO regression analysis.Subsequently, we assessed the diagnostic accuracy of key genes for COVID-19 with stroke using the AUC score of the ROC curve (https://www.omicshare.com/tools/). To further validate the accuracy of DEGs in stroke diagnosis and to refine the selection of key genes, we conducted screening based on hub genes using RF and SVM-RFE. With the genes selected through screening, we constructed a neural network and intersected it with the key genes identified by LASSO regression, ultimately confirming them as key genes.The analysis employed the "randomForest" R package for random forest, "e1071" and "caret" R packages for SVM-RFE, and the "neuralnet" R package for neural network.

2.1.5. Exploration of potential drugs

DGIdb tool helped retrieve the screened key genes (https://dgidb.genome.wustl.edu) to determine the drugs matching the key

genes. Simultaneously, the inhibition and activation effects of drugs on genes were confirmed, followed by comparing the results of activation and inhibition of key genes in differential expression. The corresponding activators or inhibitors are selected based on the upregulated or downregulated expression status of the selected genes in the disease development process.

3. Result

3.1. Subsection

3.1.1. Identification of shared DEG between the groups of COVID-19, stroke, and control group

After normalizing the gene set (figure S1A B) , the DEGs of GSE16561 and GSE213313 were screened following the above method. GSE16561 had 169 differentially expressed genes, and GSE213313 possessed 8765 differentially expressed genes. Among them, 4766 were DEGs in the critical group, and 3999 were DEGs in the non-critical group, as indicated in the volcano map (Fig. 2A–C and Table S1) . After that, 77 DEGs were shared among COVID-19 and IS samples, and the control group was screened by a Venn

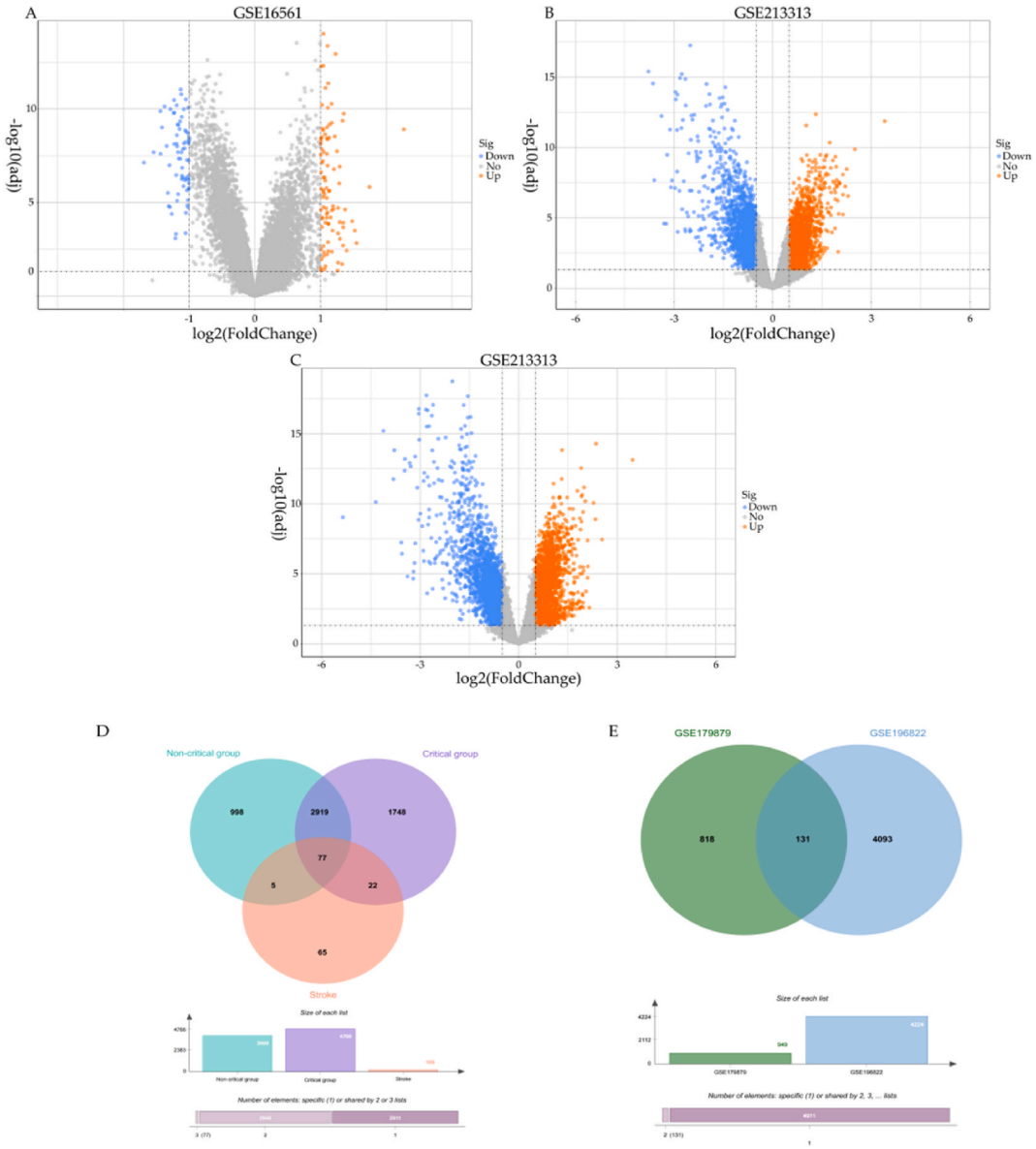

Fig. 2. Recognition of shared DEGs between COVID-19 and stroke.(A)The volcano plot of GSE16561.(B)The volcano plot of non critically ill patients in GSE213313.(C)The volcano plot of critically ill patients in GSE213313.(D)The shared DEGs between COVID-19 and stroke by overlapping DEmRNAs.(E)Venn diagram of the validation set.

diagram (Fig. 2D) .

Simultaneously, we performed differential expression gene analysis on the validation datasets (GSE179879 and GSE196822) using online tools to confirm whether the DEGs observed in the training dataset exhibit similar differential expression in different datasets. In the validation datasets, we identified a set of DEGs, including but not limited to CD3D, IL2RB, and others. The expression patterns of these genes were consistent with those observed in the training dataset, further supporting the conclusions drawn from our study (Figure S1C D,Fig. 2E).

3.1.2. Gene enrichment analysis of shared DEG

GO and KEGG enrichment analyses (Fig. 3A B) helped annotate the potential functions and pathways of 77 shared DEGs. Bioprocess (BP) enrichment GO terms involved T cell activation, differentiation, and selection. The GO terms for enriching cell components (CC) were specific granules, secretory granular membranes, tertiary particles, tertiary granular membranes, and specific granular membranes. The GO enrichment terms for molecular function (MF) were immune receptor activity, profilin binding, and non-membarne spanning protein tyrosine kinase. Moreover, the enriched KEGG pathway mainly involved interacting viral proteins interaction with cytokines and cytokines receptors, Th1 and Th2 cell differentiation, cell adhesion molecules, influenza virus, and *Staphylococcus aureus*.

3.1.3. PPI network and hub genes analysis

Genes do not exist in isolation, which makes connections through encoding proteins. The PPI network of DEGs was established using the STRING tool, with 54 nodes and 112 edges (Fig. 4A). Furthermore, the genes with the most connections were selected, with 34 edges and 9 nodes (Fig. 4B). Nine genes in this network, viz., FCGR3A, KLRB1, IL2RB, CD2, IL7R, CCR7, CD3D, GZMK, and ITK, were hub genes since this network has the highest score in the MCODE tool (Fig. S2).

3.1.4. Identification of key genes

Through LASSO regression, we identified four key genes: KLRB1, IL2RB, CD3D, and ITK, as biomarkers for diagnosing stroke (Fig. 5A B). Subsequently, we used ROC curve AUC scores (Fig. 5C) to determine whether the selected key genes could accurately diagnose stroke. The AUC scores were CD3D: 0.923, ITK: 0.938, KLRB1: 0.853, IL2RB: 0.842. Both CD3D and ITK had AUC scores exceeding 0.9. To select more accurate biomarkers, we employed RF (Fig. 5D E) and SVM-RFE (Fig. 5F) based on hub genes. RF identified five hub genes: ITK, CD2, IL7R, CD3D, and KLRB1, while SVM-RFE identified four hub genes: ITK, CD2, IL7R, and CD3D. Using these five selected genes, we constructed a neural network (Fig. 5G). The results indicated that the chosen genes could accurately diagnose stroke. In the final step, we chose the intersection of all the identified results to determine the key biomarkers, ultimately identifying CD3D and ITK as the final key biomarkers for diagnosing COVID-19-associated ischemic stroke (IS).

3.1.5. Exploration of potential drugs

Based on screening using the DGIdb tool, 23 drugs associated with key genes were selected. Among them, Blinatumomab were in the opposite direction to the activity of key genes in DEGs (Table 1). While taking Blinatumomab, it is important to be aware of immune response and neurologically related side effects. To prevent severe consequences from an overdose of Blinatumomab, research into wearable devices, similar to those developed for avoiding opioid overuse, could be beneficial. These devices can enable timely reporting of patient conditions to medical professionals, thus minimizing potential [35].

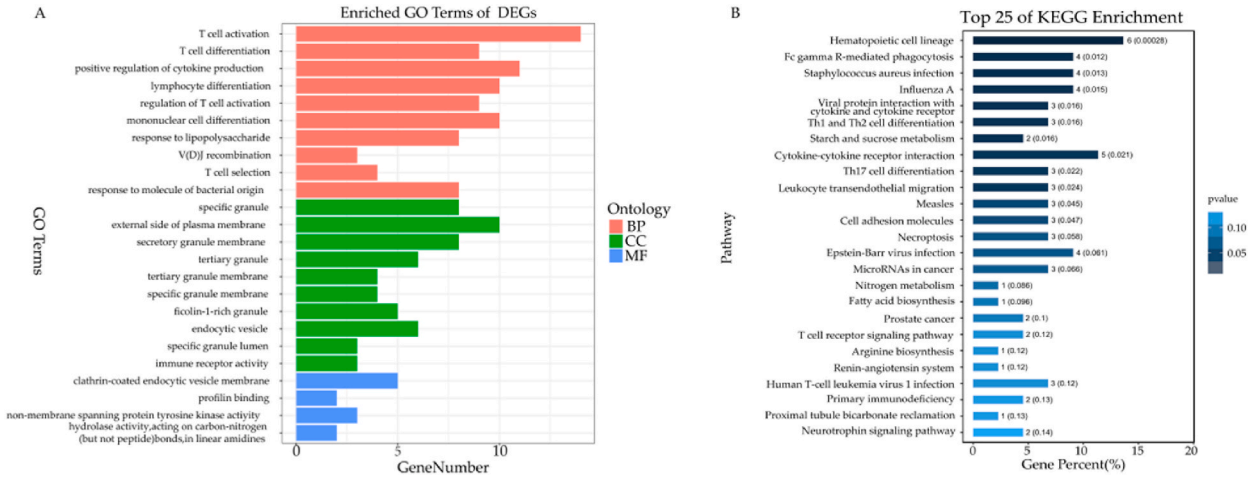

Fig. 3. enrichment analyses of the shared DEGs.(A)Gene Ontology (GO) analysis of COVID-19 related stoke of DEGS.(B)Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of COVID-19 related stroke of DEGs.

3.2. Figures, tables and schemes

Fig. 2.
Fig. 3.
Fig. 4.
Fig. 5.
Table 1.

4. Discussion

Since December 2019, the massive spread of COVID-19 due to the SARS-CoV-2 virus has burdened healthcare systems worldwide, affecting almost every country. In response to this emergency, numerous research outcomes have emerged regarding the prevention, diagnosis, and treatment of COVID-19. Among them, early diagnosis of COVID-19 proves effective in preventing further exacerbation of the disease and curbing the further spread of the epidemic [36,37]. Apart from the respiratory system response to COVID-19, the increased risk of stroke should also attract much attention, particularly IS. Stroke is the second highest cause of mortality globally, and the incidence of IS is significantly higher. Epidemiological studies indicated that IS incidence is higher than before the COVID-19 pandemic. However, the biological mechanisms behind these remain unknown. Therefore, the present study aims to investigate the possible genetic relationship between COVID-19 and IS to reveal the pathogenicity correlation and identify relevant biological diagnostic markers and therapeutic targets. We also attempt to prevent and treat IS early, reduce its risk, alleviate related neurological symptoms, and enhance the prognosis and quality of life.

This study identified 169 stroke-related differentially expressed mRNA and 8765 differentially expressed mRNAs associated with COVID-19. This also included 4766 differentially expressed genes in the critical group and 3999 in the non-critical group. Then, depending on cross-analysis, 77 shared DEGs were identified. GO enrichment analysis was performed on shared DEGs, with results related to cell functions, such as T cell activation, differentiation, selection, and lymphocyte activation. According to the analysis, the primary reason behind the increased IS risk due to COVID-19 is associated with immune function [38–41]. After acute infection with SARS-CoV-2, the peripheral blood circulating lymphocytes were decreased by about 80 %. In this case, the body was in a special mode – coexisting suppression and activation. The CD8+T cells proliferated, although peripheral lymphocytes decreased significantly [42]. Other than CD8+T cells, CD4+T cells demonstrated a corresponding proliferative response during infection. Moreover, related research has indicated that $CD4^{+}$, $CD8^{+}$, and CD25+T cells are related to an increased stroke risk. The secretion of interleukin (IL)-10, IL-4, etc., can elevate the stroke risk and aggravate nerve damage post-stroke [39,42]. Among them, the content of CD3+T cells was found to be increased in intracranial thrombus of those suffering COVID-19 and were not limited to the edge of the thrombus. Therefore, the CD3+T cells accumulated at the embolic site during the thrombus formation stages. Moreover, CD3+T cells are reconstructed in COVID-19 patients after eliminating viral infection. Therefore, CD3+T cells can be utilized as diagnostic criteria and an important biomarker for the early stage of COVID-19 and invisible stroke diagnosis [43,44]. The cell particles in the cell structure are primarily stress particles whose formation is enhanced by miR-335, significantly inhibiting brain cell apoptosis. Studies have indicated that miR-335 treatment upregulated SG formation, alleviated ischemia-induced infarction, and decreased the expression of ROcK2 protein and cell apoptosis level [45,46]. The KEGG enrichment analysis involved the viral protein interactions with cytokines and cytokine receptors, Th1 and Th2 cell differentiation, cell adhesion molecules, and the co-infection of influenza virus and *Staphylococcus aureus* during COVID-19. These cases can lead to atherosclerosis and vascular embolism. The interaction of viral protein with

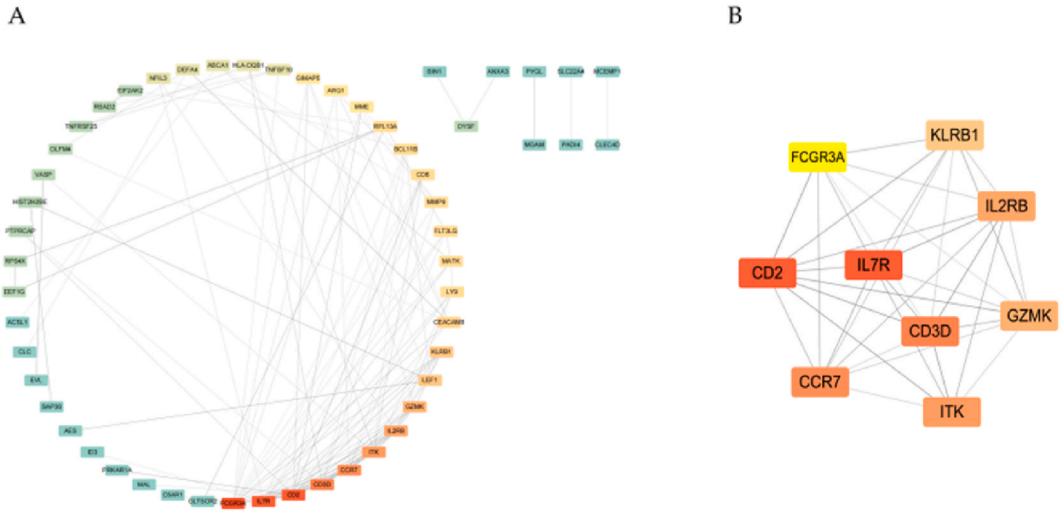

Fig. 4. PPI network analysis.(A)The PPI network of DEGs.(B)The first module of the PPI network.

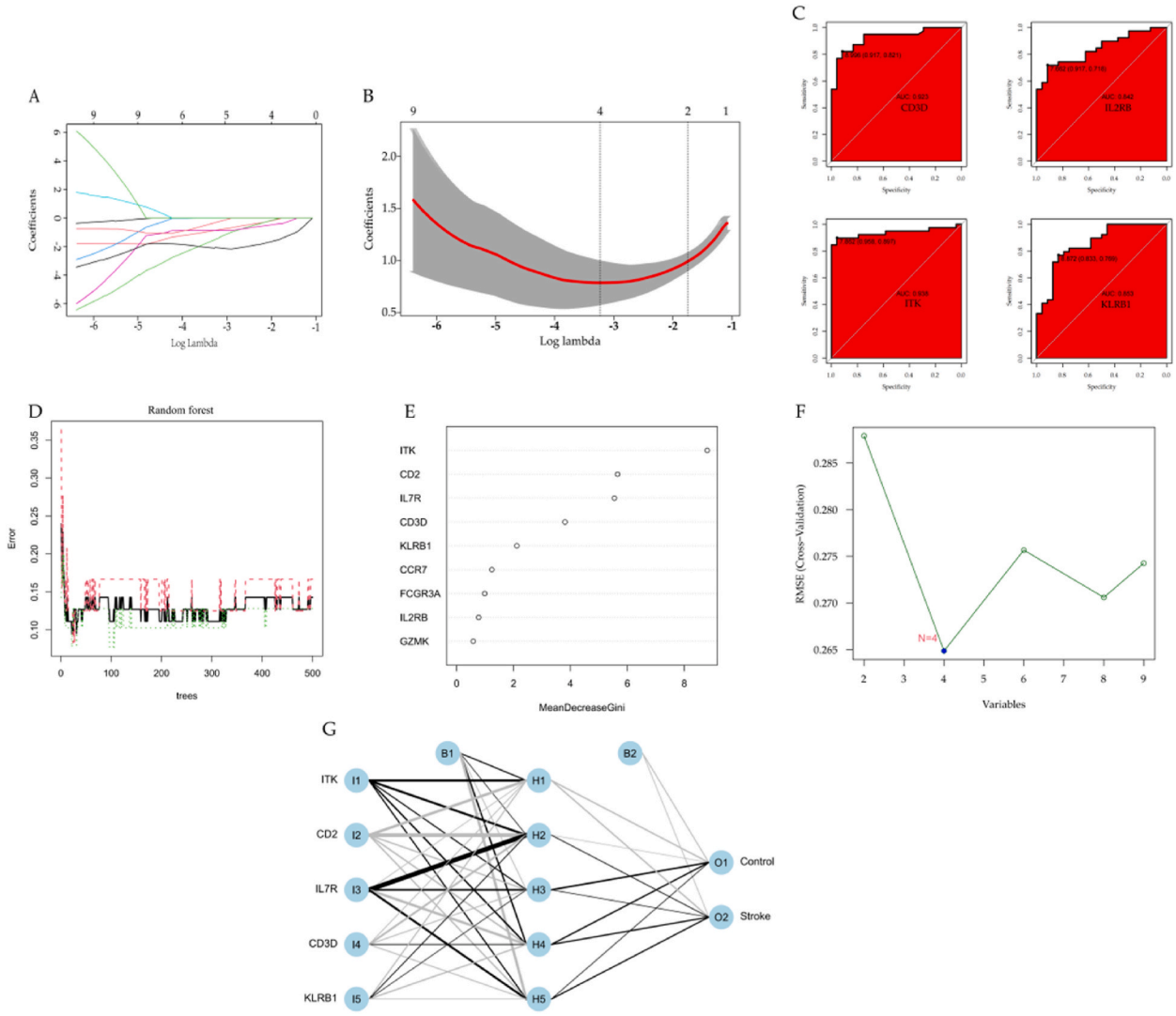

Fig. 5. Identification of the key genes. (A B) Further screening of hub genes by machine learning LASSO regression, KLRB1,CD3D,IL2RB and ITK were identified as key genes. (C) ROC curves for evaluating diagnostic accuracy of KLRB1, CD3D, IL2RB and ITK. (D) The effect of decision tree number on CV error of RF classifier. Green, red, and black curves represent the error of control, COVID-19 with stroke,and total groups respectively. (E)The mean decrease Gini coefficients of genes in the RF classifier. ITK, CD2, L7R, CD3D and KLRB1 were the key genes with mean decrease Gini larger than 2. (F) After SVM-RFE analysis, ITK, CD2, L7R,and CD3D were identified as key genes under the condition of minimum CV error. (G) NN constructed with five neurons in the input layer, five neurons in the hidden layer, and two neurons in the output layer. I, input layer; H, hidden layer; O, output layer; B, bias.

Table 1
Targeted drug.

Gene	Drug	Interaction types
CD3D	Blinatumomab	Activator

cytokines and cytokine receptors refer to the case that SARS-CoV-2 has a protein shell called spinous protein (SP). The shell binds to the angiotensin-converting enzyme 2 (ACE2) surface receptor on the host cell to promote coronavirus entry inside the host cell [47]. Afterward, SP binds to ACE2 on platelets and/or endothelial cells to activate the clotting cascade. Additionally, isolated circulating SP can induce a hypercoagulable state by interacting with fibrin/fibrinogen [48,49]. SP is an important biological recognition target for the CoV virus and can activate Th1 and Th2 cell differentiation. Microglia/macrophages and type 1 helper cells (Th1) appear harmful, while type 2 helper cells (Th2) and regulatory T cells (T-reg) are protective. Based on the cytokine environment, CD4 Th0 differentiates into Th1, Th2, or T-reg lymphocytes. The Th1 immune response is characterized by pro-inflammatory cytokine secretion, including

interferon-gamma (IFN-γ). This promotes cellular immune responses and can adversely affect stroke pathogenesis. The Th1 response enhances microglial/macrophage activation, improving TNF-α and IL-1β secretion. In contrast, the Th2 and T-reg immune responses are characterized by anti-inflammatory cytokine secretion, including IL-4, IL-10, and transforming growth factor β (TGF-β), regulating the innate immune response while playing a protective role [50]. Cell adhesion molecules mainly refer to E-selectin. COVID-19 directly mediates the damage of endothelial cells (EC) due to the COVID-19 virus itself and indirectly regulates EC activation by cytokines. E-selectin (CD 62 E) is produced by EC during activation, which releases E-selectin of a truncated form called SE-selectin. Experiments indicated that the E-selectin content in COVID-19 serum patients with thrombosis was much higher than in ordinary COVID-19 patients and healthy people [51]. A PRIME study of healthy middle-aged men evaluated nine hemostasis, inflammation, and endothelial activation biomarkers. The results indicated that 98 had a stroke during 10 years of follow-up. Moreover, plasma samples from 95 stroke patients were compared with 190 healthy individuals, showing significantly elevated E-selectin levels in the case group. Thus, e-selectin can be utilized as a biomarker to diagnose related cerebrovascular diseases [52]. During COVID-19, the number of peripheral circulating lymphocytes in patients was significantly reduced. Therefore, co-infection risks, such as influenza and *Staphylococcus aureus* infection, significantly increased. During influenza virus infection, activating protein C and plasminogen activator inhibitor type-1 production, which are key inhibitors of coagulation and fibrinolysis, is reduced, stimulating the external clotting pathway [53]. *Staphylococcus aureus* is the most common pathogen leading to coinfection and reinfection among COVID-19 patients. These patients were more likely to develop late-onset (seven days) ICU-acquired bloodstream infections than non-COVID-19 patients [54]. Bacteremia is characterized by an inflammatory and pro-coagulant state, which may be accompanied by endothelial and blood vessel damage, leakage, and hypotension, leading to arterial occlusion. Moreover, infection can cause atherosclerotic plaques to become unstable and rupture, inducing blood clots in the arteries.

Afterward, the selected shared DEGs helped build the PPI network, and nine hub genes were selected. The hub genes selected by the PPI network were processed to identify stroke-related genes and select the most appropriate biomarkers. LASSO regression helped process hub genes, providing four essential genes: KLRB1, IL2RB, CD3D, and ITK. After validating these genes with ROC curve AUC scores, both CD3D and ITK scored above 0.9. Subsequently, five diagnostic-related genes, KLRB1, CD2, CD3D, IL7R, and ITK, were jointly selected by random forest and SVM-RFE. They were accurately classified in the neural network between the stroke and control groups. Finally, through multiple validation methods, CD3D and ITK were chosen as key genes. CD3D encodes the δ subunit of the transmembrane CD3 antigen complex, which is a crucial component of the T-cell receptor/CD3 complex (TCR/CD3 complex). This complex, comprising CD3D and four other CD3 subunits, is essential for T-cell development and signal transduction. CD3D plays a significant role in T-cell immune-related pathways, and its involvement in abnormal activation has been confirmed through epigenetic and genomic analyses [55]. In parallel, the ITK gene, located on chromosome 5q, encodes the ITK protein, which is critical for adaptive immune responses. ITK participates in T-cell receptor activation and subsequent signaling pathways, including the phosphorylation of PLCγ1. This event leads to substrate cleavage, the release of calcium from the endoplasmic reticulum, and the activation of the nuclear factor of activated T-cells (NFAT), thereby promoting T-cell proliferation and differentiation [56].

Both genes have a direct connection to T-lymphocyte immunity. During COVID-19 infection, the immune response may involve the activation of CD3D and ITK, resulting in T-lymphocyte activation. However, this immune activation may contribute to a hypercoagulable state, potentially increasing the risk of ischemic stroke (IS) in COVID-19 patients. To date, there is insufficient research evidence detailing the mechanism by which CD3D and ITK may lead to IS in COVID-19 patients. This study is in the preliminary hypothesis phase, and subsequent in vivo and in vitro experiments are planned to validate this hypothesis.

Finally, Blinatumomab was selected from 23 drugs by comparing the activity of key genes in patients. Blinatumomab is the CD3D stimulant and commonly used to treat acute lymphoblastic leukemia (ALL) [57]. The mechanism of action of this drug is closely associated with the immune system of the human body, making it a potential therapeutic drug for treating stroke caused by COVID-19 infection. Nevertheless, this discovery is primarily derived from bioinformatics and machine learning predictions and currently lacks direct experimental or clinical corroboration. While these findings present promising avenues for future therapeutic strategies, the clinical applicability and efficacy of this approach warrant further investigation through comprehensive in vivo and in vitro studies, along with rigorous clinical trials. Future research will focus on assessing the therapeutic potential of Blinatumomab in the context of COVID-19-associated stroke, to establish its efficacy and safety.

This study predominantly involved samples from the European population, resulting in limited representation of Asian cohorts. Therefore, the generalizability of these findings to other populations remains uncertain. To address this limitation, future research will aim to expand the sample size through multicenter collaborations. These efforts will include patients from diverse regions, ethnicities, and clinical profiles to ensure that the results are applicable and robust across a broader range of populations.

5. Conclusions

CD3D, and ITK were identified as critical genes through bioinformatics in IS related to COVID-19. They mainly involve immunomodulatory functions and could be potential biomarkers for such disease conditions. In addition, the DGIdb tool helped predict targeted drugs: Blinatumomab, which can effectively treat IS associated with COVID-19. Our study could provide novel ideas for further research on mechanisms and treatment regimens.

CRediT authorship contribution statement

Sixian Wang: Data curation, Methodology, Writing – original draft, Writing – review & editing. **Yuxing Tai:** Writing – review & editing, Writing – original draft, Data curation. **Xiaoqian Yang:** Validation, Software. **Peizhe Li:** Software, Visualization. **Han Wang:**

Writing – review & editing. **Yi Tan:** Formal analysis, Data curation. **Tianjiao Gao:** Formal analysis, Data curation. **Mingrui Chu:** Formal analysis, Data curation. **Mingjun Liu:** Writing – review & editing, Supervision, Project administration, Investigation.

Data availability statement

The data that support the findings of this study are available in GEO database (http://www.ncbi.nlm.nih.gov/geo), reference number [GSE16561, GSE213313, GSE179879 and GSE196822].

Funding

National Administration of Traditional Chinese Medicine, LPGZS2012-11.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e39039.

References

[1] P. Pourciau, B.C. Smith, Stroke risk related to coronavirus disease-2019, Crit. Care Nurs. Clin. 35 (1) (2023) 53–65, https://doi.org/10.1016/j.cnc.2022.10.001.
[2] Lippi G, Sanchis-Gomar F, Henry BM. COVID-19 and its long-term sequelae: what do we know in 2023? Polish Archives of Internal Medicine. Published online January 9, 2023. doi:10.20452/pamw.16402.
[3] A. Nalbandian, K. Sehgal, A. Gupta, et al., Post-acute COVID-19 syndrome, Nat Med 27 (4) (2021) 601–615, https://doi.org/10.1038/s41591-021-01283-z.
[4] P. Tater, S. Pandey, Post-stroke movement disorders: clinical spectrum, pathogenesis, and management, Neurol. India 69 (2) (2021) 272, https://doi.org/10.4103/0028-3886.314574.
[5] Vetriselvan Subramaniyan, et al., COVID-19-associated mucormycosis and treatments, Asian Pac. J. Tropical Med. 14 (9) (2021) 401–409.
[6] G.J. Hankey, Secondary stroke prevention, Lancet Neurol. 13 (2) (2014) 178–194, https://doi.org/10.1016/S1474-4422(13)70255-2.
[7] J. Gutierrez, C. Esenwa, Secondary stroke prevention: challenges and solutions, VHRM 437 (August 2015), https://doi.org/10.2147/VHRM.S63791. Published online.
[8] S. Wu, B. Wu, M. Liu, et al., Stroke in China: advances and challenges in epidemiology, prevention, and management, Lancet Neurol. 18 (4) (2019) 394–405, https://doi.org/10.1016/S1474-4422(18)30500-3.
[9] G.A. Roth, G.A. Mensah, C.O. Johnson, et al., Global burden of c ardiovascular diseases and risk factors, 1990–2019, J. Am. Coll. Cardiol. 76 (25) (2020) 2982–3021, https://doi.org/10.1016/j.jacc.2020.11.010.
[10] D. Zhang, J. Ren, Y. Luo, et al., T cell response in ischemic stroke: from mechanisms to translational insights, Front. Immunol. 12 (2021) 707972, https://doi.org/10.3389/fimmu.2021.707972.
[11] W.J. Wang, J.J. Lu, Y.J. Wang, et al., Clinical characteristics, management, and functional outcomes in Chinese patients within the first year after intracerebral hemorrhage: analysis from China national stroke registry, CNS Neurosci. Ther. 18 (9) (2012) 773–780, https://doi.org/10.1111/j.1755-5949.2012.00367.x.
[12] the ChinaQUEST Investigators, J.W. Wei, Y. Huang, et al., Current management of intracerebral haemorrhage in China: a national, multi-centre, hospital register study, BMC Neurol. 11 (1) (2011) 16, https://doi.org/10.1186/1471-2377-11-16.
[13] G. Chen, A. Wang, X. Zhao, et al., Frequency and risk factors of spontaneous hemorrhagic transformation following ischemic stroke on the initial brain CT or MRI: data from the China National Stroke Registry (CNSR), Neurol. Res. 38 (6) (2016) 538–544, https://doi.org/10.1080/01616412.2016.1187864.
[14] D. Wang, J. Liu, M. Liu, C. Lu, M. Brainin, J. Zhang, Patterns of stroke between university hospitals and nonuniversity hospitals in mainland China: prospective multicenter hospital-based registry study, World Neurosurgery 98 (2017) 258–265, https://doi.org/10.1016/j.wneu.2016.11.006.
[15] J. Liu, D. Wang, R. Yuan, Y. Xiong, M. Liu, Prognosis of 908 patients with intracerebral hemorrhage in Chengdu, Southwest of China, Int. J. Neurosci. 127 (7) (2017) 586–591, https://doi.org/10.1080/00207454.2016.1216414.
[16] J. Zhou, Y. Zhang, H. Arima, et al., Sex differences in clinical characteristics and outcomes after intracerebral haemorrhage: results from a 12-month prospective stroke registry in Nanjing, China, BMC Neurol. 14 (1) (2014) 172, https://doi.org/10.1186/s12883-014-0172-5.
[17] Suganya Venkateshan, et al., Anti-oxidant and anti-hyperlipidemic activity of Hemidesmus indicus in rats fed with high-fat diet, Avicenna journal of phytomedicine 6 (5) (2016) 516–525.
[18] J.W. Wei, H. Arima, Y. Huang, et al., Variation in the frequency of intracerebral haemorrhage and ischaemic stroke in China: a national, multicentre, hospital register study, Cerebrovasc. Dis. 29 (4) (2010) 321–327, https://doi.org/10.1159/000278927.
[19] F. Shehjar, B. Maktabi, Z.A. Rahman, et al., Stroke: molecular mechanisms and therapies: update on recent developments, Neurochem. Int. 162 (2023) 105458, https://doi.org/10.1016/j.neuint.2022.105458.
[20] D. Ramel, S. Gayral, M.K. Sarthou, N. Augé, A. Nègre-Salvayre, M. Laffargue, Immune and smooth muscle cells interactions in atherosclerosis: how to target a breaking bad dialogue? Front. Pharmacol. 10 (2019) 1276, https://doi.org/10.3389/fphar.2019.01276.
[21] Thanh N. Nguyen, et al., Global impact of the COVID-19 pandemic on stroke volumes and cerebrovascular events: a 1-year follow-up, Neurology 100 (4) (2023) e408–e421, https://doi.org/10.1212/WNL.0000000000201426.
[22] D. Sagris, A. Papanikolaou, A. Kvernland, et al., COVID-19 and ischemic stroke, Euro J of Neurology 28 (11) (2021) 3826–3836, https://doi.org/10.1111/ene.15008.
[23] D. Bano, E. Munarriz, H.L. Chen, et al., The plasma membrane Na+/Ca2+ exchanger is cleaved by distinct protease families in neuronal cell death, Ann. N. Y. Acad. Sci. 1099 (1) (2007) 451–455, https://doi.org/10.1196/annals.1387.006.
[24] Velmurugan Chinnasamy, et al., Antiarthritic activity of Achyranthes aspera on formaldehyde - induced arthritis in rats, Open access Macedonian journal of medical sciences 7 (30 Aug. 2019) 17 2709–2714, https://doi.org/10.3889/oamjms.2019.559.
[25] A. Gasmi, T. Tippairote, P.K. Mujawdiya, et al., Neurological involvements of SARS-CoV2 infection, Mol. Neurobiol. 58 (3) (2021) 944–949, https://doi.org/10.1007/s12035-020-02070-6.

[26] Geeta Joshi, et al., FDA-approved artificial intelligence and machine learning (AI/ML)-enabled medical devices: an updated landscape, Electronics 13 (3) (2024) 498.
[27] T.L. Barr, Y. Conley, J. Ding, et al., Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling, Neurology 75 (11) (2010) 1009–1014, https://doi.org/10.1212/WNL.0b013e3181f2b37f.
[28] C. Prebensen, Y. Lefol, P.L. Myhre, et al., Longitudinal whole blood transcriptomic analysis characterizes neutrophil activation and interferon signaling in moderate and severe COVID-19, Sci. Rep. 13 (1) (2023) 10368, https://doi.org/10.1038/s41598-023-37606-y.
[29] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (6) (2012) 882–883, https://doi.org/10.1093/bioinformatics/bts034.
[30] K. McCann, A. Von Witzleben, J. Thomas, et al., Targeting the tumor mutanome for personalized vaccination in a TMB low non-small cell lung cancer, J Immunother Cancer 10 (3) (2022) e003821, https://doi.org/10.1136/jitc-2021-003821.
[31] U. Banerjee, P. Rao, M. Reddy, et al., A 9-gene biomarker panel identifies bacterial coinfections in culture-negative COVID-19 cases, Mol Omics 18 (8) (2022) 814–820, https://doi.org/10.1039/D2MO00100D.
[32] Gene Ontology Consortium, The gene Ontology (GO) project in 2006, Nucleic Acids Res. 34 (90001) (2006) D322–D326, https://doi.org/10.1093/nar/gkj021.
[33] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs, Nucleic Acids Res. 45 (D1) (2017) D353–D361, https://doi.org/10.1093/nar/gkw1092.
[34] R. Petegrosso, Z. Li, R. Kuang, Machine learning and statistical methods for clustering single-cell RNA-sequencing data, Briefings Bioinf. 21 (4) (2020) 1209–1223, https://doi.org/10.1093/bib/bbz063.
[35] N.K. Singh, G.K. Sidhu, K. Gupta, Current and future perspective of devices and diagnostics for opioid and OIRD, Biomedicines 10 (4) (2022) 743, https://doi.org/10.3390/biomedicines10040743.
[36] S.K. Pandey, G.C. Mohanta, V. Kumar, K. Gupta, Diagnostic tools for rapid screening and detection of SARS-CoV-2 infection, Vaccines 10 (8) (2022) 1200, https://doi.org/10.3390/vaccines10081200.
[37] Velmurugan Chinnasamy, et al., Antiarthritic activity of Achyranthes aspera on formaldehyde - induced arthritis in rats, Open access Macedonian journal of medical sciences 7 (30 Aug. 2019) 17 2709–2714, https://doi.org/10.3889/oamjms.2019.559.
[38] Y.R. Wang, W.Q. Cui, H.Y. Wu, X.D. Xu, X.Q. Xu, The role of T cells in acute ischemic stroke, Brain Res. Bull. 196 (2023) 20–33, https://doi.org/10.1016/j.brainresbull.2023.03.005.
[39] X. Luo, Y. Zhu, J. Mao, R. Du, T cell immunobiology and cytokine storm of COVID-19, Scand. J. Immunol. 93 (3) (2021), https://doi.org/10.1111/sji.12989.
[40] X. Wang, W. Xuan, Z.Y. Zhu, et al., The evolving role of neuro-immune interaction in brain repair after cerebral ischemic stroke, CNS Neurosci. Ther. 24 (12) (2018) 1100–1114, https://doi.org/10.1111/cns.13077.
[41] A.T. DiPiazza, B.S. Graham, T.J. Ruckwardt, T cell immunity to SARS-CoV-2 following natural infection and vaccination, Biochem. Biophys. Res. Commun. 538 (2021) 211–217, https://doi.org/10.1016/j.bbrc.2020.10.060.
[42] P. Moss, The T cell immune response against SARS-CoV-2, Nat. Immunol. 23 (2) (2022) 186–193, https://doi.org/10.1038/s41590-021-01122-w.
[43] C. Dargazanli, V. Rigau, O. Eker, et al., High CD3+ cells in intracranial thrombi represent a biomarker of atherothrombotic stroke, in: X. Cheng (Ed.), PLoS ONE, vol. 11, 2016 e0154945, https://doi.org/10.1371/journal.pone.0154945, 5.
[44] C. Ibeh, M.S.V. Elkind, Stroke prevention after cryptogenic stroke, Curr. Cardiol. Rep. 23 (12) (2021) 174, https://doi.org/10.1007/s11886-021-01604-1.
[45] Wenwen Si, et al., miR-335 promotes stress granule formation to inhibit apoptosis by targeting ROCK2 in acute ischemic stroke, Int. J. Mol. Med. 43 (3) (2019) 1452–1466, https://doi.org/10.3892/ijmm.2019.4073.
[46] Riya Thapa, et al., Recent developments in the role of protocatechuic acid in neurodegenerative disorders, EXCLI journal 22 (7 Jul. 2023) 595–599, https://doi.org/10.17179/excli2023-5940.
[47] P. Natesan Pushparaj, L.A. Damiati, I. Denetiu, et al., Deciphering SARS CoV-2-associated pathways from RNA sequencing data of COVID-19-infected A549 cells and potential therapeutics using in silico methods, Medicine 101 (35) (2022) e29554, https://doi.org/10.1097/MD.0000000000029554.
[48] Manuela De Michele, et al., Evidence of SARS-CoV-2 spike protein on retrieved thrombi from COVID-19 patients, J. Hematol. Oncol. 15 (2022), https://doi.org/10.1186/s13045-022-01329-w.
[49] GBD 2021 Fertility and Forecasting Collaborators, Global fertility in 204 countries and territories, 1950-2021, with forecasts to 2100: a comprehensive demographic analysis for the Global Burden of Disease Study 2021, Lancet (London, England) 403 (2024) 10440, https://doi.org/10.1016/S0140-6736(24)00550-6, 2057-2099.
[50] M. Poittevin, N. Deroide, F. Azibani, et al., Glatiramer Acetate administration does not reduce damage after cerebral ischemia in mice, J. Neuroimmunol. 254 (1–2) (2013) 55–62, https://doi.org/10.1016/j.jneuroim.2012.09.009.
[51] M.M. Watany, S. Abdou, R. Elkolaly, N. Elgharbawy, H. Hodeib, Evaluation of admission levels of P, E and L selectins as predictors for thrombosis in hospitalized COVID-19 patients, Clin. Exp. Med. 22 (4) (2022) 567–575, https://doi.org/10.1007/s10238-021-00787-9.
[52] C. Prugger, G. Luc, B. Haas, et al., Multiple biomarkers for the prediction of ischemic stroke: the PRIME study, ATVB 33 (3) (2013) 659–666, https://doi.org/10.1161/ATVBAHA.112.300109.
[53] R. Rubino, C. Imburgia, S. Bonura, M. Trizzino, C. Iaria, A. Cascio, Thromboembolic events in patients with influenza: a scoping review, Viruses 14 (12) (2022) 2817, https://doi.org/10.3390/v14122817.
[54] G. De Pascale, F. De Maio, S. Carelli, et al., Staphylococcus aureus ventilator-associated pneumonia in patients with COVID-19: clinical features and potential inference with lung dysbiosis, Crit. Care 25 (1) (2021) 197, https://doi.org/10.1186/s13054-021-03623-4.
[55] Junqin Lu, et al., CD3D, GZMK, and KLRB1 are potential markers for early diagnosis of rheumatoid arthritis, especially in anti-citrullinated protein antibody-negative patients, Front. Pharmacol. 12 (16 Sep) (2021) 726529, https://doi.org/10.3389/fphar.2021.726529.
[56] Sujal Ghosh, et al., Interleukin-2-inducible T-cell kinase (ITK) deficiency - clinical and molecular aspects, J. Clin. Immunol. 34 (8) (2014) 892–899, https://doi.org/10.1007/s10875-014-0110-8.
[57] A.S. Advani, A. Moseley, K.M. O'Dwyer, et al., Dasatinib/prednisone induction followed by blinatumomab/dasatinib in Ph+ acute lymphoblastic leukemia, Blood Advances 7 (7) (2023) 1279–1285, https://doi.org/10.1182/bloodadvances.2022008216.