Gene Co-Expression Networks Offer New Perspectives on Sepsis Pathophysiology

P. Martínez-Paz, J. Gomez-Pilar, M. Martín-Fernández, F. C. Ceballos, E. Gómez-Sánchez, Roberto Hornero, *Member*, *IEEE*, Eduardo Tamayo

Abstract— Sepsis is one of the most common causes of death in intensive care units. Septic shock is a type of circulatory shock that shows signs and symptoms that are similar to non-septic shock. Despite the impact of shock in patients and the economic burden, knowledge on the pathophysiology of septic shock is scarce. In this context, weighted gene co-expression network analysis can help to elucidate the molecular mechanisms of this condition. The gene expression dataset used in this study was downloaded from the Gene Expression Omnibus, which contains 80 patients with septic shock, 33 patients with non-septic shock, and 15 healthy controls. Our novel analysis revealed five gene modules specific for patients with septic shock and three specific gene modules for patients with non-septic shock. Interestingly, genes related to septic shock were mainly involved in the immune system and endothelial cells, while genes related to non-septic shock were mostly associated with endothelial cells. Together, the results revealed the specificity of the genes related to immune system in the septic shock. The novel approach developed here showed its potential to identify critical pathways for the occurrence and progression of these conditions while offering new treatment strategies and effective therapies.

Index Terms— Biology and genetics, Gene co-expression network analysis, Sepsis

1 Introduction

THE last definition of sepsis states that is an organ dysfunction caused by a dysregulated host response to infection [1]. This condition is one of the main health care problems in the intensive care units (ICUs) [2] and represents a challenge for physicians due to its high mortality rate. Despite the advances in the care of patients, the incidence of sepsis has increased while, fortunately, the mortality rate has decreased [3]. In fact, a recent study has estimated around 31.5 million cases of sepsis worldwide, with 19.4 million cases being considered severe sepsis, and 5.3 million deaths annually [4]. Moreover, sepsis represents the first cause of mortality in non-coronary ICUs [5,6], with a mortality rate of 38% in the case of septic shock in Europe and North America [7]. In addition to the negative impact of sepsis in patients, the economic burden of sepsis has been increasing over the last several years and represents a challenge for health care systems, with an increase in cost due to longer hospital stays. Supporting this, the average hospital cost per stay was estimated at $37,424, $32,421, $13,292, and $24,384 for Europe, the United States, Asia, and South America, respectively [8]. For these reasons, the World Health Organization recognizes sepsis as a global health priority [9].

In spite of the significant health problem that sepsis represents, and the advances made in understanding its pathophysiology in the last several years, the knowledge about the dysregulation of the complex molecular signaling network in patients with sepsis and septic shock is scarce. Currently, one of the methods used to know the specific pathological state of sepsis is the analysis of gene expression patterns [10–13], allowing the identification of diagnostic and prognostic gene signatures, as well as novel therapeutic targets. However, this has not been enough to elucidate the molecular mechanisms of this condition. In this sense, weighted gene co-expression network analysis (WGCNA) applied to the gene expression values of patients with septic shock can help to uncover the underlying biological functions of genes and describe the huge and complex relationships in this condition. Thereby, the use of WGCNA could reveal relevant routes in the context of septic shock that, until now, have remained hidden. In recent years, an increasing number of works have successfully applied this method to discover the genes associated with various diseases, such as chronic kidney disease [14], cancer [15–17], asthma [18], and diabetes [19]. These

- *P. Martínez-Paz is with Department of Surgery, Faculty of Medicine, University of Valladolid. Avenida de Ramón y Cajal 7, 47005 Valladolid, with Group for Biomedical Research in Critical Care Medicine (BioCritic). Valladolid, Spain, and with Biomedical Research Networking Center in Infectious Diseases (CIBERINFEC), Carlos III Institute of Health, Spain*
- *J. Gomez-Pilar (corresponding author) and R. Hornero are with the Biomedical Engineering Group, University of Valladolid. Paseo de Belén 15, 47011 Valladolid, Spain, and also with the Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Carlos III Institute of Health, Spain; R. Hornero is also with the IMUVA Mathematical Institute, University of Valladolid. Paseo de Belén s/n, 47011 Valladolid, Spain*
- *M. Martínez-Fernández is with the Group for Biomedical Research in Critical Care Medicine (BioCritic). Valladolid, Spain, with the Biomedical Research Networking Center in Infectious Diseases (CIBERINFEC), Carlos III Institute of Health, Spain, and with Department of Medicine, Dermatology and Toxicology, Faculty of Medicine, University of Valladolid. Avenida de Ramón y Cajal 7, 47005 Valladolid, Spain*
- *F. C. Ceballos is with Unit of Viral Infection and Immunity, National Center for Microbiology (CNM), Health Institute Carlos III (ISCIII). Carretera de Pozuelo 28, 28222 Majadahonda, Spain*
- *E. Gómez-Sánchez and E. Tamayo are with Department of Surgery, Faculty of Medicine, University of Valladolid. Avenida de Ramón y Cajal 7, 47005 Valladolid, Spain, with the Group for Biomedical Research in Critical Care Medicine (BioCritic). Valladolid, Spain, with the Biomedical Research Networking Center in Infectious Diseases (CIBERINFEC), Carlos III Institute of Health, Spain, and with Anesthesiology and Resuscitation Service, University Clinical Hospital of Valladolid. Avenida de Ramón y Cajal 3, 47003 Valladolid, Spain*

precedents indicate that this method could be also used to identify key genes as novel candidate biomarkers or therapeutic targets in septic shock. Moreover, sepsis lacks a quick and accurate gold standard for its diagnosis, making it difficult to differentiate between septic and non-septic shock following surgery, conditions that show similar signs and symptoms [20]. Hence, this technique could allow us to identify specific and unique genetic patterns of septic shock, opening new doors for the personalized treatment in this pathology.

Previous reports have revealed the gene co-expression patterns in sepsis [21–23], but these studies were focused on medical sepsis. Thus, the goal of the present study is to increase the knowledge about the correlation network in patients with septic shock and non-septic shock by the identification of specific gene clusters with the aim to understand the pathophysiology of septic shock (i.e., the correlation between genes of unknown function with biological processes or distinguish transcriptional regulatory programs). Therefore, the potential findings of this study can help advance the understanding of the septic shock and non-septic shock transcriptomes and provide novel therapeutic targets.

2 METHODS

2.1 The Gene Expression Omnibus dataset

A microarray dataset with accession number GSE131761 was obtained from the Gene Expression Omnibus (GEO) public database. This dataset includes 129 samples comprising 15 healthy controls, 80 patients with septic shock, and 33 patients with non-septic shock. Pre-processing was performed as previously described by Martínez-Paz et al. [10]. Briefly, dataset files were imported into the Bioconductor R package ecosystem and were normal-exponential background corrected. Normalization was performed by the quantile method, and gene expression values were calculated using the *lmFit* function from the limma package.

2.2 Network analysis

After pre-processing, a correlation network was generated independently for each group by WGCNA. This method is a data mining method based on representing each gene as nodes and pairwise correlations between them as network links [24]. Pearson's rank correlations were used to recurrently assess the relationship between all pairs of nodes using the Matlab 'Statistics and Machine Learning Toolbox'. Due to the large size of the resulting networks (more than 500 million connections), the Cohen's threshold for large correlations [25] or very strong correlations [26] was applied. Thus, two weighted networks per group were generated, one representing high positive correlations (R>0.8) and the other high negative correlations (R<-0.8). Before thresholding, the generated correlation networks had the same number of nodes, only differing in the value of each correlation. However, after thresholding, each network can show different number of connections.

To assess the specific genes involved in septic shock, we were interested in analyzing the group-specific strong relationships, that is, those correlations between specific genes above the threshold (0.8) in the septic shock group but not in the non-septic shock group and vice versa. These networks are called differential networks. For this purpose, we developed a novel approach consisting of obtaining the characteristic and specific gene expression pattern of each group. Thus, after applying the threshold $|R|>0.8$, the networks were binarized, that is, a value of 1 was assigned to those connections higher than the threshold and a value of 0 the rest. Finally, the shared links between the groups were removed. In this way, new networks consisting of non-shared links were obtained. These networks only show specific strong connections (above 0.8) from each group, allowing analyzed the particular patters of each group.

With the aim of increase the robustness of the results while statistically comparing the properties of the networks, a previously validated bootstrap procedure [27,28] was applied for the first time in genetic data. Thereby, for each group, 100 random selections of 33 subjects with possible repetition (the number of subjects in the group of patients with non-septic shock, i.e., the more restrictive of the two groups) were used to generate the networks. Seven complementary graph parameters derived from Complex Network Theory were then calculated on each resulting network, including number of links, node degree, characteristic path length, diameter, average clustering coefficient, modularity, and eigen-vector centrality [29–32].

Finally, Gephi software (version 0.9.2) was used for network visualization [33]. Depending on the nature of the networks, two different force-based algorithms were used. The ForceAtlas2 algorithm [34] was applied to the weighted networks, which considers both the distance and the node degree of the connected nodes. On the other hand, the Fruchterman-Rheingold algorithm [35] was used to represent the binary networks, which uses custom forces of attraction and repulsion, depending only on the distance between the connected nodes. Despite the non-deterministic nature of these methods, they usually reach stable stages (as in the case of our networks) and have the advantage of turning structural proximities into visual proximities. Thereby, genetic communities or clusters emerge spatially separated, providing information about hidden genetic structures [36].

2.3 Pathway enrichment analysis

Pathway enrichment analysis identifies biological pathways that are enriched in a gene list more than it would be expected by chance. Analysis was developed using g:Profiler [37], database for annotation, visualization and integrated discovery (DAVID) [38], and protein annotation through evolutionary relationship (PANTHER) [39]. These techniques search a collection of gene sets representing Gene Ontology (GO) terms, pathways, networks, regulatory motifs, and disease phenotypes. Pathway enrichment methods use Fisher's exact test or binomial test, with Bonferroni correction for multiple testing, by considering all annotated protein-coding genes as background genes for comparison purposes. The general study design is summarized in Figure 1.

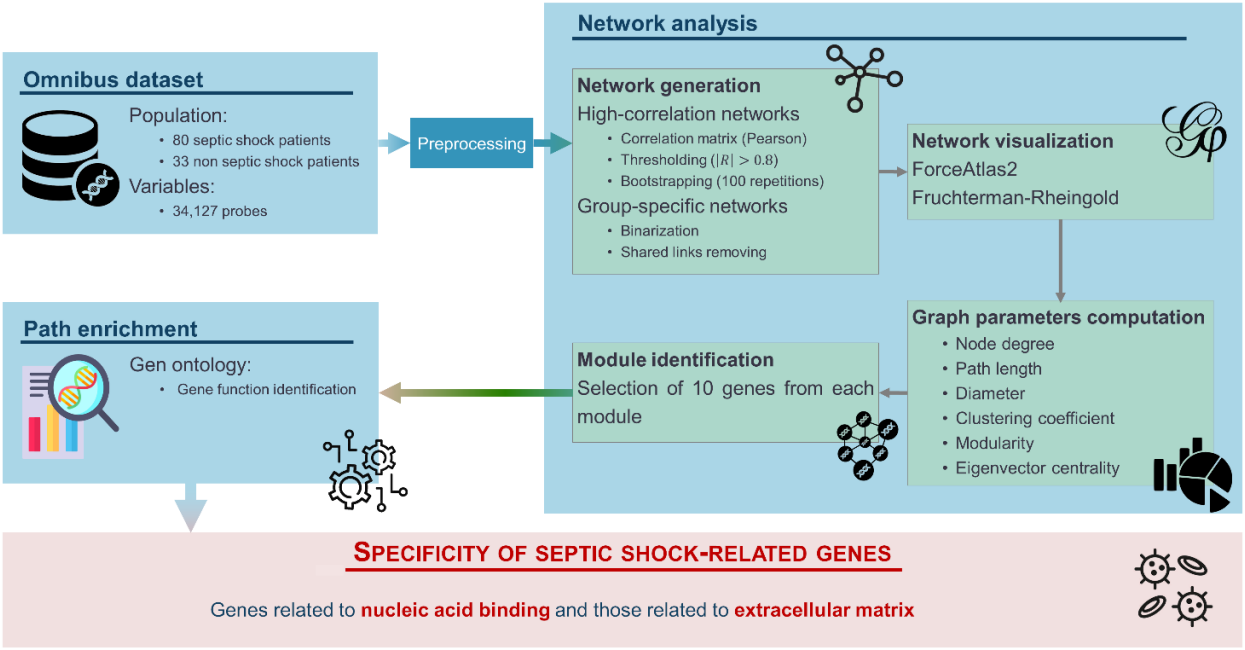

Fig. 1. Study workflow. Path enrichment after correlation network analysis identified genes particularly relevant in septic shock.

3 Results

A total of 113 patients from the Gene Expression Omnibus database (GSE131761) were included in the current study, of which 80 patients had septic shock and 33 patients had non-septic shock. The clinical characteristics of the postsurgical patients that were enrolled have been described previously [10].

3.1 Consensus network construction and module detection

In this work, we applied WGCNA using 34,127 probes from the microarrays of 133 patients to construct the gene modules from the matrix of gene expression values. The first step of the present study was to analyze the correlation structure in postsurgical patients with septic shock and non-septic shock, with the aim to evaluate the behavior of gene clusters and identify changes in gene-to-gene interactions that can be associated with these conditions. Table 1 shows the graph-theory-related parameters of the high positive and high negative weighted correlation networks considering all the possible pairs of nodes. The focus was on different complementary characteristics, including basic features, integration, segregation, and centrality, to compressively characterize the networks. Two basic features of the network were provided. First, the number of links in the network indicates the number of correlations higher than 0.8 for high positive networks or lower than -0.8 for high negative networks. The node degree also was calculated, which provides information on the connectedness of the considered gene by adding all the correlations that start from that node in a single index [29]. Thus, the average node degree summarizes the density of the network. The integration measures give an estimate of the degree of compactness of the network. Here, the characteristic path length and the diameter were reported. While the characteristic path length is the average shortest path length between all pairs of nodes [30], the diameter is the shortest distance between the two most distant nodes in the network [24]. On the other hand, network segregation is the capability of the network to be divided in different units with high intra-unit connectivity. The average clustering coefficient [31] and Blondel's modularity [32] were used for this work. The average clustering coefficient measures the presence of clusters inside the network by computing the ratio between the existing triangles and the total number of triangles that could exist. The modularity index provides information on how different modules inside the network are separated from each other in terms of correlations. Finally, the centrality of a node provides an estimate of the degree of relevance of that node within that network. In this context, if a node is very relevant (usually named 'hub'), it means that it is well connected and, therefore, many paths pass through it. The degree of centrality of the network gives an idea of its global topology. In particular, the eigenvector centrality [40] measures the average influence of all the nodes, that is, its connectedness to other important/highly connected genes.

The results obtained from this analysis show that patients with non-septic shock presented with larger high correlation networks as indexed by the number of links, both positive and negative (Table I). In addition, the overall connectivity of the network is diminished in septic shock, meaning a lower global relationship between gene expressions. This is particularly noticeable in the negative correlation network. Differences between networks are also evident regarding the network integration and segregation. Finally, higher degrees of centrality are shown by non-septic shock, supporting the lower number of hubs (high connected nodes) in them. In summary, the high positive correlation network presented a higher size and

TABLE 1
GRAPH-THEORY PARAMETERS ASSOCIATED WITH THE HIGH-CORRELATION NETWORKS (POSITIVE AND NEGATIVE) FOR EACH GROUP

Feature	Graph parameter	High positive correlation networks			High negative correlation networks		
		Septic shock	Non-septic shock	*p*-value	Septic shock	Non-septic shock	*p*-value
Basic	Number of links	502,031	616,045	<0.05	531	28,783	<0.05
	Average node degree	29.421	36.103	<0.05	2.855	8.489	<0.05
Integration	Characteristic path length	8.060	7.116	<0.05	3.131	5.651	<0.05
	Diameter	24	24	N.S.	8	20	<0.05
Segregation	Average clustering coefficient	0.570	0.422	N.S.	0.000	0.000	N.S.
	Modularity	0.257	0.349	<0.05	0.527	0.563	N.S.
Centrality	Eigenvector centrality	0.030	0.051	<0.05	0.007	0.074	<0.05

N.S.: Non-significant

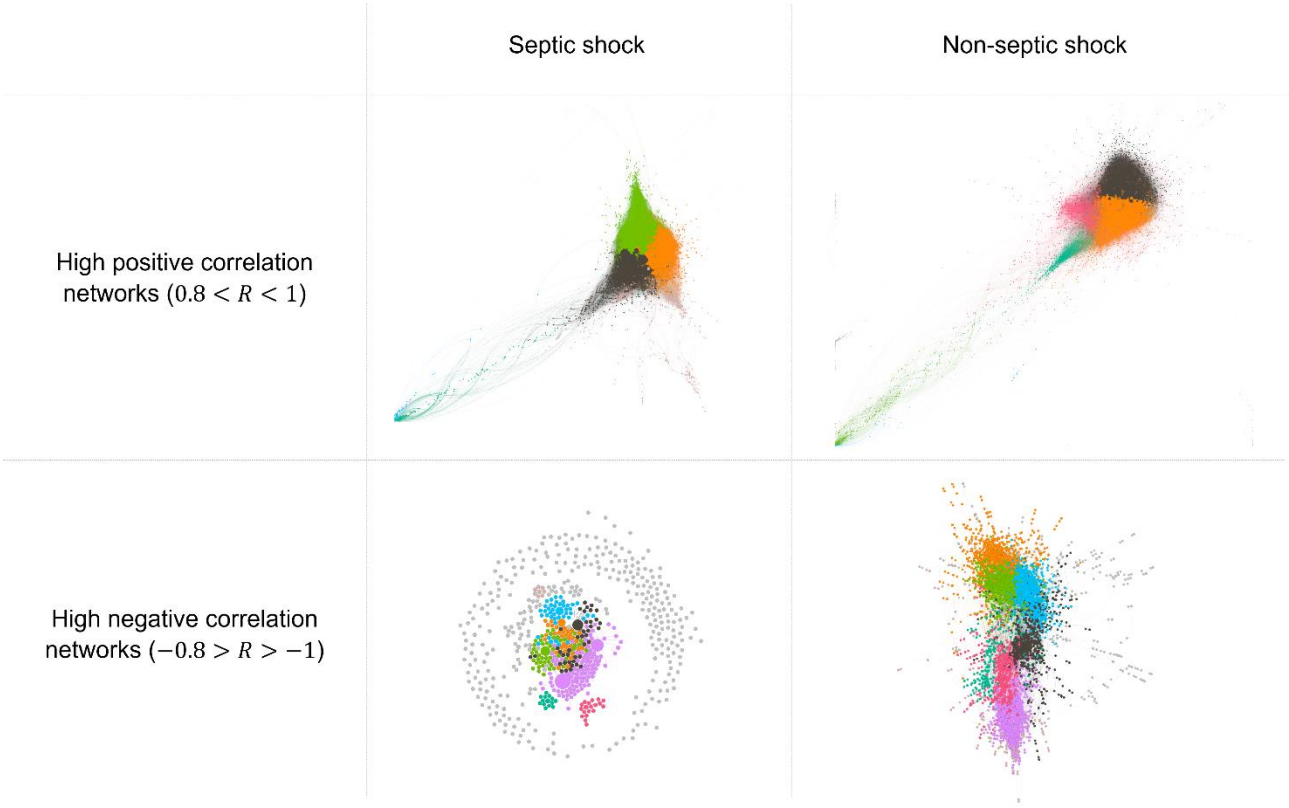

Fig. 2. High positive and high negative correlation networks for patients with septic shock and non-septic shock. The color of each node represents the membership of a specific module obtained by applying Blondel's modularity [32]. The arrangement of the nodes follows the ForceAtlas2 algorithm [34], which is based on attractive and repulsive forces between nodes, appropriate for weighted networks. In this way, the nodes (and modules) with the greatest relationship between them tend to appear spatially close.

connectedness in patients with non-septic shock. Similarly, when the high negative correlation network was analyzed these parameters where higher in patients with non-septic shock. These results are depicted in Figure 2, where the structure of these correlation networks and the presence of different clusters is shown, allowing the possibility to characterize it and to inspect the differences between these kinds of postsurgical patients.

3.2 Septic shock and non-septic shock network analyses without shared links

With the aim to analyze the specific relations between genes particularized for each group, new correlation networks, called differential networks, were performed by removing the non-shared links between groups and binarizing the resulting weighted networks. Thus, these networks consisted of a variety of links representing high correlations ($|R|>0.8$) that only appear in that group of patients but not in the other. The visual representation of the new correlation networks of patients with shock after removing the shared links confirm the existence of a reduced number of well-defined and separated clusters in each condition (Figure 3). Concerning these correlations, the septic shock network has a higher number of links, average of node degree, path length, and diameter when compared with the non-septic shock network, showing a high degree of specificity with the septic shock network (Table 2).

TABLE 2
GRAPH-THEORY PARAMETERS ASSOCIATED WITH THE DIFFERENTIAL CORE NETWORK

Feature	Graph parameter	Septic shock	Non-septic shock
Basic	Number of links	259,222	106,437
	Average node degree	15.192	6.238
Integration	Characteristic path length	5.772	3.180
	Diameter	31	24
Segregation	Average clustering coefficient	0.464	0.473
	Modularity	0.423	0.570
Centrality	Eigenvector centrality	0.078	0.396

N.S.: Non-significant

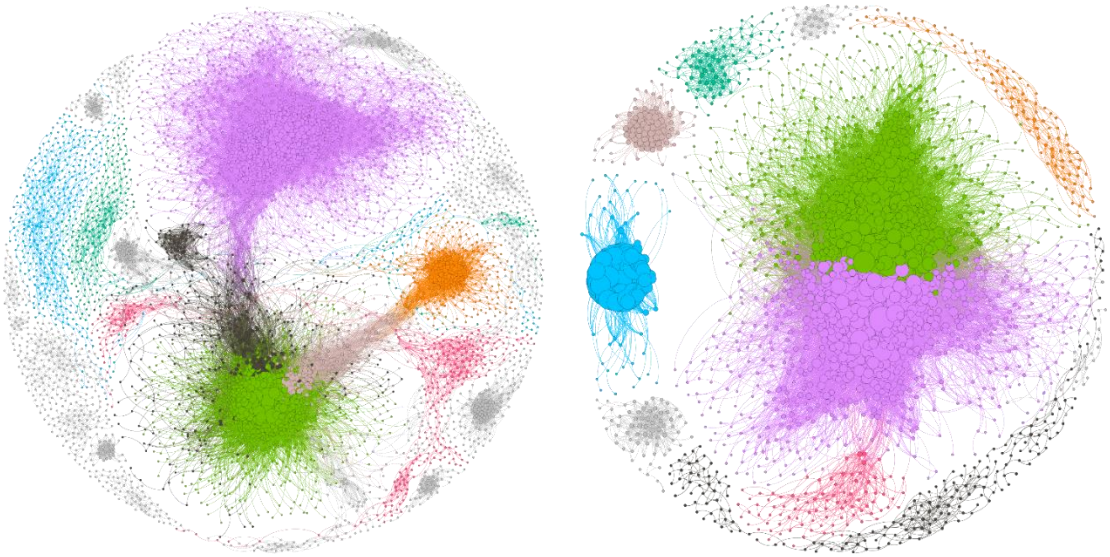

Fig. 3. Correlation networks (|R|>0.8) after removing shared links in patients with septic shock (left) and non-septic shock (right). The color of each node represents the membership of a specific module obtained by applying Blondel's modularity [32]. The arrangement of the nodes follows the Fruchterman-Reingold [35] algorithm, which is based on attractive and repulsive forces between nodes, appropriate for binary networks. The first goal of this study was to evaluate the behavior of gene clusters and identify changes in gene-to-gene interactions that can be associated to postsurgical patients with septic shock and non-septic shock through gene correlation networks. Non-septic patients presented a higher number of positive and negative links, connectivity, and centrality than septic patients. These results show marked genomic differences between both patient groups. However, with the aim to reinforce the hypothesis of a differentiated genetic signature in postsurgical patients with shock, a new network analysis was performed without shared links. In this case, the septic shock network presents a higher degree of specific for its correlations. Therefore, the comparison between the networks with shared and non-shared links shows that the correlation network of patients with non-septic shock was significantly changed after the exclusion on shared links.

3.3 Cluster analysis

Cluster analysis of septic shock and non-septic shock specific modules allows to study its characteristics and to identify the differences between both groups.

A GO enrichment analysis on the genes in modules was firstly performed. The main processes obtained in this analysis were related to nucleic acid binding and the extracellular matrix (see Table S1 in the Supplementary Material for details).

In addition to the previous analyses, genes with the highest node degree for each cluster from these non-shared networks were assessed (Figure 3). The shortlist of the top 10 genes in each cluster is shown in Table 3. From the patients with septic shock, the genes with higher node degree for the purple, green, blue, gray, and orange modules were IGLV5-48, COLEC10, MICU2, NDNF, and ST7-OT4, respectively; for non-septic shock, the genes with highest node degree for the purple and green modules were FLJ36000 and AGGF1, respectively.

4 DISCUSSION

Most previous studies have focused on the transcriptional profiling of sepsis and septic shock using microarrays to identify biomarker candidate genes [10–13]. However, while there are previous reports that have analyzed gene co-expression patterns in sepsis [21–23], there are no works about the scenario involving septic shock vs. non-septic shock to assess their pattern specificity. Moreover, compared with the classical analysis of transcriptional profiles, the study of gene networks-based methods allows one to gain insight into the pathophysiology of both septic shock and non-septic shock, as well as global biology activity, considering that both conditions show similar signs and symptoms [20].

The differences between specific modules with regard to patients with septic shock and non-septic shock open the possibility to study the particular characteristics of those clusters and to identify the arising differences between

TABLE 3
GENES WITH THE HIGHEST NODE DEGREE IN THE DIFFERENTIAL NETWORK

Septic shock					Non-septic shock		
Purple	**Green**	**Blue**	**Gray**	**Orange**	**Purple**	**Green**	**Blue**
IGLV5-48	*COLEC10*	*MICU2*	*NDNF*	*ST7-OT4*	*FLJ36000*	*AGGF1*	A_33_P3875570
HSPG2	*KATNBL1P6*	*CCAR1*	*APOL5*	*ETNK2*	*TNXB*	*PRRT1*	*C7orf65*
SAMD11	*WBP2NL*	*USP1*	*IGF2BP1*	*SCN10A*	*ZBTB3*	A_33_P3289456	A_33_P3278211
A_33_P3209321	*FAM154B*	*RDH14*	A_33_P3380783	A_33_P3312034	*FBXL17*	A_33_P3376026	*SLC18A1*
GPR25	*TMEM207*	*GOPC*	A_33_P3375299	*CCT7P2*	A_33_P3222664	A_33_P3300877	*SP5*
CDKN2A	A_33_P3555368	*TDG*	*IL26*	*GPX8*	*ATP6V1G2*	*RIMBP2*	A_33_P3420347
SLC22A11	*ADH1B*	*ATF1*	*TMCO5B*	*LGALS14*	A_24_P170726	*TEAD1*	A_33_P3383292
CELA2B	A_33_P3223059	*ZNF721*	A_33_P3266609	*RGS13*	A_33_P3306802	*AAK1*	*CCDC40*
MUC3A	*TTPA*	*FNTA*	*ACTRT2*	*NUTM1*	*BCL2L15*	A_33_P3713035	*AK127999 / KCNIP4*
TNFRSF14	*BSN*	A_32_P45493	A_33_P3392213	*PM20D1*	*CEP104*	*CHD1L*	*DEFB136*

them. Related to both conditions, the GO results show that the most prominent processes in the detected modules are related to nucleic acid binding and the extracellular matrix. However, previous reports show that the pathways involved in septic shock and non-septic shock were mainly related to the immune system, inflammatory processes or endothelial barriers [41–45]. Despite inconsistency in the module's annotation, these differences could be due to the way that genes were obtained for the analysis. While the previous works explored the GO using the differentially expressed genes, in the present study, the analyzed genes were obtained from correlation networks. Particularly, genes were selected as those with a higher node degree in each module (i.e., genes highly correlated with genes of the same module). These results could indicate the existence of crossover effects that could have been hidden by the classical analysis and revealed for the first time. In this sense, it has been reported the modulation of the extracellular matrix into the immune cell function [46] and how the composition of the extracellular matrix undergoes changes during infections [47]. Regarding the nucleic acid binding pathways obtained, previous studies have shown that nucleic acid binding proteins are associated with poor prognosis in septic patients [48] and are required for interferon production in response to viral infection [49]. Moreover, bacterial and viral nucleic acids can act as inductors of inflammation [50]. Thus, the present routes reported in this work have received less attention than other pathways in the infection processes and can offer a new research line in the pathophysiology of sepsis.

In this study, WGCNA was used with a bootstrapping procedure to analyze, in a robust way, those genes with higher node degree for each cluster from differential networks. Genes IGLV5-48, COLEC10, MICU2, NDNF, and ST7-OT4 presented the most node degree for septic shock. IGLV5-48 encodes for immunoglobulin lambda variable 5-48, involved in the immune response; however, its function and molecular mechanism are not clear [51]. In addition, the gene expression of members of the IGLV family were upregulated in patients with cardiogenic shock and septic shock [45]. COLEC10 encodes for a protein C-lectin family member, collectin subfamily member, with one of its functions being binding to antigens on microorganisms facilitating their recognition and removal. It has been reported that vascular endothelial cells have receptors for collectins [52]. As a result, these cells play a major role in the systemic response to bacterial infections [53]. In addition, the protein encoded by this gene activates the complement system [54]. MICU2 encodes for a transporter protein called mitochondrial calcium uptake 2. It has been reported its upregulated expression in cells in vitro after infection with Salmonella enterica Serovar Typhimurium [55].

Further, mitochondrial calcium uptake 2 plays an important role in the regulation of the Pseudomonas aeruginosa-dependent inflammatory response [56]. This protein has been associated with the induction of autophagy and apoptotic cell death in endothelial cells in response to oxygen-glucose deprivation [57]. NDNF encodes for neuron derived neurotrophic factor, which is secreted in cultured endothelial cells stimulated by hypoxia, promotes endothelial cell survival and vessel formation, and plays an important role in the process of revascularization [58]. ST7-OT4 encodes for a long non-coding RNA whose expression is upregulated in cardiac CCR2- macrophages [59]. Altogether, the node genes fit with the pathophysiology of sepsis, where this condition is defined as organ dysfunction caused by a host response to infection [1]. In this sense, the endothelial cells play a central role in the systemic response to bacterial infection, leading to multiorgan failure syndrome [53,60]. Moreover, these node genes are involved in the immune system, which is consistent with the key role of this system in sepsis and with previous

studies that suggest that this condition is accompanied by overall immune dysregulation [10,61–63]. Regarding patients with non-septic shock, the most node degree genes for the purple and green modules were FLJ36000 and AGGF1. In the case of the blue module, the most degree gene corresponded with an uncharacterized DNA sequence. FLJ36000 is a lncRNA, while AGGF1 encodes an angiogenic factor that acts as an anti-inflammatory factor by suppressing endothelial activation responses to TNF-α [64]. On the other hand, a brief description of the role of the top 10 genes of each cluster is described in Table S1 (Supplementary Material). Overall, these results can help to identify new gene signatures that help to understand the pathophysiology of septic shock and non-septic shock. However, the 10 genes with the highest node degree for each cluster from these non-shared networks were analyzed, aiming at finding their relation to septic shock and non-septic shock. As shown in Table S1, these genes maintained the relationship with each condition. While septic shock genes are mostly involved in inflammatory processes, the immune system, and endothelial cells, the non-septic shock genes are mainly related to endothelial cells.

This work presents limitations that we must acknowledge. First, no distinction of different subgroups within non-septic shock was made. This would allow an even more specific study of the patterns within this heterogeneous group. However, due to the sample size of the database, this would also reduce the statistical power of studying each group separately. Second, it is a single-center study; therefore, a multi-center study would provide possible inter-hospital variation, providing more robustness to the results. Third, a compatible database with characteristics similar to the one reported here was not found. This would have made it possible to validate the results, which would help to reinforce their robustness. However, for that reason, a bootstrap-based approach with 100 iterations was used. Since each iteration produces different networks, the results found here are more robust and generalizable than with a classical approach. The last limitation is about the nature of the samples, where the peripheral blood provides the gene expression patterns of white blood cells and offers mainly an insight into immune pathways. Thus, future works should keep in mind the kind of shock to analyze the relationship with its specific gene expression patterns and analyze its mRNA level in the endothelial cells, which appear to be the target tissue for these conditions.

5 Conclusions

The present study identified novel genetic modules from correlation networks associated with septic shock and non-septic shock in post-surgical patients using gene co-expression network analysis. This was achieved by using a novel procedure that combines correlation network, differential networks, and a bootstrap procedure to increase the robustness of the results. Of each module, the most representative genes in septic shock were mainly related to the immune system and endothelial cells, while genes encoding aspects related to endothelial cells were the most representative for non-septic shock. This novel way of selecting the most relevant genes could provide new pathways that might have remained hidden until now. Therefore, these results offer a new insight into patients with shock with the aim of promoting the identification of critical pathways and providing new treatment strategies in future clinical studies.

Acknowledgment

This work was supported by the Carlos III Institute of Health from Ministerio de Ciencia, Innovación y Universidades (Spain) [grant number PI18/01238], by Ministerio de Ciencia, Innovación y Universidades (Spain) [grant numbers PID2020-115468RB-I00 and PDC2021-120775-I00], by Fundación Ramón Areces (Spain), and by Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN) and Biomedical Research Networking Center in Infectious Diseases (CIBERINFEC) through Carlos III Institute of Health (Spain). Javier Gomez-Pilar is the corresponding author. Roberto Hornero and Eduardo Tamayo contributed equally to the work.

References

[1] M. Singer *et al.*, 'The Third International Consensus Definitions for sepsis and septic shock (Sepsis-3)', *JAMA*, vol. 315, no. 8, p. 801, Feb. 2016, doi: 10.1001/jama.2016.0287.

[2] J.-L. Vincent *et al.*, 'Sepsis in European intensive care units: Results of the SOAP study', *Crit Care Med*, vol. 34, no. 2, p. 10, 2006.

[3] C. Rhee and M. Klompas, 'Sepsis trends: increasing incidence and decreasing mortality, or changing denominator?', *J. Thorac. Dis.*, vol. 12, no. S1, pp. S89–S100, Feb. 2020, doi: 10.21037/jtd.2019.12.51.

[4] C. Fleischmann *et al.*, 'Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations', vol. 193, no. 3, p. 14, 2016.

[5] R. Ferrer *et al.*, 'Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain', *JAMA*, vol. 299, no. 19, pp. 2294–2303, May 2008, doi: 10.1001/jama.299.19.2294.

[6] M. M. Levy *et al.*, 'Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study', *Crit. Care Med.*, vol. 43, no. 1, pp. 3–12, Jan. 2015, doi: 10.1097/CCM.0000000000000723.

[7] J.-L. Vincent, G. Jones, S. David, E. Olariu, and K. K. Cadwell, 'Frequency and mortality of septic shock in Europe and North America: a systematic review and meta-analysis', *Crit. Care*, vol. 23, no. 1, p. 196, Dec. 2019, doi: 10.1186/s13054-019-2478-6.

[8] H. Arefian *et al.*, 'Hospital-related cost of sepsis: A systematic review', *J. Infect.*, vol. 74, no. 2, pp. 107–117, Feb. 2017, doi: 10.1016/j.jinf.2016.11.006.

[9] K. Reinhart, R. Daniels, N. Kissoon, F. R. Machado, R. D. Schachter, and S. Finfer, 'Recognizing sepsis as a global health priority — A WHO resolution', *N. Engl. J. Med.*, vol. 377, no. 5, pp. 414–417, Aug. 2017, doi: 10.1056/NEJMp1707170.

[10] P. Martínez-Paz *et al.*, 'Distinguishing septic shock from non-septic shock in postsurgical patients using gene expression', *J. Infect.*, vol. 83, no. 2, pp. 147–155, Aug. 2021, doi: 10.1016/j.jinf.2021.05.039.

[11] P. Severino *et al.*, 'Patterns of gene expression in peripheral blood mononuclear cells and outcomes from patients with sepsis

secondary to community acquired pneumonia', *PLoS ONE*, vol. 9, no. 3, p. e91886, Mar. 2014, doi: 10.1371/journal.pone.0091886.

[12] L. McHugh *et al.*, 'A molecular host response assay to discriminate between sepsis and infection-negative systemic inflammation in critically ill patients: discovery and validation in independent cohorts', *PLOS Med.*, vol. 12, no. 12, p. e1001916, Dec. 2015, doi: 10.1371/journal.pmed.1001916.

[13] K. L. Burnham *et al.*, 'Shared and distinct aspects of the sepsis transcriptomic response to fecal peritonitis and pneumonia', *Am. J. Respir. Crit. Care Med.*, vol. 196, no. 3, pp. 328–339, Aug. 2017, doi: 10.1164/rccm.201608-1685OC.

[14] Z. Zuo *et al.*, 'Weighted Gene Correlation Network Analysis (WGCNA) detected loss of MAGI2 promotes Chronic Kidney Disease (CKD) by podocyte damage', *Cell. Physiol. Biochem.*, vol. 51, no. 1, pp. 244–261, 2018, doi: 10.1159/000495205.

[15] F. Magani *et al.*, 'Identification of an oncogenic network with prognostic and therapeutic value in prostate cancer', *Mol. Syst. Biol.*, vol. 14, no. 8, Aug. 2018, doi: 10.15252/msb.20188202.

[16] J. Tang *et al.*, 'Weighted gene correlation network analysis identifies RSAD2, HERC5, and CCL8 as prognostic candidates for breast cancer', *J. Cell. Physiol.*, vol. 235, no. 1, pp. 394–407, Jan. 2020, doi: 10.1002/jcp.28980.

[17] Y. Lai, G. OuYang, L. Sheng, Y. Zhang, B. Lai, and M. Zhou, 'Novel prognostic genes and subclasses of acute myeloid leukemia revealed by survival analysis of gene expression data', *BMC Med. Genomics*, vol. 14, no. 1, p. 39, Dec. 2021, doi: 10.1186/s12920-021-00888-0.

[18] L.-L. He, F. Xu, X.-Q. Zhan, Z.-H. Chen, and H.-H. Shen, 'Identification of critical genes associated with the development of asthma by co-expression modules construction', *Mol. Immunol.*, vol. 123, pp. 18–25, Jul. 2020, doi: 10.1016/j.molimm.2020.01.015.

[19] L. Li, Z. Pan, and X. Yang, 'Key genes and co-expression network analysis in the livers of type 2 diabetes patients', *J. Diabetes Investig.*, vol. 10, no. 4, pp. 951–962, Jul. 2019, doi: 10.1111/jdi.12998.

[20] J.-L. Vincent and D. De Backer, 'Circulatory shock', *N. Engl. J. Med.*, vol. 369, no. 18, pp. 1726–1734, Oct. 2013, doi: 10.1056/NEJMra1208943.

[21] Z. Zhang, L. Chen, P. Xu, L. Xing, Y. Hong, and P. Chen, 'Gene correlation network analysis to identify regulatory factors in sepsis', *J. Transl. Med.*, vol. 18, no. 1, p. 381, Dec. 2020, doi: 10.1186/s12967-020-02561-z.

[22] Y. Li, Y. Li, Z. Bai, J. Pan, J. Wang, and F. Fang, 'Identification of potential transcriptomic markers in developing pediatric sepsis: a weighted gene co-expression network analysis and a case–control validation study', *J. Transl. Med.*, vol. 15, no. 1, p. 254, Dec. 2017, doi: 10.1186/s12967-017-1364-8.

[23] R. Godini, H. Fallahi, and E. Ebrahimie, 'Network analysis of inflammatory responses to sepsis by neutrophils and peripheral blood mononuclear cells', *PLOS ONE*, vol. 13, no. 8, p. e0201674, Aug. 2018, doi: 10.1371/journal.pone.0201674.

[24] P. J. Gutiérrez-Díez, J. Gomez-Pilar, R. Hornero, J. Martínez-Rodríguez, M. A. López-Marcos, and J. Russo, 'The role of gene to gene interaction in the breast's genomic signature of pregnancy', *Sci. Rep.*, vol. 11, no. 1, p. 2643, Dec. 2021, doi: 10.1038/s41598-021-81704-8.

[25] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2. ed., Reprint. New York, NY: Lawrence Erlbaum Associates, 1988.

[26] Y. H. Chan, 'Biostatistics 104: correlational analysis', *Singapore Med. J.*, vol. 44, no. 12, pp. 614–619, Dec. 2003.

[27] N. Jimeno *et al.*, 'Main Symptomatic Treatment Targets in Suspected and Early Psychosis: New Insights From Network Analysis', *Schizophr. Bull.*, vol. 46, no. 4, pp. 884–895, Jul. 2020, doi: 10.1093/schbul/sbz140.

[28] G. C. Gutiérrez-Tobal, L. Kheirandish-Gozal, D. Gozal, and R. Hornero, 'Editorial: Unraveling Sleep and Its Disorders Using Novel Analytical Approaches', *Front. Neurosci.*, vol. 16, p. 924359, May 2022, doi: 10.3389/fnins.2022.924359.

[29] M. Rubinov and O. Sporns, 'Complex network measures of brain connectivity: Uses and interpretations', *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, Sep. 2010, doi: 10.1016/j.neuroimage.2009.10.003.

[30] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, 'Random graphs with arbitrary degree distributions and their applications', *Phys. Rev. E*, vol. 64, no. 2, p. 026118, Jul. 2001, doi: 10.1103/PhysRevE.64.026118.

[31] M. Latapy, 'Main-memory triangle computations for very large (sparse (power-law)) graphs', *Theor. Comput. Sci.*, vol. 407, no. 1–3, pp. 458–473, Nov. 2008, doi: 10.1016/j.tcs.2008.07.017.

[32] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, 'Fast unfolding of communities in large networks', *J. Stat. Mech. Theory Exp.*, vol. 2008, no. 10, p. P10008, Oct. 2008, doi: 10.1088/1742-5468/2008/10/P10008.

[33] M. Bastian, S. Heymann, and M. Jacomy, 'Gephi: an open source software for exploring and manipulating networks', 2009, doi: 10.13140/2.1.1341.1520.

[34] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, 'ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software', *PLoS ONE*, vol. 9, no. 6, p. e98679, Jun. 2014, doi: 10.1371/journal.pone.0098679.

[35] T. M. J. Fruchterman and E. M. Reingold, 'Graph drawing by force-directed placement', *Softw. Pract. Exp.*, vol. 21, no. 11, pp. 1129–1164, Nov. 1991, doi: 10.1002/spe.4380211102.

[36] A. Noack, 'Modularity clustering is force-directed layout', *Phys. Rev. E*, vol. 79, no. 2, p. 026102, Feb. 2009, doi: 10.1103/PhysRevE.79.026102.

[37] J. Reimand, M. Kull, H. Peterson, J. Hansen, and J. Vilo, 'g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments', *Nucleic Acids Res.*, vol. 35, no. Web Server issue, pp. W193-200, Jul. 2007, doi: 10.1093/nar/gkm226.

[38] G. Dennis *et al.*, 'DAVID: Database for Annotation, Visualization, and Integrated Discovery', *Genome Biol.*, vol. 4, no. 5, p. P3, 2003.

[39] H. Mi, A. Muruganujan, J. T. Casagrande, and P. D. Thomas, 'Large-scale gene function analysis with the PANTHER classification system', *Nat. Protoc.*, vol. 8, no. 8, pp. 1551–1566, Aug. 2013, doi: 10.1038/nprot.2013.092.

[40] M. J. Zaki and W. Meira, Jr, *Data mining and analysis: fundamental concepts and algorithms*, 1st ed. Cambridge University Press, 2014. doi: 10.1017/CBO9780511810114.

[41] Y. Tang *et al.*, 'Bioinformatic analysis identifies potential biomarkers and therapeutic targets of septic-shock-associated acute kidney injury', *Hereditas*, vol. 158, no. 1, p. 13, Apr. 2021, doi: 10.1186/s41065-021-00176-y.

[42] J. Yang *et al.*, 'Identification of key genes and pathways using bioinformatics analysis in septic shock children', *Infect. Drug Resist.*, vol. 11, pp. 1163–1174, 2018, doi: 10.2147/IDR.S157269.

[43] X. Zeng *et al.*, 'Screening of key genes of sepsis and septic shock using bioinformatics analysis', *J. Inflamm. Res.*, vol. 14, pp. 829–841, 2021, doi: 10.2147/JIR.S301663.

[44] K. S. Kim, D. W. Jekarl, J. Yoo, S. Lee, M. Kim, and Y. Kim, 'Immune gene expression networks in sepsis: A network biology approach', *PLOS ONE*, vol. 16, no. 3, p. e0247669, Mar. 2021, doi: 10.1371/journal.pone.0247669.

[45] D. Braga *et al.*, 'A longitudinal study highlights shared aspects of the transcriptomic response to cardiogenic and septic shock', *Crit. Care Lond. Engl.*, vol. 23, no. 1, p. 414, Dec. 2019, doi: 10.1186/s13054-019-2670-8.

[46] G. Maiti *et al.*, 'Matrix lumican endocytosed by immune cells controls receptor ligand trafficking to promote TLR4 and restrict

TLR9 in sepsis', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 27, p. e2100999118, Jul. 2021, doi: 10.1073/pnas.2100999118.

[47] H. Tomlin and A. M. Piccinini, 'A complex interplay between the extracellular matrix and the innate immune response to microbial pathogens', *Immunology*, vol. 155, no. 2, pp. 186–201, Oct. 2018, doi: 10.1111/imm.12972.

[48] Y. Zhou, H. Dong, Y. Zhong, J. Huang, J. Lv, and J. Li, 'The Cold-Inducible RNA-Binding Protein (CIRP) level in peripheral blood predicts sepsis outcome', *PloS One*, vol. 10, no. 9, p. e0137721, 2015, doi: 10.1371/journal.pone.0137721.

[49] Y. Chen, X. Lei, Z. Jiang, and K. A. Fitzgerald, 'Cellular nucleic acid-binding protein is essential for type I interferon-mediated immunity to RNA virus infection', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 26, p. e2100383118, Jun. 2021, doi: 10.1073/pnas.2100383118.

[50] J. Lee, J. W. Sohn, Y. Zhang, K. W. Leong, D. Pisetsky, and B. A. Sullenger, 'Nucleic acid-binding polymers as anti-inflammatory agents', *Proc. Natl. Acad. Sci.*, vol. 108, no. 34, pp. 14055–14060, Aug. 2011, doi: 10.1073/pnas.1105777108.

[51] X. Guan, Z.-Y. Xu, R. Chen, J.-J. Qin, and X.-D. Cheng, 'Identification of an immune gene-associated prognostic signature and its association with a poor prognosis in gastric cancer patients', *Front. Oncol.*, vol. 10, p. 629909, 2020, doi: 10.3389/fonc.2020.629909.

[52] U. Holmskov, R. Malhotra, R. B. Sim, and J. C. Jensenius, 'Collectins: collagenous C-type lectins of the innate immune defense system', *Immunol. Today*, vol. 15, no. 2, pp. 67–74, Feb. 1994, doi: 10.1016/0167-5699(94)90136-8.

[53] J. Joffre, J. Hellman, C. Ince, and H. Ait-Oufella, 'Endothelial responses in sepsis', *Am. J. Respir. Crit. Care Med.*, vol. 202, no. 3, pp. 361–370, Aug. 2020, doi: 10.1164/rccm.201910-1911TR.

[54] E. Axelgaard *et al.*, 'Investigations on Collectin Liver 1', *J. Biol. Chem.*, vol. 288, no. 32, pp. 23407–23420, Aug. 2013, doi: 10.1074/jbc.M113.492603.

[55] K.-C. Wang, C.-H. Huang, C.-J. Huang, and S.-B. Fang, 'Impacts of Salmonella enterica Serovar Typhimurium and its speG gene on the transcriptomes of in vitro M cells and Caco-2 cells', *PloS One*, vol. 11, no. 4, p. e0153444, 2016, doi: 10.1371/journal.pone.0153444.

[56] A. Rimessi, V. Bezzerri, S. Patergnani, S. Marchi, G. Cabrini, and P. Pinton, 'Mitochondrial Ca2+-dependent NLRP3 activation exacerbates the Pseudomonas aeruginosa-driven inflammatory response in cystic fibrosis', *Nat. Commun.*, vol. 6, p. 6201, Feb. 2015, doi: 10.1038/ncomms7201.

[57] V. Natarajan *et al.*, 'Oxygen glucose deprivation induced prosurvival autophagy is insufficient to rescue endothelial function', *Front. Physiol.*, vol. 11, p. 533683, Sep. 2020, doi: 10.3389/fphys.2020.533683.

[58] K. Ohashi *et al.*, 'Neuron-derived neurotrophic factor functions as a novel modulator that enhances endothelial cell function and revascularization processes', *J. Biol. Chem.*, vol. 289, no. 20, pp. 14132–14144, May 2014, doi: 10.1074/jbc.M114.555789.

[59] G. Bajpai *et al.*, 'The human heart contains distinct macrophage subsets with divergent origins and functions', *Nat. Med.*, vol. 24, no. 8, pp. 1234–1245, Aug. 2018, doi: 10.1038/s41591-018-0059-x.

[60] C. Ince *et al.*, 'The endothelium in sepsis', *Shock Augusta Ga*, vol. 45, no. 3, pp. 259–270, Mar. 2016, doi: 10.1097/SHK.0000000000000473.

[61] R. Almansa *et al.*, 'Transcriptomic correlates of organ failure extent in sepsis', *J. Infect.*, vol. 70, no. 5, pp. 445–456, May 2015, doi: 10.1016/j.jinf.2014.12.010.

[62] T. E. Sweeney and H. R. Wong, 'Risk stratification and prognosis in sepsis', *Clin. Chest Med.*, vol. 37, no. 2, pp. 209–218, Jun. 2016, doi: 10.1016/j.ccm.2016.01.003.

[63] G. P. Parnell *et al.*, 'Identifying key regulatory genes in the whole blood of septic patients to monitor underlying immune dysfunctions', *Shock*, vol. 40, no. 3, pp. 166–174, Sep. 2013, doi: 10.1097/SHK.0b013e31829ee604.

[64] F.-Y. Hu *et al.*, 'AGGF1 is a novel anti-inflammatory factor associated with TNF-α-induced endothelial activation', *Cell. Signal.*, vol. 25, no. 8, pp. 1645–1653, Aug. 2013, doi: 10.1016/j.cellsig.2013.04.007.

Pedro Martínez-Paz Pedro Martínez-Paz holds a Biology degree (2008) and a Postgraduate in Genetics (2009) from Complutense University of Madrid. He received his PhD in Science with the National Distance Education University in 2014, where he was awarded with the Outstanding Thesis Award, and he completed a Master in Bioinformatics with the Valencia International University in 2022. Pedro is currently working as a Postdoctoral Research Assistant at the Queen Mary University of London (United Kingdom). His research career has been focused on the search for molecular biomarkers, mainly at genetic level, but applied in environmental toxicology during his predoctoral and the beginning of his postdoctoral stage and in the field of septic patients during the advanced postdoctoral stage.

Javier Gomez-Pilar received the M.S. degree in Telecommunication Engineering (2012) from the University of Valladolid (Spain), where he also obtained the Master of Advanced Studies on Biomedical Engineering (2013) and the PhD (2018). He is currently a researcher on the Biomedical Engineering Group, associated with the Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine. His research is primarily focused on biomedical signal processing of electroencephalograms using time-frequency analysis and complex network theory to help in the diagnosis of several pathologies. Currently, he is involved in different studies applying correlation networks to integrate large sets of heterogeneous data, such us genetic data.

Marta Martín-Fernández received her Bachelor's Degree in Biology at the University of Salamanca (2015), the Master in Biomedical Research at the University of Valladolid (2016) and the PhD at the University of Valladolid (2021) in BioSepsis group with the Mention towards Excellence by the MEC. Her PhD was focused on the investigation and identification of diagnostic and prognostic biomarkers in severe infection and sepsis. Nowadays, she is a Postdoctoral Researcher at the University of Valladolid where she is involved in infection and molecular biology in collaboration with international researchers from Italy, United Kingdom, and Swiss. Her main research interests include molecular biology, systems biology and infection.

Francisco C. Ceballos Francisco has a PhD in Genetics at the University of Santiago de Compostela (2012) and a master's degree in statistics (2014). Francisco motivations as a geneticist are to understand natural demographic history and to decipher the genetic basis of complex traits. As a postdoc Francisco joined several teams in South Africa, Edinburgh, Ankara and Spain. Currently he is a postdoc at the Spanish Health Institute Carlos III in Spain under the IMPaCT project.

Esther Gómez-Sánchez is an associate professor at University of Valladolid, Spain, as well as an Anesthesiologist with more than 10 years of experience at Hospital Clínico of Valladolid, a tertiary Spanish hospital. She is PhD in Medicine since 2011. Furthermore, she has authored more than 40 publications in indexed scientific journals, related to sepsis, genetic polymorphisms, and critical care medicine.

Roberto Hornero received the M.S. degree in telecommunication engineering and the Ph.D. degree from the University of Valladolid (Spain), in 1995 and 1998, respectively. He is currently Professor in the Department of Signal Theory and Communications at University of Valladolid. His main research interest is spectral and nonlinear analysis of biomedical signals to help physicians in the clinical diagnosis. He founded the Biomedical Engineering Group in 2004, whose research interests are connected to the field of nonlinear dynamics, chaotic theory, and wavelet transform with applications in biomedical signal and image processing. He is a senior member of the IEEE.

Eduardo Tamayo was graduated in Medicine and Surgery at the University of Valladolid. He is specialist in Anesthesiology and Resuscitation and Doctor of Medicine and Surgery. Currently, he is Professor of Anesthesiology at the University of Valladolid, with assistive labor at the Hospital Clínico Universitario (Valladolid). His research interest includes sepsis, renal dysfunction, and cognitive alterations in postoperative patients.