Co-expression analysis to identify key modules and hub genes associated with COVID19 in Platelets

Ahmed B. Alarabi¹, Attayeb Mohsen², Kenji Mizuguchi^{2,3}, Fatima Z. Alshbool¹, Fadi T. Khasawneh⁴

¹Department of Pharmacy Practice, Irma Lerma Rangel College of Pharmacy, Texas A&M University, Kingsville, Texas, USA. ²Laboratory of Bioinformatics, Artificial Intelligence Center for Health and Biomedical Research (ArCHER), National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki, Osaka 567-0085, Japan. ³Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka, Japan. ⁴Department of Pharmaceutical Sciences, Irma Lerma Rangel College of Pharmacy, Texas A&M University, Kingsville, Texas, USA.

ABSTRACT

The severe acute respiratory syndrome corona virus 2 (SARS-CoV-2) is a highly contagious virus that causes a severe respiratory disease known as Corona virus disease 2019 (COVID19). Indeed, COVID19 increases the risk of cardiovascular occlusive/thrombotic events and is linked to poor outcomes. The pathophysiological processes underlying COVID19-induced thrombosis are complex, and remain poorly understood. To this end, platelets play important roles in regulating our cardiovascular system, including via contributions to coagulation and inflammation. There is an ample of evidence that circulating platelets are activated in COVID19 patients, which is a primary driver of the thrombotic outcome observed in these patients. However, the comprehensive molecular basis of platelet activation in COVID19 disease remains elusive, which warrants more investigation. Hence, we employed gene co-expression network analysis combined with pathways enrichment analysis to further investigate the aforementioned issues. Our study revealed three important gene clusters/modules that were closely related to COVID19. Furthermore, enrichment analysis showed that these three modules were mostly related to platelet metabolism, protein translation, mitochondrial activity, and oxidative phosphorylation, as well as regulation of megakaryocyte differentiation, and apoptosis, suggesting a hyperactivation status of platelets in COVID19. We identified the three hub genes from each of three key modules according to their intramodular connectivity value ranking, namely: COPE, CDC37, CAPNS1, AURKAIP1, LAMTOR2, GABARAP MT-ND1, MT-ND5, and MTRNR2L12. Collectively, our results offer a new and interesting insight into platelet involvement in COVID19 disease at the molecular level, which might aid in defining new targets for treatment of COVID19-induced thrombosis.

Keywords– Platelets, COVID19, WGCNA.

key points

- Co-expression analysis of platelet RNAseq from COVID19 patients show distinct clusters of genes (modules) that are highly correlated to COVID19 disease.
- Identifying these modules might help in understanding the mechanism of thrombosis in COVID19 patients

INTRODUCTION

The coronavirus SARS-CoV-2 is a highly contagious infection that causes a severe respiratory disease known as COVID19. This disease that has reached a pandemic level, is impacting tens of millions of people worldwide. In the United States, there are around 34 million reported cases, over two million hospital admissions, and half a million deaths as of July 2021¹. It is now known that COVID19-induced thrombosis increases the incidence of cardiovascular occlusive events in infected patients, a fact that has been reported in several studies²⁻⁴. Indeed, abnormal hemostasis responses were observed in COVID19 hospitalized patients, which was linked to poor prognosis^{2,5,6}. In addition, studies have shown that COVID19 leads to increase in platelet activation through alterations of platelet transcriptome and proteome^{7,8}. In this connection, it is now well established that platelets play roles beyond vascular hemostasis, including innate immunity and tumor metastasis⁹. Moreover, platelets

were shown to be activated in the septic state, and antiplatelet therapy has been used as a strategy to prevent organ damage in sepsis¹⁰. To this end, evidence have indicated that viral infections are associated with coagulation disorders, and thrombotic cardiovascular events^{11,12}, which is consistent with the thrombotic phenotype seen in COVID19 patients/SARS-CoV-2 viral infection. While there has been some progress, our understanding of the pathways that govern platelet participation in COVID19-induced thrombosis remains limited, but clearly warrants investigation.

To obtain a comprehensive insight into the pathogenesis of specific disease states, several computational and research methods have been developed¹³. Some of these approaches were employed to examine the potential gene networks, which are very instrumental to guide understanding of diseases and their mechanistic pathways. Notably, co-expression analysis is one such approach, which clusters genes into coexpressed groups known as modules. These genes that belong to the same module are thought to share functional properties¹⁴. This approach relies

on using graph theory concepts that allow researchers to understand in a systematic way the relations between the genes of a module and the phenotype based on the module eigengene¹⁴. In fact, co-expression using weighted correlation network analysis (WGCNA) has been used for analyzing a number of biological processes, including cancer^{15,16} and cognitive and mental disorders^{17,18}. In short, gene networks provide the utility to move beyond individual-gene comparisons and comprehensively identify biologically meaningful relationships between gene products and phenotypes.

Previous studies on the mechanisms of thrombosis in COVID19 disease have primarily concentrated on specific pathophysiological functions, with relatively fewer studies identifying comprehensive regulatory networks. Therefore, in the present study, WGCNA was used to determine gene networks associated with COVID19 disease in platelets. PRJNA634489 dataset which contained a total of 15 samples from COVID19 patients and health controls⁷ was used in the present study. Three modules with the highest level of significance in correlation with COVID19 disease were identified, and the three genes with the highest intramodular connectivity were selected as the hub genes in the respective modules for COVID19. Gene enrichment analysis was also conducted to determine enrichments in the key modules. The results of this study may provide novel information/insight into the underlying mechanisms of COVID19 disease and may assist in the identification of potential biomarkers for diagnosis and targets for treatment.

METHODS

Data Preprocessing and Differentially Expressed Genes Screening

RNAseq data were downloaded from BioProject accession #PRJNA634489⁷. Data comprised of ten COVID19 patients in addition to age- and sex-matched five healthy controls. The Kallisto program was employed for pseudoalignment of reads and quantification to obtain the counts and the transcript per million (TPM)¹⁹. Log2CPM (log transformed counts per million) was used for the differential expression analysis by employing Voom normalization²⁰ and Limma R package²¹ TPM normalized and filtered to exclude low variance transcripts (≤ 0.001)²² was used for the weighted gene co-expression network analysis.

Weighted Gene Coexpression Network Analysis

The weighted co-expression network was produced using R package "WGCNA"¹⁴. To weight highly correlated genes, the soft thresholding power (β) was set at 12, and the minimal module size was set at 30. To define clusters of genes in the dataset, the adjacency matrix was used to calculate the topological overlap matrix (TOM), which shows the degree of overlap in shared neighbors between pairs of genes in the network. The resulting gene network was visualized as a heatmap.

Screening for key modules and hub genes

Correlation between module eigengenes and the COVID19 status was calculated to identify key modules that have significant correlation. The correlation values were displayed within a heatmap. The modules that correlated with COVID19 most significantly were considered as the key modules. Gene significance

(GS) was defined as the correlation between gene expression and the COVID19 status. Module membership (MM) was defined as the correlation between gene expression and each module's eigengene, and intramodular connectivity (K.in), which measures how connected a given gene with respect to the genes of a particular module, was also calculated using WGCNA. Subsequently, the correlation between GS and MM as well as GS and K.in were examined to verify module-COVID19 status associations. The correlation analyses in this study were performed using the Pearson correlation as described in the "WGCNA" package¹⁴. All module genes were ranked according to their intramodular connectivity, and only the top three genes were selected as hub genes.

Functional enrichment analysis of key modules

The genes in each key modules were extracted from the network and enrichment analysis was performed to further explore the functions of the respective modules. Targetmine²³ which is a web-based integrative data analysis platform for target prioritisation and broad-based biological knowledge discovery- was used to perform Gene Ontology (GO) and Reactome pathway enrichment analysis. In this analysis, a benjamini hochberg adjusted P-value of 0.05 was set as the significance threshold to identify the most significant functional pathways/GO terms. Only top results of enriched terms are reported.

Statistical and visualization tools

We used R statistical programming language²⁴ version 4.1.0, with the following packages: "WGCNA"¹⁴ for coexpression analysis; "circlize"²⁵ for chord diagram building; "ggplot2" for visualization²⁶; "lgraph" for network analysis²⁷ and "ggraph"²⁸ for network visualization.

RESULTS

Construction of Co-expression Network

The transcript per million (TPM) gene expression dataset were filtered based on variance, and 7119 genes in the 15 samples of ten COVID19 patients and five healthy controls were used to construct the co-expression network. The results of cluster analysis of the samples are demonstrated in (Fig. 1A). To construct the network, a soft-threshold of 12 was used to obtain the approximate scale-free topology (Supplementary Fig. 1). Genes across the 15 samples were hierarchically clustered based on topological overlap (Fig. 1C & 1D). We identified 16 modules in which genes are coexpressed, random colors were assigned to the modules to distinguish between them. The size (number of genes/module) of each module is presented in (Fig 1B). To demonstrate how these modules were relatively distinctive, we plotted the network heatmap of 400 randomly selected genes based on topological overlap matrix dissimilarity and their cluster dendrogram (Fig. 2A) indicating relative independence among these clusters.

Correlation between modules and COVID19 disease status

To examine the relation of COVID19 status with the emerged modules, we built the eigengene adjacency matrix by calculating the correlation of the eigengenes matrix after inserting COVID19 status to the matrix. The heatmap (Fig. 2B) showed the modules'

Figure 1: Construction of weighted co-expression network. (A) Sample dendrogram and trait heatmap. (B) Module size. (C and D) Cluster dendrogram block 1 and block 2. Each color represents one specific co-expression module, and branches above represent genes.

relationship and the correlation between the modules namely black, cyan, yellow, blue, and magenta and COVID19 status.

Identification of key modules in relationship to COVID19 disease status

To further determine the closest modules to COVID19 status, we re-clustered the eigengenes using single linkage method with absolute correlation as a distance function; the single linkage clustering algorithm looks for closest modules to form a cluster, then cluster them with the next nearest module progressively until one cluster is formed²⁹. As demonstrated in Fig. 2C, the closest three modules to COVID19 status are magenta, yellow and black. Three essential measurements can help confirm the importance of the module to a specific trait, 1) Module membership (MM), which increases for a particular gene, when the module eigengene accurately represents this gene, 2) gene significance (GS) is measured by calculating the correlation of gene expression with the specific trait and 3) intramodular connectivity (K.in) for a gene within the module, reflecting the centrality of the gene to the module expression network. Based on WGCNA, if a gene is higher with GS, MM, and K.in, it is more meaningful to the clinical trait of interest^{30,31}. Explicitly, the higher the correlation between gene significance of genes in a module and their module membership, the higher its importance. Similarly, when the gene centrality in the network increases in parallel with gene significance, that also is strong evidence that key modules are essential in that trait. The correlations between gene significance and module membership as well as between gene significance and intramodular connectivity show that yellow, black, and magenta modules have the highest correlation values with a substantial difference to the

next nearest module (Blue $R=0.61$) (Supplementary Fig. 2). For those reasons, we selected yellow, black, and magenta modules for further investigation and will refer to them using the term key modules.

Key modules show high correlation to COVID19 disease status

The module-trait relationship was determined by correlating module eigengenes with COVID19 disease status to identify significant correlation. The yellow and the black modules exhibited the highest positive correlation ($R=0.91$; $p\text{-value}=3 \times 10^{-6}$, and $R=0.86$; $p\text{-value}=3 \times 10^{-5}$, respectively; Fig. 2D). On the other hand, the magenta module ($R=-0.96$; $p\text{-value}=1 \times 10^{-8}$) exhibited the highest negative correlation (Fig. 2D). Therefore, these three modules were identified as key modules for COVID19 disease and its impact on platelets. The significant correlations between the different GS, MM, and K.in for COVID19 are illustrated in (Fig. 3A & 3B). We also showed the GS, MM, and K.in of the green module that showed the low correlation to COVID19 disease status (Fig. 3A & 3B).

Gene hub detection and visualization of module networks.

Genes in the selected key modules were ranked according to the intramodular connectivity and the top 20 genes of each key modules was used to visualize the network of each specific module (Fig. 4). Subsequently, the top three genes of the yellow, black, and magenta modules were labeled as the hub genes in their modules that are important for COVID19 disease. Thus, the protein coding genes COPE, CDC37 and CAPNS1 were selected as the hub genes in the yellow module, whereas AURKAIP1, LAMTOR2, and GABARAP protein coding genes

Figure 2: Co-expression module analysis. (A) Interaction of co-expression genes based on TOM dissimilarity and the cluster dendrogram of 1,000 randomly selected genes. The colors of the axes represent respective modules. The intensity of the yellow inside the heatmap represents the degree of overlap, with a darker yellow representing an increased overlap. (B) Eigengene adjacency heatmap of different gene co-expression modules. (C) An eigengene dendrogram identified groups of correlated modules. (D) Heatmap of the correlation between COVID19 status and module eigengenes. Column corresponds to a clinical trait, and each row corresponds to a module. Each row contains the correlation coefficients which correspond to the cell color; green represents negative correlation and red represents positive correlation. The P-values are stated in the brackets.

Gene	LogFC	AveExpr	Adj.P.Val	Module
MT-ND5	-3.2	14.0	2.12×10^{-5}	magenta
MT-ND1	-3.1	15.2	2.32×10^{-6}	magenta
MTRNR2L12	-3.1	15.5	5.83×10^{-7}	magenta
AURKAIP1	1.6	3.1	4.29×10^{-3}	black
GABARAP	1.2	6.4	2.75×10^{-3}	black
COPE	0.9	4.5	3.04×10^{-2}	yellow
LAMTOR2	0.9	3.4	4.08×10^{-2}	black
CAPNS1	0.6	8.3	1.14×10^{-1}	yellow
CDC37	0.6	6.9	1.33×10^{-1}	yellow

Table 1: Differential expression of hub genes identified in the key modules. (LogFC: log2 transformed fold change; AveExpr: Average expression; Adj.P.Val: Adjusted P-value).

were selected as the hub genes in the black module. Regarding the magenta module, MT-ND1, MT-ND5, and MTRNR2L12 are selected as hub genes. All of these hub genes exhibited a high intramodular connectivity, which established their network centrality and potentially vital roles in the COVID19 disease. We also observed that not all of hub genes have shown differential gene expression (Table 1).

Enrichment Analysis of key modules

Gene ontology (GO) pathway enrichment analyses were performed on the yellow, black, and magenta modules using Targetmine platform, and the top relevant terms of each category are presented in (Fig. 5A). The pathway enrichment results demonstrated that the genes in both yellow and black modules were primarily enriched in pathways associated with metabolic process, protein translation, energy substance metabolism, mitochondrial activity, and oxidative phosphorylation. Genes in the magenta module were enriched in several pathways that are primarily associated with regulation of megakaryocyte differentiation and apoptosis, including the regulation of the execution phase of apoptosis. Reactome showed enriched pathways of metabolism, platelet degranulation, and response to elevated platelet cytosolic Ca^{2+} in the yellow module. The black module shows enrichment of respiratory electron transport, ATP synthesis by chemiosmotic coupling, heat production by uncoupling proteins, citric acid (TCA) cycle, and respiratory electron trans-

Figure 3: Module features of GS, MM and K.in. (A) Module Membership (MM) and Gene significance (GS) are significantly correlated in the key modules (magenta, black, and yellow). Each point represented an individual gene within each module, which was plotted by GS on the y-axis and MM on the x-axis. The regression line, correlation value, and p-value were shown for each plot. (B) Correlation of the K.in (x-axis) and the GS (y-axis). Green module is not a key module and was added here for sake of comparison.

port just to name a few (Fig. 5B) (More detailed results are shown in supplementary Fig. 3).

DISCUSSION

The underlying pathophysiological mechanisms of thrombosis in COVID19 are extremely complicated³², and hence clearly re-

Figure 4: Interaction of gene co-expression patterns in the key identified module and hub gene abundance. The module was visualized using R package "ggraph" software. The node size corresponds to the K.in level, and the thickness of the link represents the strength of correlation between genes. For sake of visualization clarity, edges of weight less than 0.6 were not drawn.

in the present study, gene co-expression analysis was performed on COVID19 platelet RNAseq dataset containing gene expression data from ten COVID19 patients and five healthy controls. There were three modules that were identified as the key modules in COVID19, with the highest level of significant association. The top three genes of each key module with the highest intramodular connectivity were identified as hub genes for COVID19 in platelets. The results of the enrichment analysis suggest that the key modules and the pathological processes underlying the disease are associated with energy metabolism, mitochondrial process, and apoptosis. Furthermore, we also saw enrichment of platelet secretion and activation pathways. These results provide at least in part- an illustration of the comprehensive platelet regulatory network in COVID19, which should improve the current understanding of the mechanisms underlying immunothrombosis in COVID19 patients. Ultimately, these findings might help in finding appropriate therapeutic targets.

The present study used the data in BioProject accession #PRJNA634489⁷ to perform the co-expression analysis using WGCNA. The data used in this study, which was generated by Manne et al⁷, revealed that COVID19 disease leads to changes in platelet transcriptional profiles in comparison to control. Manne et al showed that platelet differential gene expression in COVID19 is associated with enrichment of protein ubiquitination, antigen presentation, and mitochondrial dysfunction. The major differences in the genes or modules obtained in the present study, compared with the results from other studies including the one by Manne et. al.⁷ is that the present study used a more comprehensive method by employing WGCNA. Using this method, we were able to identify/pull-out co-expression modules of genes, namely the yellow and black modules, which represent an important regulatory modules of platelet function in COVID19. In addition, we were able to identify the magenta module, which represents genes that are negatively correlated with COVID19 disease state and show enrichment in megakaryocyte differentiation and apoptotic pathways. This systematic and in-depth analysis performed should complement results obtained from conventional DEGs analysis, and therefore allow for a better understanding of the pathophysiological mechanisms of COVID19 disease.

Notably, the co-expression analysis revealed a total cluster of 16 modules, with the yellow and black modules exhibiting the strongest positive correlation with COVID19 disease. Enrichment analysis indicated that the genes in the yellow and black modules were primarily associated with platelet metabolism, energy, and oxidative phosphorylation. Furthermore, the analysis of the yellow module showed enrichment of a host of platelet functional responses/activities, such as platelet degranulation/secretion and increased platelet response to Ca²⁺. Indeed, other studies showed the COVID19 disease to be associated with platelet activation and to increase platelet alpha granule secretion, which are critical in the development of thrombosis seen in those patients^{7,8}. It is noteworthy that the platelet alpha granule secretion response is not only important for thrombus formation, but also in inflammation by releasing receptors that facilitate adhesion of platelets with other vascular cells as well as releasing a wide range of inflammatory chemokines³³.

The yellow and black modules show strong enrichment in platelet metabolic processes, which is in agreement with the in-

quire more examination. Inspecting gene co-expression patterns is proven to be an effective method to analyze and uncover complicated genetic networks. To address the aforementioned issues,

A

B

Figure 5: Gene ontology (GO) terms and Reactome enrichment analysis of the key modules. Chord plots depicting the enriched top significant GO terms of biological process (A), and top significant Reactome pathways (B) in the key modules. Thickness of the connection is corresponding to the number of involved genes.

crease in platelet activation. For instance, previous data have shown that platelet transition from inactive to active state requires alteration in ATP availability³⁴, furthermore, substrate metabolism (e.g. glucose) was shown to be essential for platelet activation³⁵, and thrombosis³⁶. This seems to suggest that altered platelet metabolism may play a critical role in the pathophysiology of thrombosis in COVID19 patients. It is important to note that reports have suggested that a state of hypermetabolic demand is one of COVID19 disease features, especially when sepsis develops³⁷. Like other viruses that can impact cellular metabolism in human cells and utilize them to their advantage, SARS-CoV-2 virus appears to have the ability to localize proteins to mitochondria and hijack the host's mitochondrial function³⁸. This mechanism might explain the enrichment of platelet mitochondrial processes we observed in the yellow and black modules. This finding is in fact supported by a recent study that reported that SARS-CoV-2 impacts mitochondria in platelets, which affects their involvement in the pathophysiology of thrombosis in COVID19 patients³⁹. The enrichment of protein translation in the yellow and black modules suggests an alteration in protein synthesis and possible hijacking of the translation machinery of platelet by the virus. In line with this observation, one study suggested that the cells infected with SARS-CoV-2 might exhibit a faster protein synthesis rate, which implies a high translation rate⁴⁰. This notion requires further investigation to determine the exact mechanism underlying enhancement of translation in platelets of COVID19 patient.

One particular characteristic of platelet apoptotic process is the phosphatidylserine (PS) exposure, which is essential for the generation of thrombin⁴¹. PS exposure is found to be downregulated in activated platelets from COVID19 patients due to mitochondrial dysfunction³⁹. This observation is supported by the negative regulation of apoptotic processes in platelets enrichment in the negatively correlated magenta module. On the contrary, another report showed that COVID19 increases PS externalization, which is linked to thrombosis⁴². The impact of platelets mitochondrial damage on hemostasis depends on its severity. Thus, it leads to bleeding by progressing toward apoptosis if it is severe; or toward platelet activation pathways and development of thrombosis risk in case of mild damage⁴³. Based on this reasoning, COVID19 disease-caused mitochondrial damage in platelets is probably mild; and hence the thrombotic phenotype still prevails in these patients. Based on these considerations, more investigation is needed to confirm these observations and to understand the underlying mechanisms.

We identified hub genes in each of the key modules. For example, in the yellow module the COPE, CDC37, and CAPNS1, which are protein coding genes involved in vesicle-mediated transport, positive regulation of cellular processes, and regulation of interferons. Furthermore, some of these protein coding genes have also been investigated in platelets and shown to regulate important aspects of their function⁴⁴⁻⁴⁷. Interestingly, although our co-expression analysis showed that CAPNS1 is an important hub gene in the yellow module, this gene was not differentially expressed in our differential gene expression analysis. Furthermore, CAPNS1 has shown to play a significant role in regulating platelet activity and thrombosis under hypoxia⁴⁶, a condition commonly seen in severe COVID19 patients⁴⁸. This

observation might indicate that some of the important genes in establishing thrombotic phenotype in COVID19 may not necessarily be differentially expressed.

The hub genes of the black module, AURKAIP1, LAMTOR2, and GABARAP are linked to regulation of mitochondrial activity, regulation of signaling processes, and protein targeting. Data on the role of these genes in platelets is limited, however, further investigation is warranted based on our findings. It is noteworthy that LAMTOR2 is a known regulator of the MAPK/Erk and mTOR signaling pathways^{49,50}, both of which are shown to be important in regulating platelet function^{51,52}. Moreover, the p14/LAMTOR2 deficiency- which is associated with one of the primary immunodeficiency diseases that also include "Hermansky-Pudlak syndrome type 2"- has been linked to platelet defects⁵³. However, more needs to be done to examine the exact role of LAMTOR2 in platelets of COVID19 patients.

In the magenta module, MT-ND1, MT-ND5, and MTRNR2L12 protein coding genes are related to NADH dehydrogenase activity and apoptotic processes. According to our analysis, all hub genes in the magenta module have differentially expressed and are downregulated in COVID19 patients in comparison to healthy controls. Down regulation of MT-ND1, MT-ND5 protein coding gene might at least in part explain the mitochondrial dysfunction seen in platelets of COVID19 patients. With respect to MTRNR2L12, it was observed that it is one of the differentially expressed genes in bronchoalveolar lavage fluid samples from patients with severe COVID19 in comparison to control⁵⁴. MTRNR2L12 is a paralog of the protein coding gene MTRNR2L8, and both are expressed in platelets⁵⁵. It is of note that MTRNR2L12 was shown to be among the top 10 RNA with differential splice junctions in platelets of patients of multiple sclerosis⁵⁶.

In addition to the identified hub genes, a number of other canonical platelet genes in the yellow and black modules were also associated with platelet function. For example, SLEB and ITGA2B protein coding genes were present in the yellow module with high intramodular connectivity (ranked in the top 50) and both proteins are critical for platelet function⁵⁷. Moreover, another canonical platelet gene that was also identified in the yellow module, namely ITGB3 was ranked 132 with regard to its intramodular connectivity, which is considered high in the yellow module of 681 genes. Furthermore, we also noticed that the protein coding gene IFITM3 shows high module membership (black module). The protein encoded by this gene is an interferon-induced membrane protein that was shown to be helpful in immunity against influenza A H1N1 virus, West Nile virus, and dengue virus⁵⁸⁻⁶⁰. Most recently, IFITM3 has also been shown to be upregulated protein in COVID19 disease^{7,61}.

The present study has certain limitations that should be noted. Firstly, the analysis focused on only one dataset, due to limited access to platelet gene expression data that were collected from COVID19 patients. Therefore, additional datasets should be analyzed, if available, to validate our findings and/or obtain more representative results. Also, the number of samples was 15, which may be associated with some noise, albeit it is the minimum number of samples recommended for co-expression analysis by WGCNA. Finally, any limitations in the original study, from which the data was obtained will also be reflected in the results of this study.

In conclusion, our co-expression analysis of a platelet RNAseq dataset from COVID19 patients and healthy controls revealed 16 modules, amongst which the yellow, black, and magenta were identified as the most critical in COVID19 disease. Additionally, 9 hub genes were determined to potentially serve key roles in pathophysiological mechanisms of COVID19 in the context of platelet biology. The positively associated yellow and black modules were identified to be involved in platelet degranulation, energy metabolism, and mitochondria. The negatively associated magenta module was associated with interactive pathways of apoptosis. These data should help expand our understanding of the underlying mechanisms of thrombosis in COVID19 disease and help promote and guide future experimental studies to investigate the roles of the protein coding genes in the pathophysiology of this disease. Additionally, these genes may serve as novel therapeutic targets for treating patients.

ACKNOWLEDGEMENTS

The authors acknowledge Matthew T. Rondina and Robert A. Campbell from the department of internal medicine, University of Utah, Salt Lake City, Utah, USA, for their contribution and suggestions that improved the work substantially. The statements made here are solely the responsibility of the authors. No external funding was utilized in this study.

AUTHORSHIP

Contribution: AA, AM: conceptualization; AA, AM: analyzed results and made the figures; AA: drafted the manuscript; AA, AM, KM, FA, FK: edited the manuscript and approved the submission; Conflict of interest disclosure: Authors declare no conflict of interest related to this work.

ORCID profiles: AA: 0000-0002-4278-5067; AM: 0000-0003-0690-8012; KM: 0000-0003-3021-7078; FA: 0000-0002-2765-3250; FK: 0000-0002-4027-2997.

Correspondence: Fadi Khasawneh: Department of Pharmaceutical Sciences, Irma Lerma Rangel College of Pharmacy, Texas A&M University, Kingsville, Texas, USA, fkhassawneh@tamu.edu.

REFERENCES

- [1] CDC. COVID Data Tracker Weekly Review; 2021. (visited 2021-08-06). Available from: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>.
- [2] Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *Journal of Thrombosis and Haemostasis*. 2020 Apr;18(4):844–847. Available from: <https://doi.org/10.1111/jth.14768>.
- [3] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. 2020 Feb;395(10223):497–506. Available from: [https://doi.org/10.1016/s0140-6736\(20\)30183-5](https://doi.org/10.1016/s0140-6736(20)30183-5).
- [4] Jie Guan W, Yi Ni Z, Hu Y, Hua Liang W, Quan Ou C, Xing He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*. 2020 Apr;382(18):1708–1720. Available from: <https://doi.org/10.1056/nejmoa2002032>.
- [5] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*. 2020 Mar;395(10229):1054–1062. Available from: [https://doi.org/10.1016/s0140-6736\(20\)30566-3](https://doi.org/10.1016/s0140-6736(20)30566-3).
- [6] Bowles L, Platten S, Yartey N, Dave M, Lee K, Hart DP, et al. Lupus Anticoagulant and Abnormal Coagulation Tests in Patients with Covid-19. *New England Journal of Medicine*. 2020 Jul;383(3):288–290. Available from: <https://doi.org/10.1056/nejmc2013656>.
- [7] Manne BK, Denorme F, Middleton EA, Portier I, Rowley JW, Stubben C, et al. Platelet gene expression and function in patients with COVID-19. *Blood*. 2020 Sep;136(11):1317–1329. Available from: <https://doi.org/10.1182/blood.202007214>.
- [8] Zaid Y, Puhm F, Allaey I, Naya A, Oudghiri M, Khalki L, et al. Platelets Can Associate With SARS-CoV-2 RNA and Are Hyperactivated in COVID-19. *Circulation Research*. 2020 Nov;127(11):1404–1418. Available from: <https://doi.org/10.1161/circresaha.120.317703>.
- [9] Smyth SS, McEver RP, Weyrich AS, Morrell CN, Hoffman MR, Arepally GM, et al. Platelet functions beyond hemostasis. *J Thromb Haemost*. 2009 Nov;7(11):1759–1766.
- [10] Bashour TT, Myler RK, Andreae GE, Stertz SH, Clark DA, Ryan CJ. Current concepts in unstable myocardial ischemia. *Am Heart J*. 1988 Apr;115(4):850–861.
- [11] Goeijenbier M, van Wissen M, van de Weg C, Jong E, Gerdes VEA, Meijers JCM, et al. Review: Viral infections and mechanisms of thrombosis and bleeding. *Journal of Medical Virology*. 2012 Aug;84(10):1680–1696. Available from: <https://doi.org/10.1002/jmv.23354>.
- [12] Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, et al. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. *New England Journal of Medicine*. 2018 Jan;378(4):345–353. Available from: <https://doi.org/10.1056/nejmoa1702090>.
- [13] Cai Y, González JV, Liu Z, Huang T. Computational Systems Biology Methods in Molecular Biology, Chemistry Biology, Molecular Biomedicine, and Biopharmacy. *BioMed Research International*. 2014;2014:1–2. Available from: <https://doi.org/10.1155/2014/746814>.
- [14] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008 Dec;9(1). Available from: <https://doi.org/10.1186/1471-2105-9-559>.
- [15] Tian Z, He W, Tang J, Liao X, Yang Q, Wu Y, et al. Identification of Important Modules and Biomarkers in Breast Cancer Based on WGCNA. *Oncotargets and Therapy*. 2020 Jul;Volume 13:6805–6817. Available from: <https://doi.org/10.2147/ott.s258439>.
- [16] Bai KH, He SY, Shu LL, Wang WD, Lin SY, Zhang QY, et al. Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index. *Cancer Medicine*. 2020 Apr;9(12):4290–4298. Available from: <https://doi.org/10.1002/cam4.3047>.
- [17] Liang JW, Fang ZY, Huang Y, Liuyang Zy, Zhang XL, Wang JL, et al. Application of Weighted Gene Co-Expression Network Analysis to Explore the Key Genes in Alzheimer's Disease. *Journal of Alzheimer's Disease*. 2018 Sep;65(4):1353–1364. Available from: <https://doi.org/10.3233/JAD-180400>.
- [18] Zeng D, He S, Ma C, Wen Y, Song W, Xu Q, et al. Network-based approach to identify molecular signatures in the brains of depressed suicides. *Psychiatry Research*. 2020 Dec;294:113513. Available from: <https://doi.org/10.1016/j.psychres.2020.113513>.
- [19] Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. 2016 Apr;34(5):525–527. Available from: <https://doi.org/10.1038/nbt.3519>.
- [20] Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*. 2014;15(2):R29. Available from: <https://doi.org/10.1186/gb-2014-15-2-r29>.
- [21] Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*. 2004 Jan;3(1):1–25. Available from: <https://doi.org/10.2202/1544-6115.1027>.
- [22] Kadarmideen HN, Watson-haigh NS. Building gene co-expression networks using transcriptomics data for systems biology investigations: Comparison of methods using microarray data. *Bioinformatics*. 2012 Sep;8(18):855–861. Available from: <https://doi.org/10.6026/97320630008855>.

- [23] Chen YA, Tripathi LP, Fujiwara T, Kameyama T, Itoh MN, Mizuguchi K. The TargetMine Data Warehouse: Enhancement and Updates. *Frontiers in Genetics*. 2019 Oct;10. Available from: <https://doi.org/10.3389/fgene.2019.00934>.
- [24] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2021. Available from: <https://www.R-project.org/>.
- [25] Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014 Jun;30(19):2811–2812. Available from: <https://doi.org/10.1093/bioinformatics/btu393>.
- [26] Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available from: <https://ggplot2.tidyverse.org>.
- [27] Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal*. 2006;Complex Systems:1695. Available from: <https://igraph.org>.
- [28] Pedersen TL. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks; 2021. R package version 2.0.5. Available from: <https://CRAN.R-project.org/package=ggraph>.
- [29] Bu J, Liu W, Pan Z, Ling K. Comparative Study of Hydrochemical Classification Based on Different Hierarchical Cluster Analysis Methods. *International Journal of Environmental Research and Public Health*. 2020 Dec;17(24):9515. Available from: <https://doi.org/10.3390/ijerph17249515>.
- [30] Langfelder P, Mischel PS, Horvath S. When Is Hub Gene Selection Better than Standard Meta-Analysis? *PLoS ONE*. 2013 Apr;8(4):e61505. Available from: <https://doi.org/10.1371/journal.pone.0061505>.
- [31] Lou Q, Chen L, Mei H, Xu K, Wei H, Feng F, et al. Root Transcriptomic Analysis Revealing the Importance of Energy Metabolism to the Development of Deep Roots in Rice (*Oryza sativa* L.). *Frontiers in Plant Science*. 2017 Jul;8. Available from: <https://doi.org/10.3389/fpls.2017.01314>.
- [32] Loo J, Spittle DA, Newnham M. COVID-19, immunothrombosis and venous thromboembolism: biological mechanisms. *Thorax*. 2021 Jan;76(4):412–420. Available from: <https://doi.org/10.1136/thoraxjnl-2020-216243>.
- [33] May AE, Seizer P, Gawaz M. Platelets: Inflammatory Firebugs of Vascular Walls. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2008 Mar;28(3). Available from: <https://doi.org/10.1161/atvbaha.107.158915>.
- [34] Aibibula M, Naseem KM, Sturmey RG. Glucose metabolism and metabolic flexibility in blood platelets. *Journal of Thrombosis and Haemostasis*. 2018 Sep;16(11):2300–2314. Available from: <https://doi.org/10.1111/jth.14274>.
- [35] Fidler TP, Campbell RA, Funari T, Dunne N, Angeles EB, Middleton EA, et al. Deletion of GLUT1 and GLUT3 Reveals Multiple Roles for Glucose Metabolism in Platelet and Megakaryocyte Function. *Cell Reports*. 2017 Jul;20(4):881–894. Available from: <https://doi.org/10.1016/j.celrep.2017.06.083>.
- [36] Fidler TP, Marti A, Gerth K, Middleton EA, Campbell RA, Rondina MT, et al. Glucose Metabolism Is Required for Platelet Hyperactivation in a Murine Model of Type 1 Diabetes. *Diabetes*. 2019 Feb;68(5):932–938. Available from: <https://doi.org/10.2337/db18-0981>.
- [37] Shenoy S. Coronavirus (Covid-19) sepsis: revisiting mitochondrial dysfunction in pathogenesis, aging, inflammation, and mortality. *Inflammation Research*. 2020 Aug;69(11):1077–1085. Available from: <https://doi.org/10.1007/s00011-020-01389-z>.
- [38] Ajaz S, McPhail MJ, Singh KK, Mujib S, Trovato FM, Napoli S, et al. Mitochondrial metabolic manipulation by SARS-CoV-2 in peripheral blood mononuclear cells of patients with COVID-19. *American Journal of Physiology-Cell Physiology*. 2021 Jan;320(1):C57–C65. Available from: <https://doi.org/10.1152/ajpcell.00426.2020>.
- [39] Denorme F, Manne BK, Portier I, Petrey AC, Middleton EA, Kile BT, et al. COVID-19 patients exhibit reduced procoagulant platelet responses. *Journal of Thrombosis and Haemostasis*. 2020 Oct;18(11):3067–3073. Available from: <https://doi.org/10.1111/jth.15107>.
- [40] Dasari CM, Bhukya R. Comparative analysis of protein synthesis rate in COVID-19 with other human coronaviruses. *Infection, Genetics and Evolution*. 2020 Nov;85:104432. Available from: <https://doi.org/10.1016/j.meegid.2020.104432>.
- [41] Schoenwaelder SM, Yuan Y, Josefsson EC, White MJ, Yao Y, Mason KD, et al. Two distinct pathways regulate platelet phosphatidylserine exposure and procoagulant function. *Blood*. 2009 Jul;114(3):663–666. Available from: <https://doi.org/10.1182/blood-2009-01-200345>.
- [42] Althaus K, Marini I, Zlamal J, Pelzl L, Singh A, Häberle H, et al. Antibody-induced procoagulant platelets in severe COVID-19 infection. *Blood*. 2021 Feb;137(8):1061–1071. Available from: <https://doi.org/10.1182/blood.202008762>.
- [43] Valentino ML, Barboni P, Ghelli A, Bucchi L, Rengo C, Achilli A, et al. The ND1 gene of complex I is a mutational hot spot for Leber's hereditary optic neuropathy. *Annals of Neurology*. 2004;56(5):631–641. Available from: <https://doi.org/10.1002/ana.20236>.
- [44] Burkhardt JM, Vaudel M, Gambaryan S, Radau S, Walter U, Martens L, et al. The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. *Blood*. 2012 Oct;120(15):e73–e82. Available from: <https://doi.org/10.1182/blood-2012-04-416594>.
- [45] Chen G, Cao P, Goeddel DV. TNF-Induced Recruitment and Activation of the IKK Complex Require Cdc37 and Hsp90. *Molecular Cell*. 2002 Feb;9(2):401–410. Available from: [https://doi.org/10.1016/s1097-2765\(02\)00450-1](https://doi.org/10.1016/s1097-2765(02)00450-1).
- [46] Pawlinski R. Inhibit the calpain to climb the mountain. *Blood*. 2014 Feb;123(8):1123–1124. Available from: <https://doi.org/10.1182/blood-2013-12-543397>.
- [47] Azam M, Andrabi SS, Sahr KE, Kamath L, Kuliopulos A, Chishti AH. Disruption of the Mouse μ -Calpain Gene Reveals an Essential Role in Platelet Function. *Molecular and Cellular Biology*. 2001 Mar;21(6):2213–2220. Available from: <https://doi.org/10.1128/mcb.21.6.2213-2220.2001>.
- [48] Nitsure M, Sarangi B, Shankar GH, Reddy VS, Walimbe A, Sharma V. Mechanisms of Hypoxia in COVID-19 Patients: A Pathophysiologic Reflection. *Indian Journal of Critical Care Medicine*. 2020;24(10):967–970. Available from: <https://doi.org/10.5005/jp-journals-10071-23547>.
- [49] Thauerer B, Voegelé P, Hermann-Kleiter N, Thuille N, de Araujo MEG, Offterdinger M, et al. LAMTOR2-Mediated Modulation of NGF/MAPK Activation Kinetics during Differentiation of PC12 Cells. *PLoS ONE*. 2014 Apr;9(4):e95863. Available from: <https://doi.org/10.1371/journal.pone.0095863>.
- [50] Sparber F, Scheffler JM, Amberg N, Tripp CH, Heib V, Hermann M, et al. The late endosomal adaptor molecule p14 (LAMTOR2) represents a novel regulator of Langerhans cell homeostasis. *Blood*. 2014 Jan;123(2):217–227. Available from: <https://doi.org/10.1182/blood-2013-08-518555>.
- [51] Flevaris P, Li Z, Zhang G, Zheng Y, Liu J, Du X. Two distinct roles of mitogen-activated protein kinases in platelets and a novel Rac1-MAPK-dependent integrin outside-in retractile signaling pathway. *Blood*. 2009 Jan;113(4):893–901. Available from: <https://doi.org/10.1182/blood-2008-05-155978>.
- [52] Aslan JE, McCarty OJT. Regulation of the mTOR-Rac1 axis in platelet function. *Small GTPases*. 2012 Jan;3(1):67–70. Available from: <https://doi.org/10.4161/sgtp.19137>.
- [53] Zamani R, Shahkarami S, Rezaei N. Primary immunodeficiency associated with hypopigmentation: A differential diagnosis approach. *Allergologia et Immunopathologia*. 2021 Mar;49(2):178–190. Available from: <https://doi.org/10.15586/aei.v49i2.61>.
- [54] Chow RD, Majety M, Chen S. The aging transcriptome and cellular landscape of the human lung in relation to SARS-CoV-2. *Nature Communications*. 2021 Jan;12(1). Available from: <https://doi.org/10.1038/s41467-020-20323-9>.
- [55] Eicher JD, Wakabayashi Y, Vitseva O, Esa N, Yang Y, Zhu J, et al. Characterization of the platelet transcriptome by RNA sequencing in patients with acute myocardial infarction. *Platelets*. 2015 Sep;27(3):230–239. Available from: <https://doi.org/10.3109/09537104.2015.1083543>.
- [56] Sol N, Leurs CE, Veld SGI, Strijbis EM, Vancura A, Schweiger MW, et al. Blood platelet RNA enables the detection of multiple sclerosis. *Multiple Sclerosis Journal - Experimental, Translational and Clinical*. 2020 Jul;6(3):205521732094678. Available from: <https://doi.org/10.1177/2055217320946784>.
- [57] Rondina MT, Voora D, Simon LM, Schwartz H, Harper JF, Lee O, et al. Longitudinal RNA-Seq Analysis of the Repeatability of Gene Expression and Splicing in Human Platelets Identifies a Platelet SELP Splice QTL. *Circulation Research*. 2020 Feb;126(4):501–516. Available from: <https://doi.org/10.1161/circresaha.119.315215>.

- [58] Everitt AR, , Clare S, Pertel T, John SP, Wash RS, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature*. 2012 Mar;484(7395):519–523. Available from: <https://doi.org/10.1038/nature10921>.
- [59] Zhu X, He Z, Yuan J, Wen W, Huang X, Hu Y, et al. IFITM3-containing exosome as a novel mediator for anti-viral response in dengue virus infection. *Cellular Microbiology*. 2014 Aug;17(1):105–118. Available from: <https://doi.org/10.1111/cmi.12339>.
- [60] Gorman MJ, Poddar S, Farzan M, Diamond MS. The Interferon-Stimulated Gene Ifitm3 Restricts West Nile Virus Infection and Pathogenesis. *Journal of Virology*. 2016 Sep;90(18):8212–8225. Available from: <https://doi.org/10.1128/jvi.00581-16>.
- [61] Hachim MY, Heialy SA, Hachim IY, Halwani R, Senok AC, Maghazachi AA, et al. Interferon-Induced Transmembrane Protein (IFITM3) Is Upregulated Explicitly in SARS-CoV-2 Infected Lung Epithelial Cells. *Frontiers in Immunology*. 2020 Jun;11. Available from: <https://doi.org/10.3389/fimmu.2020.01372>.

SUPPLEMENTARY MATERIAL

Supplementary Figure 1

Supplementary Figure 1: Network Soft threshold

Supplementary Figure 2

Supplementary Figure 2: Module Membership and gene significance in other modules.

Supplementary Figure 3

Supplementary Figure 3: Detailed enrichment analysis. A; Gene Ontology of biological function. B; Reactome pathways enrichment.