Identification of potential biomarkers of venous thromboembolism in patients with COVID-19 via an integrated bioinformatic-based study

Zhao-jun Chen
Zhongnan Hospital of Wuhan University

Si-chi Xu
The Central Hospital of Wuhan, Huazhong University of Science and Technology

Hai-hua Chen (chenhaihua@whu.edu.cn)
Zhongnan Hospital of Wuhan University

Research Article

Keywords: COVID-19, venous thromboembolism, biomarker, bioinformatics

Posted Date: August 2nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3203108/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background

As a result of the COVID-19 pandemic, venous and arterial thromboembolic events have increased dramatically in many patients. This study aimed to identify the potential biomarkers for COVID-19-associated venous thromboembolism (VTE).

Methods

The datasets about COVID-19 and VTE were downloaded from the Gene Expression Omnibus (GEO) dataset. Weighted gene co-expression network analysis (WGCNA) was conducted to identify the most critical module associated with the clinic results. The intersection of common modules was processed for further enrichment analysis. The hub genes were identified by combining the differentially expressed genes (DEGs) of VTEs and common module genes. Then, the final diagnostic value for VTE was verified through bioinformatics algorithms.

Results

As many as 78 common critical genes were summarized by the intersection of the most positive and negative modules of COVID-19 and VTE. These genes were mainly enriched in coronavirus disease, fluid shear stress and atherosclerosis, ribosome, NF-kappa B, and TNF signalling pathways. Four critical genes including GZMA, BCL2A1, CD52, and RANSE2 were selected by performing the intersection analysis with the DEGs in VTE. All these genes were found to be increased in VTE samples in GSE19151 and achieved a good diagnostic value with a relative proper area under the curve (AUC).

Conclusions

Our study found that similar changes occurred in COVID-19 and VTE. GZMA, BCL2A1, CD52, and RANSE2 can be utilized as potential diagnostic markers for COVID-19-related VTE. This study may offer new opportunities for the detection and prevention of COVID-19-induced hypercoagulable state and VTE.

Introduction

The coronavirus disease of 2019 (COVID-19), caused by the coronavirus SARS-CoV2 (Severe Acute Respiratory Syndrome Coronavirus 2), has caused tens of thousands of deaths during the past three years. Despite acting as a respiratory infectious disease, severe COVID-19 patients are usually threatened with an increased risk of venous and arterial thromboembolic formation \(^1\). Thromboembolic complications have been frequently discussed in COVID-19 patients in previous studies \(^2,3\). The current
clinical evidence indicates that both pulmonary embolism (PE) and deep vein thrombosis (DVT) are the most frequently thrombotic events in severe COVID-19. As many as 25 to 30% of patients suffered from VTE, especially in critically ill and mechanically ventilated patients. Another study also proposed that asymptomatic DVT was estimated to be as high as 85% in critically ill patients and 46% of hospitalized medical patients via ultrasound tests. A recent systematic review and meta-analysis including 86 studies reported an incidence of 7.9% in non-ICU patients and 22.7% in ICU patients with a PE prevalence of 3.5% and 13.7%, respectively. Moreover, multi-organs are also disturbed by COVID-19-induced hypercoagulopathy, including the vasculature of the lungs, legs, spleen, heart, and brain. These complications are usually accompanied with multiorgan failure and high mortality in severe cases. Jose Maria Pereira de Godoy reported that the mortality rate was 67% in the group positive for DVT and 31% in the group negative for DVT, which demonstrated DVT is associated with an increased mortality in patients with COVID-19.

In addition, in a cohort of 26 COVID-19 autopsy cases, Xiao-Hong Yao had reported that SARS-CoV-2 spike protein can be detected in CD34 positive endothelia in blood ranging from blood-air barrier or pulmonary vessels, blood–testis barrier and renal proximal convoluted tubular epithelia in serial sections of the COVID-19 patients. Autopsy evidence also confirmed that SARS-CoV2 can cause direct or indirect damage to blood vessels, too. Numerous governmental and professional associations have published guidance for the screening, prevention, and treatment of VTE in hospitalized patients with COVID-19.

In previous studies, multi-bioinformatic analysis has been conducted to elaborate the molecular mechanisms of COVI-19-related cardiac injury, stroke, and renal impairment. At the same time, the link between COVID-19 and coagulopathy is attracting attention from both clinicians and basic scientists. However, the relationship between COVID-19 and VTE has not completely been eliminated, and no new biomarker for the evaluation of high-risk patients with DVT has been proposed yet. In this study, we intend to investigate the specific relationship and mechanism between COVID-19 and VTE via multiple bioinformatic analysis.

Materials and Methods

Extraction of data and differentially expressed genes (DEGs)

The original gene expression data sets of COVID-19 datasets (GSE152418, GSE164805, GSE166253, GSE171110) and VTE datasets (GSE48000, GSE19151) were obtained from the Gene Expression Omnibus (GEO) database. The specific information of the sample and platform is presented in table-1. The probe ID in each dataset was transformed into gene symbols according to the probe annotation files. In cases where some probes have the same gene ID, the average value of the probes was considered as the final expression value. The final two databases were merged into metadata cohort and the batch effects were eliminated by using the combat function of the SVA package. The batch effects were
depicted by principal component analysis (PCA). The “limma” package was used to screen DEGs. The DEGs were calculated and displayed via the “ggplot2” package with the threshold of P value < 0.05 and |log2FC| > 0.5.

Weighted gene co-expression network analysis (WGCNA)

WGCNA was conducted to investigate the most potential functional modules via scale-free networks. The top 25% of genes of each dataset with high expression variance were processed for further analysis. Then the standard-scale free network was constructed to obtain the appropriate soft threshold power and adjacency values. The adjacency values were then transformed into a topological overlap matrix (TOM) and obtained the dissimilarity (1-TOM) values. Finally, hierarchical clustering analysis with the average value was used to construct the cluster tree and divide the genes into different modules, respectively. The correlation between different modules and clinical manifestations was evaluated by using the Pearson correlation analysis. The most positive and negative modules associated with COVID-19 and VTE were selected for further analysis.

The diagnostic value of COVID-19-related VTE biomarkers

To assess the diagnostic value of the hub genes, receiver operating characteristic (ROC) curve analysis was conducted for the dataset using the “pROC” package. Then, areas under the ROC curves (AUC) were adopted to evaluate the diagnostic sensitivity and efficiency.

Functional enrichment analysis and hub gene

To elaborate the specific function of genes during COVID-19 and VTE, the R package termed “clusterProfiler” was used to carry out Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis from the intersection of the most related COVID-19 and VTE module. The interaction of the genes was conducted by the PPI network in the string dataset(https://string-db.org/). Then the result was imported into Cystoscope software 3.9.1 (America) and the hub gene was identified by the “cytohub” plug-in components via the degree value.

Distribution of immune cells and the relationship with hub genes

CIBERSORT algorithm was used to estimate the proportion of various immune cell infiltration for each sample by using the “cibersort” package in R. Then, the differences of the 22 immune cell subtypes were constructed and the results were presented with violin plots via “vioplot” package in R. Moreover, the correlation among 22 immune cell subtypes in the database was visualized using the “corrplot” package. The correlations between the hub genes and immune cells were calculated using spearman’s rank analysis and the results were presented using the “ggplot2” package.

Results and Discussion

Data processing in both COVID-19 and VTE
The four datasets of COVID-19 (GSE152418, GSE164805, GSE166253, and GSE171110) were merged into single COVID-19 metadata, while the two datasets of VTE (GSE19151 and GSE48000) were summarized into another VTE dataset. The batch effects and interbatch variation among datasets were removed after normalization. The results before and after correction were displayed in a two-dimensional PCA cluster diagram before and after normalization (Fig. 1).

Weighted gene co-expression network analysis (WGCNA) analysis

All genes related to COVID-19 and VTE were calculated by WGCNA analysis.

Thelevant co-expressed gene modules were extracted from the metadata datasets. The scale-free topological index was 0.9 and the soft thresholds for COVID-19 and VTE were 6, respectively (Fig. 2A-D.). We also calculated the dendrogram of derived genes as well as the corresponding multi-modules. The phenotypes were presented by hierarchical clustering and Spearman correlation analysis. The relationships between the different modules and clinical information were explored and represented in the heat map in Fig. 2E-F. As shown in the brown modules ($r = 0.69, P = 1E−19$) were the most positively associated with COVID-19 and red modules ($r = -0.55, P = 2E−11$) were negatively associated with COVID-19. At the same time, the yellow modules ($r = 0.53, P = 3E−19$) were the most positively associated with VTE and the blue modules ($r = -0.37, P = 2E−09$) were negatively associated with VTE. Then we conducted the intersection between the most positive and negatively correlated modules of COVID-19 and VTE and as many as 78 common critical genes were summarized in the end.

Functional enrichment analysis of common intersecting genes

We also performed functional enrichment analysis to decipher the features of the 78 intersection genes. The consequences of the GO analysis showed that the common intersected genes were mainly involved in the regulation of leukocyte migration, positive regulation of macrophage migration, regulation of epithelial cell apoptotic process (BP); external side of the plasma membrane, cytosolic ribosome, cytosolic small ribosomal subunit, small ribosomal subunit (CC); metalloaminopeptidase activity, metalloexopeptidase activity, aminopeptidase activity, and complement receptor activity (MF) (Fig. 3A). The KEGG analysis indicated that these common genes were mainly enriched in hematopoietic cell lineage, coronavirus disease, fluid shear stress and atherosclerosis, ribosome, NF-kappa B signalling pathway, and TNF signalling pathway (Fig. 3B).

Screen and verify the hub genes

We performed the intersection analysis between the common critical genes from WGCNA analysis and DEGs in the VTE dataset. Finally, 14 intersecting genes were obtained. (Figure-4A). The interaction of all genes was constructed by the PPI network as shown in Fig. 4B. Then the hub genes were calculated by cystoscope software. Finally, four critical genes including GZMA, BCL2A1, CD52, and RANSE2 were selected as the most potentially relevant biomarkers as shown in Fig. 4C-D.

Screen the diagnostic value of crucial genes
Then we tested the diagnostic efficiency of the four hub genes. We discovered that the four genes of GZMA, BCL2A1, CD52 and RANSE2 were increased in VTE samples compared with those in control in GSE19151 (Figs. 5A-D). In addition, according to the ROC analysis, GZMA, BCL2A1, CD52, and RANSE2 had all achieved a proper diagnostic value with the relative result of AUC 0.794 (95% CI: 0.715–0.866), 0.741(95% CI: 0.654 – 0.821), 0.851(95% CI: 0.780 – 0.912) and 0.773(95% CI: 0.686 – 0.848) in GSE 19151 (Figs. 5E-H).

Assessment of immunocytes infiltration in VTE dataset

COVID-19 was closely associated with the immune and inflammatory response as well as VTE. In order to understand the correlation between GZMA, BCL2A1, CD52, and RANSE2, the immune microenvironment and the ratio of immune cell infiltration were conducted on the CIBERSORT algorithm and 22 types of immune cell profiling of VTE were accomplished in Fig. 6A. At the same time, the relationship between 22 immune cells was displayed in Fig. 6B. B cell memory was significantly associated with dendritic cells activated (r = 0.31), but inversely related to B cell naive (r = −0.35). T cells CD4 memory activated were negatively associated with T cells regulatory (Tregs) (r = −0.52). NK cells resting were negatively associated with T cells CD4 memory resting (r = −0.33). neutrophils (r = −0.34) and T cells CD4 naïve (r = −0.34). T cells CD8 were negatively correlated with T cells CD4 memory resting (r = −0.32) and neutrophils(r = −0.6). T cells regulatory (Tregs) were negatively associated with T cells CD4 memory activated (r = −0.52), T cells CD4 memory resting (r = −0.32), and monocyte (r = −0.43). Neutrophils were negatively associated with NK cells resting (r = −0.34), T cells CD8(r = −0.6). Based on the study's results, the fraction of T cells CD4 memory resting(P = 0.024), T cells CD4 memory activated (P< 0.001), and T cells regulatory (Tregs)(P < 0.001), T cells gamma delta (P = 0.009), monocytes (P < 0.001), macrophage M2 (P = 0.04), and neutrophils (P = 0.011) were obvious different from the normal samples in VTE patients (Fig. 6C).

The relationship between hub genes and immunocytes

In addition, we also assessed the relationship between GZMA, BCL2A1, CD52, and RANSE2 and the ratio of multi-infiltrating cells. GZMA was positively associated with T cells CD8, monocytes, NK cells resting T cells CD4 memory activated and inversely correlated with macrophage M0, neutrophils, T cell CD4 naïve and T cell regulatory (Tregs) (Fig. 7A and Supplementary Figure S1). BCL2A1 was strongly correlated with T cells memory resting, monocytes, neutrophils and T cells CD4 memory activated, and negatively correlated with Tregs, T cells CD8, and NK cells resting (Fig. 7B and Supplementary Figure S2). CD52 was highly related to macrophages, and T cells CD4 memory activated, T cells CD4 memory resting, and negatively associated with neutrophils, macrophage Mo, and Tregs (Fig. 7C and Supplementary Figure S3). RANSE2 was positively correlated with monocyte, T cells CD4 memory activated, monocytes, T cells CD4 memory resting and negatively correlated with B cell naive, neutrophils, T cell CD4 naive, and Tregs (Fig. 7D and Supplementary Figure S4). These results further supported the effect of GZMA, BCL2A1, CD52, and RANSE2 on the immune activity of the immune microenvironment.
Discussion

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), results in substantial pulmonary parenchyma lesions and leads to obvious manifestations outside the respiratory system \(^1^6\). COVID-19 may predispose patients to a hypercoagulable state, formation of deep vein thrombosis, and even pulmonary embolism (PE) due to persistent hypoxia, excessive inflammation response, blood stasis, platelet activation, and endothelial dysfunction \(^1^7\). Multiorgan failure and death may be the direct consequence of hypercoagulability complications in patients with COVID-19. In the previous autopsy report of 12 consecutive patients who died of COVID-19, a high incidence of DVT (58%) was detected and one-third of patients died of PE \(^1\). However, the precise ratios of the risk factors, incidence, mortality, and prognosis associated with VTE in patients with COVID-19 are limited due to only single-centre reports. VTE was estimated to occupy 4.8% and 85% of hospitalized patients diagnosed with COVID-19 \(^6, 1^8–2^1\). This variability may vary according to the positive events counted, the type of screening for VTE, the assessment setting, and the use and type of thromboprophylaxis.

Comprehensive risk assessment of thrombotic and haemorrhagic events is crucial for applying strategies to improve patient outcomes. Previous study has proposed that patients with D-dimer levels of 1.0 µg/L or higher had an almost 18-fold increased risk of death \(^2^2\). One study adopted the International Society on Thrombosis and Haemostasis definition of disseminated intravascular coagulation (DIC) and found that a score of 5 points was detected in 71% of those who died compared with 0.6% of survivors \(^2^3\). Since high-quality diagnostic accuracy studies are lacking, d-dimer levels should be interpreted carefully, considering the limited clinical evidence \(^2^4\). It still deserves our attention that elevated d-dimer levels can be associated with increased odds of bleeding, particularly accompanied with thrombocytopenia events. Other coagulation parameters such as prothrombin time (PT) and activated partial thrombin time (APTT), fibrinogen, fibrin degradation products (FDP) and antithrombin levels were barely satisfactory. Our study demonstrated that COVID-19 and VTE showed similar changes in gene expression. As many as 78 common critical genes were collected by the intersection of the most positive and negative modules in COVID-19 and VTE datasets. These genes were mainly enriched in coronavirus disease, fluid shear stress and atherosclerosis, ribosome, NF-kappa B, and TNF signalling pathways.

Aldo Bonaventura proposed that SARS-CoV-2 infection induces a process known as immunothrombosis, in which activated neutrophils and monocytes interact with platelets and the coagulation cascade leading to intravascular clot formation in small and larger vessels. Microthrombotic complications may result in serious acute respiratory distress syndrome (ARDS) and other organ dysfunctions. Several antithrombotic and immunomodulating drugs have been considered as potential drugs to treat patients with SARS-CoV-2 infection \(^2^5\).

COVID-19 may predispose to DVT in several ways. The fundamental pathophysiology of thrombosis is likely attributed to a combination of SARS-CoV-2-mediated direct endothelial injury and the second dysregulated inflammation-induced coagulation activation. In prior studies, it has been proved that many
different viruses can directly trigger the coagulation system, such as HIV, dengue virus, and Ebola virus \(^{26, 27}\). In addition, the secondary inflammatory reaction, characterized by increased levels of von Willebrand factor; by Toll-like receptor activation; and a pro-coagulatory state by tissue factor pathway activation, can directly lead to endothelial dysfunction \(^{28}\). Moreover, stubborn hypoxemia develops in some patients, especially in critical ill patients. Thrombus formation under hypoxic conditions is verified both in animal models and in human beings. The vascular response under hypoxia is adjusted primarily by the hypoxia-inducible transcription factors, whose target genes include several factors that regulate thrombus formation \(^{29}\). Finally, the dysregulated system and too much release of inflammatory cytokines can also lead to a cascade of events in the blood circulation that may trigger vascular endothelial injury and thrombosis.

Our study further demonstrated that the fraction of T cells CD4 memory resting, T cells CD4 memory activated, and T cells regulatory, T cells gamma delta, monocytes, macrophage M2, and neutrophils were significantly different from the normal samples in VTE patients. Unlike previous studies that used inflammatory factors and pro-inflammatory release factors as diagnostic markers, our study identified novel markers related to SARS-CoV2-associated vein thrombosis. GZMA, BCL2A1, CD52, and RANSE2 can be used as potential diagnostic markers for COVID-19 related VTE diagnosis, and their expression levels are strongly correlated with immune cell.

GZMA (granzymes A)

Granzyme A (GZMA, tryptase) belong to a family of homologous serine proteases primarily expressed in some cytotoxic cells, such as cytotoxic T lymphocytes (CTLs), γδ T cells, and natural killers (NK) cells. The cytotoxic effect of apoptosis was induced by recruiting the death receptor pathway or granule secretory pathway \(^{30}\). Whereas the death receptor pathway includes death cell surface receptor-ligand interaction and caspase activation, the granule secretory pathway delivers granzymes through a process under the aid of perforin to target cells. Upon recognition of a target cell, cytotoxic cells release the content of granules into the immunological synapse. GZMA is the most abundant protease present in cytotoxic granules and is reported as the dominant mediator of toxicity in vitro \(^{31}\). Moreover, cumulative clinical and biochemical evidence proved that elevated levels of extracellular GZMA in plasma, serum, synovial fluid, and bronchoalveolar lavage (BAL) fluid in patients under various viral infections, bacterial infections or other pro-inflammatory conditions \(^{32–34}\). These increased levels of extracellular GZMA could potentially display the spontaneous or inadvertent release of granzymes in response to persistent inflammation. In addition, elevated levels of GZMA were commonly detected in various inflammatory conditions, such as bacterial infections, viruses, parasites, and rheumatoid arthritis \(^{35}\). The results of our study also confirmed the elevated levels of GZMA in VTE patients and found that the increased levels were closely associated with COVID-19-related VTE. Our study also found that the elevated level of GZMA among VTE patients was closely associated multi-immune cell infiltration. The consequence was consistent with an early study that GZMA could act as a pro-inflammatory mediator by cytokine activity \(^{36}\).
In addition, extracellular GZMA promotes the production of IL-6, IL-8, and TNF-α in human peripheral blood mononuclear cells and purified monocytes without the presence of perforin 37.

Despite the common triggers for a hypercoagulable state and venous thrombosis, COVD-19 was also a potential risk factor for bleeding and DIC 38. GZMA could contribute to the enhanced activation of the coagulation cascade in sepsis by generating cytokines accompanied with endothelial cell activation and coagulation like TNF-α or IL-6 39. It has been reported that the presence of proinflammatory cytokines such as IL-6, IL-8, and TNF-α (all of them induced by GZMA) leads to the liberation of large VWF multimers that is a potent platelet aggregator 40. Curiously, GZMA was discovered to interact with the thrombin receptor (PAR1) on the surface of platelets and block thrombin-mediated responses. GZMA, this protease can also degrade some proteins of the ECM like fibronectin or collagen IV and could be released by NK cells during sepsis 41. In a word, GZMA can act as extracellular proteases that regulate the inflammatory response irrespectively of its ability to induce cell death. Indeed, studies in animal models proved that GZMA is involved in the cytokine release syndrome characteristic of sepsis. GZMA family also could regulate other biological processes involved in sepsis pathophysiology like the coagulation cascade, platelet function, endothelial barrier permeability, and the immunosuppressive stage of sepsis 40.

BCL2A1 (B-cell lymphoma 2-related protein A1)

BCL2A1 belonging to the Bcl-2 family members, rigorously regulates cell endogenous apoptosis, and targets antiapoptotic members. BCL2A1 is particularly prominent in the hematopoietic system and exerts its antiapoptotic function by sequestering proapoptotic B-cell lymphoma 2 (BCL2) proteins. BCL2A1 is a highly regulated nuclear factor κB (NF-κB) target gene that exhibits pro-survival functions 42. Despite being a critical role during tumorigenesis, BCL2A1 has also related to inflammation response. Selma Olsson Kefeldt showed that the therapeutic compound BCL2A1 might provide a new strategy in chronic inflammatory conditions associated with bone loss 43. In addition, Jun Li demonstrated that BCL2A1 has potential diagnostic and prognostic value for sepsis. The relationship between BCL2A1 and COVID-19 has also been reported that Hibah Shaath revealed that the serum level of BCL2A1 was only seen in bronchoalveolar lavage (BAL) from severe-COVID-19 patients 44. Our study first discovered that the serum level of BCL2A1 could serve as a potential biomarker for COVDI-19-related VTE.

CD52 (Campath-1 antigen)

CD52, also called as Campath-1 antigen, is a small surface glycoprotein composed of 12 amino acids, and widely expressed on the cell surface of immune cells, such as mature lymphocytes, natural killer cells (NK), eosinophils, neutrophils, monocytes/macrophages, and dendritic cells (DCs). Ligation of cell surface CD52 molecules may trigger costimulatory signals for T-cell activation and proliferation. However, soluble CD52 molecules will interact with the inhibitory sialic acid-binding immunoglobulin-like lectin 10 (Siglec10) to clearly inhibit T cell proliferation and activation 45. CD52 function is not fully understood, although experiments with anti-CD52 antibodies have shown that CD52 is crucial for lymphocyte trans-
endothelial migration and contributes to the co-stimulation of CD4+ T cells and T-cell activation and proliferation. The wide existence of CD52 on the surface of a broad spectrum of immune cells makes it a therapeutic target, especially in immune-mediated diseases. Xinru Qiu reported that CD52 is a biomarker and a therapeutic target for sepsis due to its dynamic expression in lymphocytes and correlation with improved sepsis outcome46. Pedram Shafiee-Jahani reported that CD52 depleting antibody of Alemtuzumab is a potential therapeutic target for the treatment of pulmonary inflammation, abrogation of eosinophilia, improvement of lung function, and thus therapy of allergic airway hyperreactivity (AHR)47. Moreover, CD52 was closely associated with virus replication and attack. Dhriti Chatterjee discovered that mouse hepatitis virus infection upregulated genes involved in innate immune responses such as CD11b, CD74, CD52, and CD6848. Upasana Parthasarathy reported that damage-associated molecular patterns (DAMPs) played a key role in the pathology of severe COVID-19 including Siglecs and their cognate ligands CD24 and CD52, in COVID-19. while soluble CD52 binding to Siglec-10 inhibits T cell receptor-associated kinase phosphorylation and T cell activation49,50. Treatment with the anti-human CD52 antibody of alemtuzumab can relieve the symptoms of COVID-19 and may also act through Siglec stimulation51.

RNASE2

The eosinophil-derived neurotoxin (EDN/ RNASE2) is one of the four major secretory proteins detected in the specific granules of the human eosinophilic leukocyte. The divergent orthologs, the mouse eosinophil-associated RNases (mEars), are prominent secretory proteins of eosinophilic leukocytes and are all members of the larger family of RNase A-type ribonucleases52. Although the well-known antiviral activity of EDN has been widely accepted, targeting RNA viruses via mechanisms that may require enzymatic activity, more recent studies have elucidated how these RNases may generate host defence via roles in promoting leukocyte activation, maturation, and chemotaxis. EDN and its more highly charged and cytotoxic paralog, the eosinophil cationic protein (ECP/RNase 3) are released from eosinophil granules when these cells are stimulated with cytokines and other proinflammatory mediators53. Recent interest in EDN has focused on its role as a specific biomarker in eosinophil-associated pathophysiology, including asthma exacerbations54, cow’s milk allergy55, eosinophilic esophagitis56, and vaccine-induced aberrant responses57. Our study first reported the close association about RNASE2 and COVID-19 associated-VTE.

Conclusions

Based on our bioinformatic analysis, similar changes occurred in CVID-19-associated VTE. These common changes were mainly enriched in NF-kappa B signalling pathway and TNF signalling pathway. Moreover, we also identified GZMA, BCL2A1, CD52, and RNASE2 were potential diagnostic markers of COVID-19-related VTE, and these genes were closely associated with viral replication and immune cell proportion. Our study provides a new prospect for the diagnosis and prediction of COVID-19 associated with VTE.
Declarations

Acknowledgments

We thank the GEO for providing available data for COVID-19 and VTE.

Author contributions

ZJ-C, and HH-C designed the research. ZJ-C and SC-X collected the primary data. ZJ-C and SC-X performed the bioinformatics analysis. ZJ-C and SC-X wrote the primary manuscript. HH-C wrote and revised the final manuscript. All authors listed have read and approved it for publication.

Conflicts of Interest

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

Funding Statement

None.

Data Sharing and Data Accessibility

The raw data used in this study are available in the public GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Conflict of Interest The authors declare that they have no conflict of interest.

References


receptor Siglec-10, Nature immunology 14, 741-748.


### Table 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Tissue</th>
<th>GEO dataset</th>
<th>Platform</th>
<th>Number of samples (Disease/healthy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COVID-19</td>
<td>Peripheral blood</td>
<td>GSE152418</td>
<td>GPL24676</td>
<td>34(17/17)</td>
</tr>
<tr>
<td>COVID-19</td>
<td>Peripheral blood</td>
<td>GSE164805</td>
<td>GPL26963</td>
<td>15(10/5)</td>
</tr>
<tr>
<td>COVID-19</td>
<td>Peripheral blood</td>
<td>GSE166253</td>
<td>GPL20795</td>
<td>26(16/10)</td>
</tr>
<tr>
<td>COVID-19</td>
<td>Peripheral blood</td>
<td>GSE171110</td>
<td>GPL16791</td>
<td>54(44/10)</td>
</tr>
<tr>
<td>VTE</td>
<td>Peripheral blood</td>
<td>GSE48000</td>
<td>GPL10558</td>
<td>132(25/107)</td>
</tr>
<tr>
<td>VTE</td>
<td>Peripheral blood</td>
<td>GSE19151</td>
<td>GPL571</td>
<td>133(63/70)</td>
</tr>
</tbody>
</table>
**Figures**

Two-dimensional PCA cluster plot of the COVID-19 datasets (GSE152418, GSE164805, GSE166253, GSE177110) and VTE datasets (GSE 19151, GSE48000) before and after sample correction.

---

**Figure 1**

Two-dimensional PCA cluster plot of the COVID-19 datasets (GSE152418, GSE164805, GSE166253, 171110) and VTE datasets (GSE 19151, GSE48000) before and after sample correction.
Figure 2

WGCNA. (A-C) Analysed network topology for the relevant soft thresholding powers. (B-D) Clustered dendrogram of genes. (E) Heatmap of the module-trait relationships. (F) common intersection genes of both COVID-19 and VTE.
Figure 3

(A) Gene ontology (GO) enrichment analyses of common critical genes. (C-D) KEGG analysis of the common critical genes in COVID-19 and VTE.
Figure 4

(A) The intersection of different genes from WGCNA analysis and VTE dataset. (B-D) The PPI network of intersected genes and hub genes identified by the cystoscope software.
Figure 5

The diagnostic value of the four hub genes in the GSE19151 dataset. (A-D) the expression level of the four critical genes. (E-H) The receiver operating characteristic (ROC) analysis of the GZMA, BCL2A1, CD52, and RANSE2 in GSE19151 dataset.
Figure 6

Assessment and visualized analysis of immune cell infiltration. (A) The results of immune cell infiltration in the VTE and control group by CIBERSORT. (B) Heat map of 19 types of immune cell connections. (C) Violin diagram of the proportion of 19 types of immune cells.
Figure 7

Analysis of correlation between hub genes and immune cells. (A) GZMA, (B) BCL2A1, (C) CD52, (D) RNASE2.
Figure 8

Legend not included with this version.

Supplementary Files

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