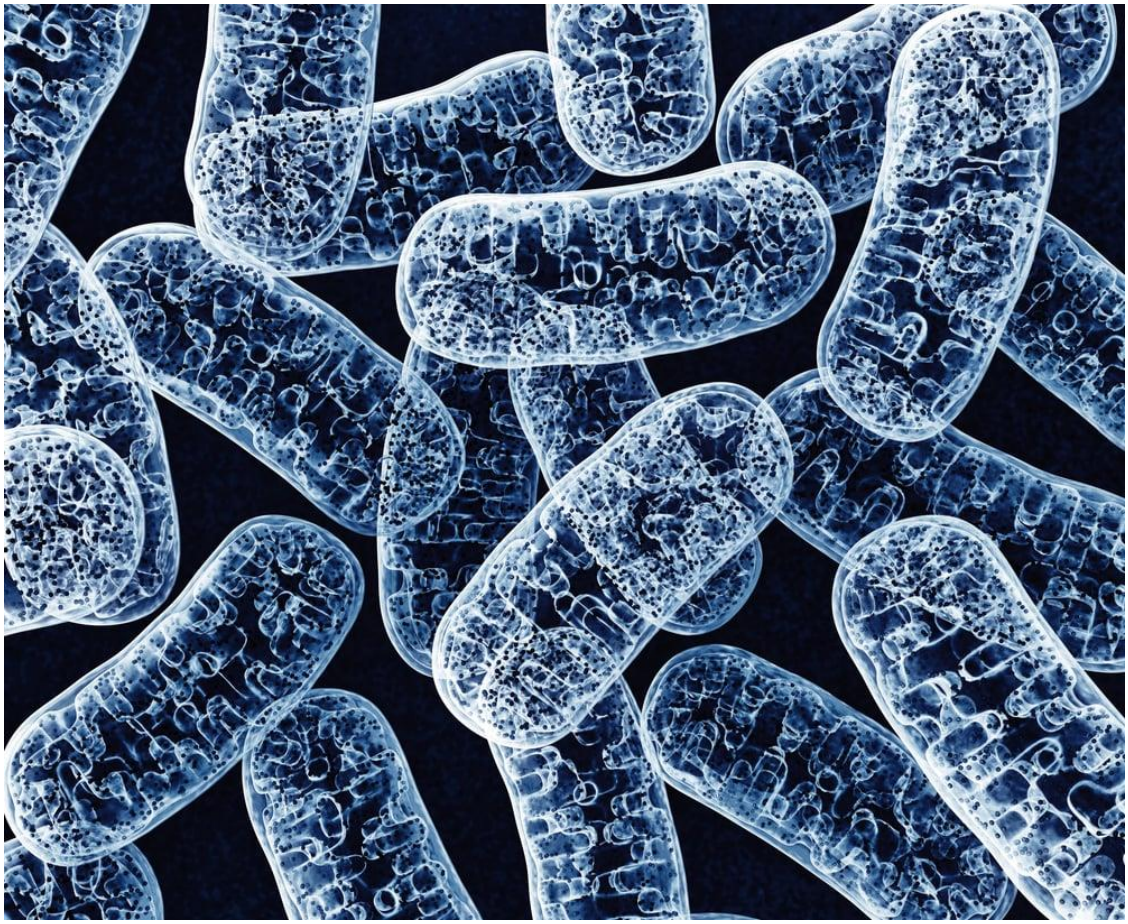
Unraveling the role of mitochondria dysfunction in early sepsis

Identifying and predicting mitochondrial-driven endotypes in early-stage sepsis

Jamie Scheper

26-01-2024



Name: Jamie Scheper
Student ID: 373689
Class: BFV-4
Date: 26-01-2024
Supervisor: Fenna Feenstra

Unraveling the role of mitochondria dysfunction in early sepsis

Identifying and predicting mitochondrial-driven endotypes in early-stage sepsis

Name:	Jamie Scheper
Student ID:	373689
Study:	Bioinformatics (BFV-4)
School:	Hanze UAS
Institute:	Life Sciences & Technology (ILST)
Supervisor Hanze:	Fenna Feenstra
Intern supervisor:	Jingyi Lu and Prof. Dr. H.R. Bouma
Date of completion:	26-01-2024
Status:	Not Confidential

Abstract

Sepsis, a major contributor to in-hospital deaths, is challenging to detect due to its heterogeneous nature, and dysfunctional mitochondria can worsen the condition. Analyzing mitochondria-related genes could aid in early detection and guide treatment.

RNA-Seq and clinical data from 348 septic patients, collected from four emergency rooms (ER) and one intensive care unit (ICU), were used and compared to 44 healthy controls. This study applied supervised and unsupervised algorithms to identify clinically relevant gene signatures.

Mitochondria-related differentially expressed genes (DEGs) were identified by comparing levels of severity based on Sequential Organ Failure Assessment (SOFA) scores. These DEGs helped to establish two unique severity-based endotypes in ER cohorts, which were distinct from the healthy controls and validated on an ICU cohort. This categorization addresses sepsis heterogeneity. Feature selection identified three gene sets that could accurately predict endotype groups. Notably, a logistic regression algorithm with L1 regularization predicted endotypes with 96% and 93% accuracy in ER and ICU cohorts using an eleven-gene set.

The gene signatures and endotypes indicated the essential role of mitochondrial genes in early sepsis detection and as potential novel biomarkers. Future research should aim to develop a multi-modal prediction tool for enhanced patient stratification.

Abbreviations

Full term	*Abbreviation*
Absolute Neutrophil Count	ANS
Acute Physiology and Chronic Health Evaluation	APACHE
Adenosine Triphosphate	ATP
Area Under the Curve	AUC
Cumulative Distribution Function	CDF
Damage-associated Molecular Pattern	DAMP
Differentially Expressed Gene	DEG
Differentially Expression	DE
Emergency Room	ER
Exploratory Data Analysis	EDA
Fold Change	FC
Intensive Care Unit	ICU
Least Absolute Shrinkage and Selection Operator	LASSO
Institute for Life Science and Technology	ILST
Median Absolute Deviation	MAD
Partition Around Medoids	PAM
Principal Component Analysis	PCA
Quick SOFA	qSOFA
Reactive Oxygen Species	ROS
Receiver Operating Characteristic	ROC
Recursive Feature Elimination with Cross-Validation	RFECV
Sequential Organ Failure Assessment	SOFA
Variance Stabilizing Transformation	VST
White Blood Cell Count	WBC
Within-Cluster Sum of Squares	WSS

Organization

This project was provided by Acutelines, a biobank and research group that focuses on acute diseases such as sepsis. The mission of Acutelines is to support acute and emergency medicine research by offering a foundation for novel clinical studies, and to enable the development of personalized medicine through data, images, and biomaterials.

This project was supervised by Prof. Dr. H.R. Bouma and daily supervisor Ph.D. Candidate Jingyi Lu.

Table of Content

Introduction

Sepsis is defined as life-threatening organ failure due to a dysregulated host response to infection. [1] Over the past few decades, mortality related to severe sepsis has gradually declined, which can be attributed to the timely use of antibiotics, prevention, and understanding of molecular pathways. [2, 3] However, sepsis remains one of the largest contributors to in-hospital mortality rates and high healthcare costs. This can be attributed to the absence of targeted medication despite various medical breakthroughs. [4] Patients with sepsis have a high rate of hospital admission; the average length of stay was 75% greater than in other conditions. [5, 6] There were a total of 48.9 million sepsis cases worldwide in which one in five patients perished, which adds up to almost 20% of global deaths. [4]

The variability in disease expression, site of infection, and host response makes early recognition and adequate treatment difficult. Such treatments are currently limited to antibiotics and supportive care. The heterogeneous nature of sepsis was a reason to redefine sepsis as a syndrome, which gives rise to subclasses of sepsis ("endotypes"). [1] Endotypes are defined by distinct biological mechanisms. Previous research indicated the existence of two to five endotypes of sepsis. [7, 8, 9] One such study, conducted by Baquir *et al*., used transcriptional and clinical data to identify gene expression profiles and to stratify patients into five distinct endotypes through unsupervised machine learning techniques. [9] According to the authors, stratifying patients through whole-blood RNA-Seq may optimize antibiotic treatment and address the heterogeneity of sepsis by identifying biomarkers. This way, more personalized therapeutics could be developed. A unique gene profile characterized each endotype and represented a different severity status of sepsis based on sequential organ failure assessment (SOFA) scores. A SOFA score is based on the assessment of six different organ systems. Each organ system gets a score between 0 and 4, with higher scores indicating severe dysfunction. A total score of 24 is, therefore, the maximum score. [10] SOFA scores are an important treatment and diagnostic tool for septic patients. In addition, a quick SOFA (qSOFA) is a simplified version based on three criteria and is used in the ER to give a quick assessment of a patient's health. [11] A qSOFA score of two or higher indicates sepsis and a need for further testing. [11] Acute Physiology and Chronic Health Evaluation (APACHE) III scores are commonly employed to evaluate overall patient severity within the first 24 hours of ICU admission. [60]

Mitochondria play an essential role in cellular homeostasis: producing adenosine triphosphate (ATP) for a continuous energy supply, managing reactive oxygen species (ROS), and regulating inflammatory and apoptosis pathways. [12] During sepsis, an increase in ROS could lead to a disturbance in the redox balance of mitochondria. This loss of balance gives rise to a state of oxidative stress, damaging mitochondrial membrane permeability and homeostasis and can potentially lead to cell death. [13] As a result of apoptosis, mitochondrial damage-associated molecular patterns (DAMPs) are released and contribute to an even larger systemic inflammatory response. [14] A well-defined relationship exists between multi-organ failure, elevated levels of oxidative stress, and mortality related to sepsis. [15] Recent studies aiming for establish endotypes have focused on immune-related gene expression profiles, and proved valuable in paving the way for

immune-specific therapeutics. [9] However, mitochondrial dysfunction significantly contributes to the heterogeneity of sepsis. If this role is not utilized, it would represent a substantial oversight. Stratifying specific mitochondrial-driven endotypes may identify distinct subgroups or stable biomarkers. When combined with immune-related endotypes, this approach served as a robust predictive tool, which could enhance early detection of sepsis, open doors to new personalized therapeutics, and ultimately improve patient outcomes. For example, a recent study on the preservation of mitochondrial membrane potential based on endotypes demonstrated improved outcomes. [16]

This study attempted to associate mitochondrial dysfunction with clinical parameters by identifying mitochondrial-driven endotypes based on analyzing whole-blood transcriptomics. Emergency room (ER) and intensive care unit (ICU) cohorts from the publicly available dataset (GSEA number: GSE185263) were used to identify mitochondrial-driven severity and mortality gene signatures. This study aimed to establish gene signatures by identifying differentially expressed genes (DEGs) by comparing patients' sepsis conditions. DEGs form a genetic understanding of how sepsis severity and outcomes manifest in different endotypes. A key objective of this study was the development of a prediction tool that could accurately predict mitochondria-driven endotypes and the severity of a patient's condition based on a refined gene set.

Materials and Methods

Study population

The recruitment of patients was based mainly on the suspicion of sepsis, which was determined by at least two Sepsis-1 criteria and the professional opinion of medical doctors. Patients were at least 18 years of age.

The number of patients with suspected sepsis added up to 348 after preprocessing and was derived from a multi-center study; 266 patients originated from four different ERs (Vancouver, Canada, Colombia, the Netherlands, and Australia), and 80 patients came from one ICU (Toronto, Canada). Also, two patients came from a ward but were added to the ICU cohort because of similar patient characteristics. Additionally, 44 healthy control samples were recruited from three hospital biobanks (Sydney, Neiva, and Toronto).

Essential clinical variables were sepsis severity and mortality. The sepsis severity variable consisted of three levels: 'High' (≥ 5 SOFA), 'Intermediate' (≥2 & <5 SOFA), and 'Low' (< 2 SOFA). Mortality data included survival, fatalities, and cases with unknown outcomes. The total amount of clinical variables was 268. The dataset had a variety of severity measurement variables, including qSOFA and SOFA scores for ER, as well as APACHE III and SOFA scores for ICU. The worst SOFA score within 72 hours was used as a benchmark for the sepsis severity variable. Our interest lies in clinical variables that indicate mitochondria dysfunction, such as lactate levels, neutrophil counts, oxygen treatment, length of admission, and microbiological results (see Table 1). [15]

RNA data collection

2.5 ml of blood for RNA sequencing was collected within 72 hours after admission. RNA-Seq was performed with Illumina NovaSeq 6000 S4, which resulted in 100 long base pair reads. After that, 'fastqc' (version 0.11.9) and 'multiqc' (version 1.6) were used for quality control, alignment was conducted with Ensembl's 'STAR' (version 2.7.9a), and read count assessment with 'htseq' (version 0.11.3). [17, 18, 19, 20] RNA-Seq data was composed of roughly 58.000 genes, with 1.690 being related to mitochondria, as determined by the definition used in the MSigDB database (via ID "GO:0005739").

Parameter	*ER (n = 266)*	*ICU (n = 82)*
Age	56,06 ± 1,26 (99,63%)	61,41 ± 1,70 (98,78%)
Gender	F: 45,49%; M: 54,51%	F: 30,49%; M: 68,29%
*In-hospital mortality**	D: 12,03%; S: 87,59%	Not available
*Severity***	H: 12,03%; I: 40,98%; L: 47,00%	H: 63,41%; I: 20,73%; L: 15,85%
ER qSOFA	1,97 ± 0,12 (100%)	Not relevant
SOFA	1,31 ± 0,14 (100%)	7,34 ± 0,55 (100%)
Lactate	2,24 ± 0,25 (18,05%)	Not available
Neutrophil count	6,59 ± 0,41 (56,02%)	Not available
Total white blood cell count	9,39 ± 0,39 (80,45%)	9,48 ± 0,67 (98,78%)
Total admission (in days)	7,53 ± 0,54 (97,74%)	12,04 ± 0,97 (100%)

*Mortality levels: Deceased (D) and Survived (S).
**Severity levels: High (H), Intermediate (I), and Low (L).

Table 1—Essential clinical parameters related to mitochondria dysfunction and sepsis severity of ER and ICU cohorts. The mean and standard error are depicted for numeric variables. For categorical variables, the distribution per level is presented. The number of available total records is presented in percentage (%) for both numeric and categorical variables.

Exploratory data analysis

The clinical and count datasets were analyzed in R (version 4.3.1). [21] The focus of the analysis was on data cleaning, which involved removing missing values or imputing them. We also removed attributes that did not contribute to our research goals and adjusted the structure of valuable attributes. Another key aspect was highlighting the differences between cohorts. We used many external packages to explore the data. For example, the packages 'dplyr' (version 1.1.3) and 'tidyverse' (version 2.0.0) were used extensively to mutate data, remove redundancies, and explore the correlations between attributes. [22] [23] [41] Other external packages had a minor function; for example, 'caret' (version 6.0-94) was used to remove near-zero variance attributes. [24] For visualization purposes, 'ggplot2' (version 3.4.3) was used throughout the project. [25] All transformations related to data structure, along with the rationale for these choices, are thoroughly explained in the provided Exploratory Data Analysis (EDA). This document as well as all research logs can be found in the accompanying Bitbucket repository. [26] Additionally, outliers were assessed by using an IsolationForest model via the Python package 'sklearn' (version 1.3.2;

also known as 'scikit-learn') and can be found in the 'outlier_detection.ipynb' notebook. Furthermore, helper functions regarding clustering, EDA, and other R scripts can be found in 'helper_functions.R'.

Differential expression and pathway analysis

Identification of differentially expressed genes was undertaken with 'DESeq2' (version 1.42.0) and, to validate our findings, we used 'egdeR' (version 4.0.2). [27] [28] We compared gene expression of patients from both the ER and ICU with a focus on sepsis severity status and mortality outcome. Low-expressed genes were filtered out if counts were less than ten over at least ten samples. In addition, genes were considered a DEG when displaying an adjusted p-value of ≤ 0.05 based on a Wald test and ≥ 1.2 absolute fold change (FC). Due to this low amount of DEGs in mortality, we primarily focused on establishing severity signatures. We compared each level from the sepsis severity variable to each other and extracted each unique DEG across all comparisons.

Pathways in which DEGs reside were identified by using the 'ReactomePA' (version 1.46.0 with database 'Reactome') and MSigDB Hallmark database via the R-based 'enrichR' (version 3.2) package. [29] [30] Pathways were considered significant when they exhibited an adjusted p-value of ≤ 0.05 (based on a hypergeometric test via enrichR's internal operations).

DESeq2's internal normalization accounted for variations in factors such as sequencing depth. To ensure a similar correction was applied for downstream analysis, we transformed the raw count data using DESeq2's variance stabilizing transformation (VST). We also employed batch correction using the 'ComBat' from the R package 'SVA' (version 3.50.0) to minimize noise introduced by batches. [31]

Unsupervised clustering

Unsupervised clustering machine learning models were employed to establish severity-based endotypes. We looked at three methods: partition around medoids (PAM), K-means, and hierarchical clustering. We built dendrograms for hierarchical clustering via 'dendextend' (version 1.17.1). [32] For each unsupervised clustering method, we compared different distances and, where appropriate, linkage methods. We determined cluster stability using several metrics, including the average silhouette score, elbow plots, and consensus clustering via the R-based package 'ConsensusClusterPlus' (version 1.66.0). [33] We determined the optimal number of clusters based on the metric results of within-cluster sum of squares (WSS), silhouette scores, and GAP scores via packages 'factoextra' (version 2.9) and build-in 'cluster' (version 2.1.6). Additionally, we used 30 different cluster scoring indices of the R package 'NbClust' (version 3.0.1), which gave the optimal number of clusters through majority vote. [34] [35] Subsequently, we performed various DE and pathway analyses to compare these endotypes between groups and healthy control samples on DEGs, but we did not filter on significance or log fold change. Clinical variables were compared across endotypes using Wilcoxon rank-sum tests for numeric attributes and Chi-squared tests for categorical variables, followed by a Benjamini-Hochberg correction for

post hoc analysis. Heatmap visualizations were made with the package 'ComplexHeatmap' (version 2.18.0). [36]

Gene set and prediction model

For this section, we utilized Python (version 3.11.2) and Jupyter for its notebook integration (version 1.0.0). [37] [38] Data manipulation was done with 'pandas' (version 2.1.3) and NumPy (version 1.26.2). [39] [40] For classification, we employed five models: support vector machine (SVM), random forest, naïve Bayes, and logistic regression, the latter including L1 (least absolute shrinkage and selection operator, or LASSO) and L2 regularization. All were applied to the DEG set on septic patients only; we did not use healthy controls. We accomplished this by using the machine-learning package 'sklearn'. [41] We measured the performance of the included classification models with scoring metrics such as F1, accuracy, area under the curve (AUC), and recall scores. In addition, we plotted a receiver operating characteristic (ROC) curve for every comparison in a one-versus-rest fashion when dealing with more than two groups. Additionally, we utilized 'seaborn' (version 0.13.0), 'matplotlib' (version 3.8.2), and 'scikit-plot' (version 0.3.7) for visualization purposes. [42] [43] [44] We trained models on 75% of ER patients; the other 25% served as a test set. The ICU cohort was used as a validation set. Cross-validation was done by splitting the ER cohort with the function 'StratifiedKFold' (k = 5) and, where appropriate, cross-validating with a leave-one-out approach due to class imbalances with High severity. Selection and extraction methods were performed to combat overfitting observed when classifying with all DEGs on extreme phenotypes (High + Intermediate vs. Low). We used the extreme phenotype classification to combat the class imbalance. We used a recursive feature elimination with cross-validation (RFECV) based on the random forest algorithm from the sklearn library. However, random noise made the most optimal gene set challenging to determine. In the end, we assessed gene importance using logistic regression with L1, L2, and ElasticNet regularization alongside model-independent mutual information. We then incrementally added the most significant gene and assessed performance based on cross-validation to refine the model. We stopped adding genes at a point when scores no longer improved and began to dwindle. Feature extraction via principal component analysis (PCA) was also undertaken on reduced gene sets. Lastly, the final model was optimized and validated on the ICU cohort. Furthermore, helper functions regarding machine learning and feature selection can be found in 'helper_functions.py'.

A flowchart summarizes all the steps mentioned above (Figure 1).

Flowchart

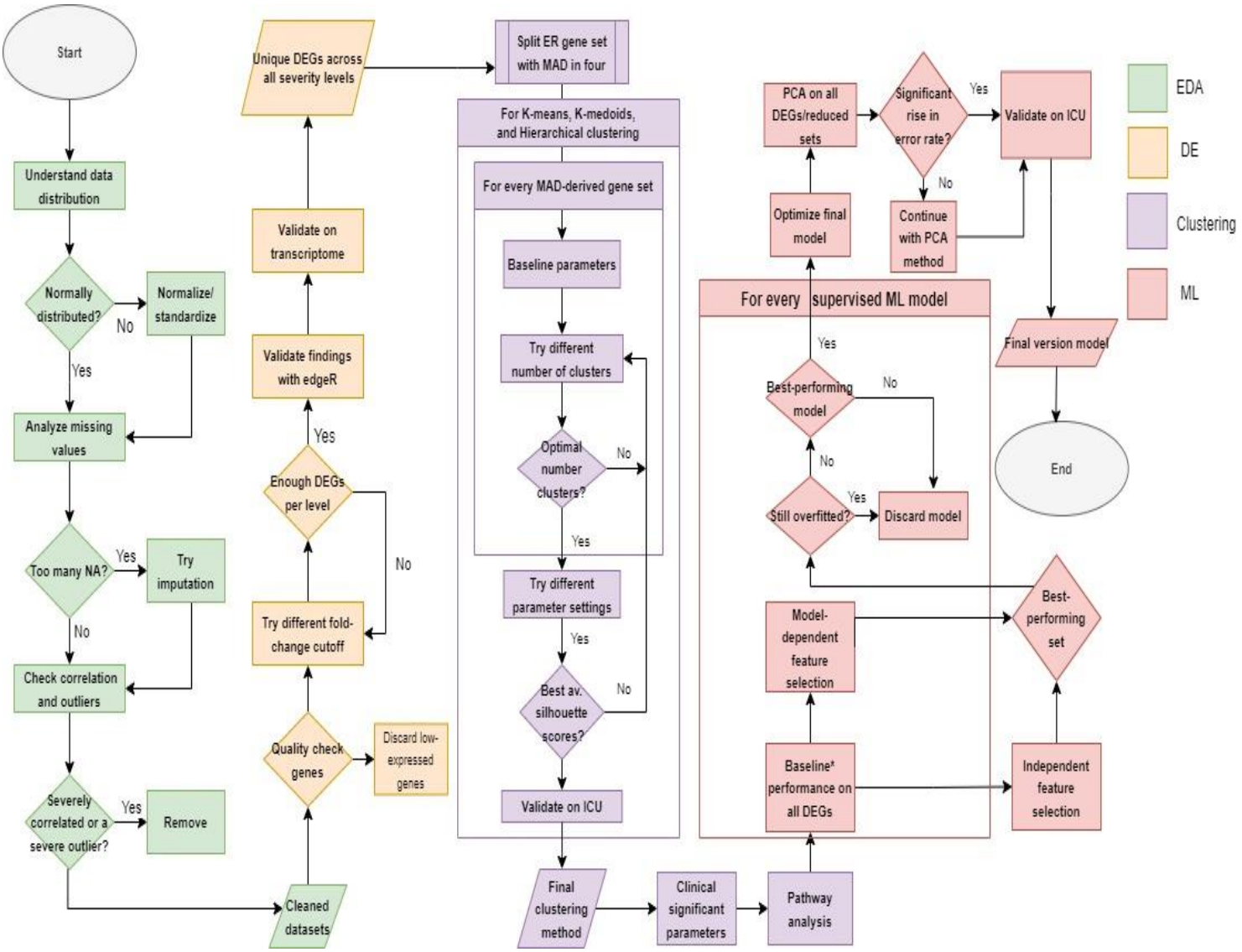


Figure 1—Flowchart depicting the project workflow. The exploratory data analysis section was highlighted in green. This section details choices made regarding handling missing values, corrections, and outliers. After that, the DE analysis section, shown in yellow, focused on determining an appropriate fold-change cutoff to identify enough DEGs for each severity level. We evaluated various unsupervised clustering methods (in purple) by utilizing multiple gene sets based on MAD (see Materials and Methods) and selected the best performing method. We used the established clusters to extract clinically significant parameters and identify upregulated and downregulated pathways. In red, machine learning was used to predict severity and endotype groups. We tested various supervised methods and selected the best-performing model, after conducting various feature selection methods.

Results

Discovery of DEGs and associated pathways

We included all 348 patients (ER and ICU) in a differential expression (DE) analysis to establish sepsis severity and mortality gene signatures. However, mortality did not exhibit sufficient DEGs and correlated with severity. Therefore, using both would not add value to our research goals and we preferred to use sepsis severity as our primary focus. We compared all groups with each other and extracted unique genes that had sufficient fold changes (1.2) and had a significant adjusted p-value of ≤0.05. When using DESeq2 we identified 201 unique DEGs out of a total 1.488 mitochondria-related genes. To validate our results, we compared our findings to edgeR (Figure 2A) while using the same parameters where possible. Both methods labeled 180 genes as differentially expressed, while edgeR found three and DESeq2 unique 21 genes, respectively. Most DEGs were identified by both methods and some discrepancies were expected due to their differing calculation methods.

Figure 2B depicts how groups of sepsis severity based on gene expression of identified DEGs, cluster together by dimension reduction through PCA. This distribution showed the challenges of sepsis heterogeneity. We highlighted the DEGs found in the High vs. Low comparison in Figure 2C, which yielded the highest number of DEGs (180) among all comparisons These 180 DEGs accounted for 88,20% of all unique DEGs. High vs. Intermediate showed 20 and Intermediate vs. Low 54 (see Supplementary S2) DEGs, respectively.

Up and downregulated pathways were identified by comparing unique DEGs per group using Reactome and MSigDB datasets (Figures 3A, B). Comparing High vs. Low severity groups yielded the most significantly altered pathways, especially those related to mitochondrial dysfunction. Examples included downregulated electron transport and mitochondrial translation pathways. However, only a few pathways were significantly altered based on High vs. Low when we used MSigDB (Figure 3B). These pathways are related to oxidative phosphorylation, cell-mediated apoptosis, and the breakdown of reactive oxygen species. In contrast to Figure 3A, the High vs. Intermediate comparison did not result in significantly altered pathways.

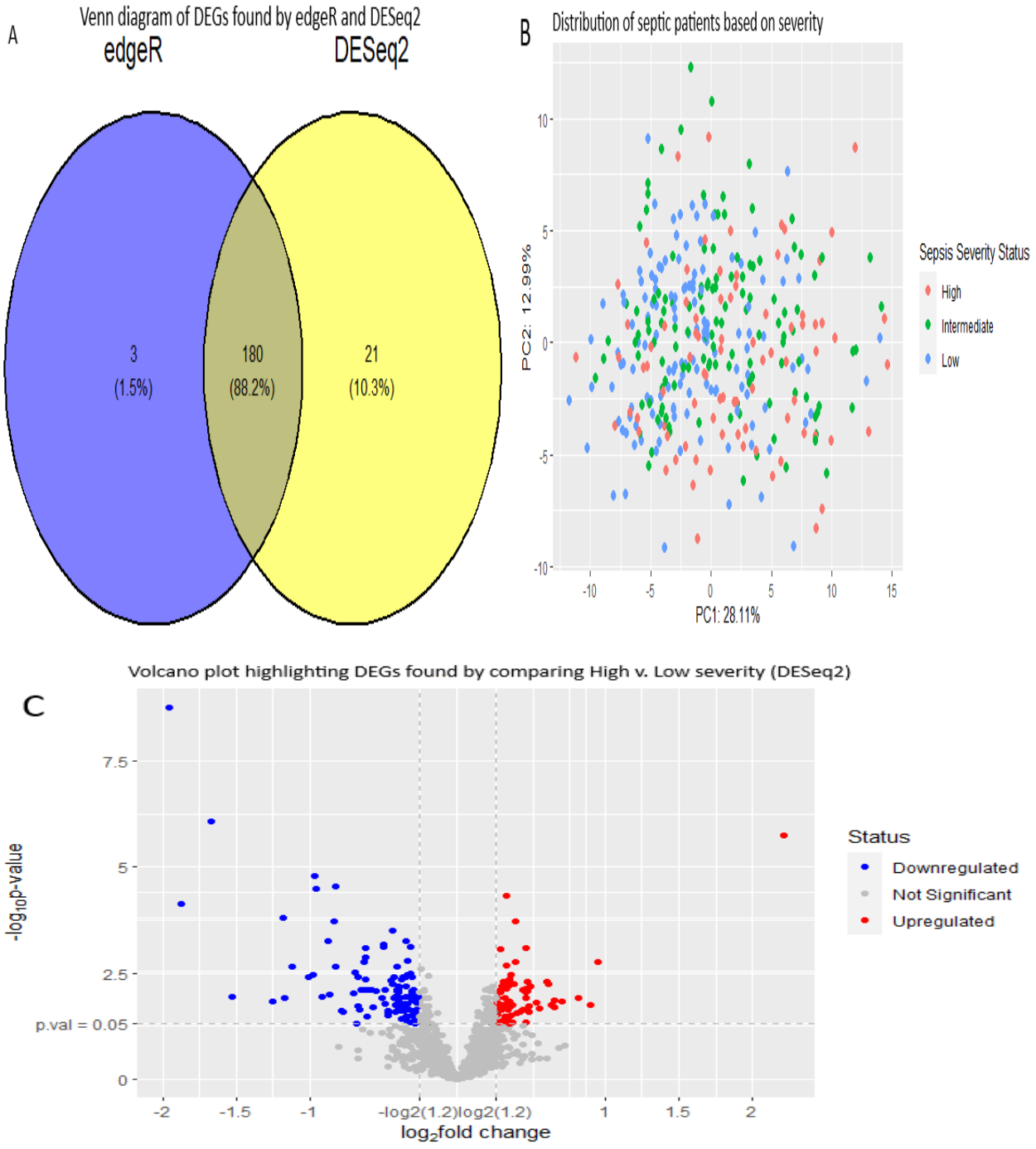


Figure 2– Discovery of DEGs related to sepsis severity. A) Venn diagram showing the DEGs identified by DESeq2 and validation method edgeR. There is a notable overlap between both methods (88,20%). B) Clustering based on the 201 DEG set found by DESeq2 on sepsis severity status indicated the instability of the groups. C) Volcano plot of found DEGs (adjusted *p-value* of 0.05; absolute FC of 1.2) when comparing High vs. Low sepsis severity groups revealed a total of 180 DEGs.

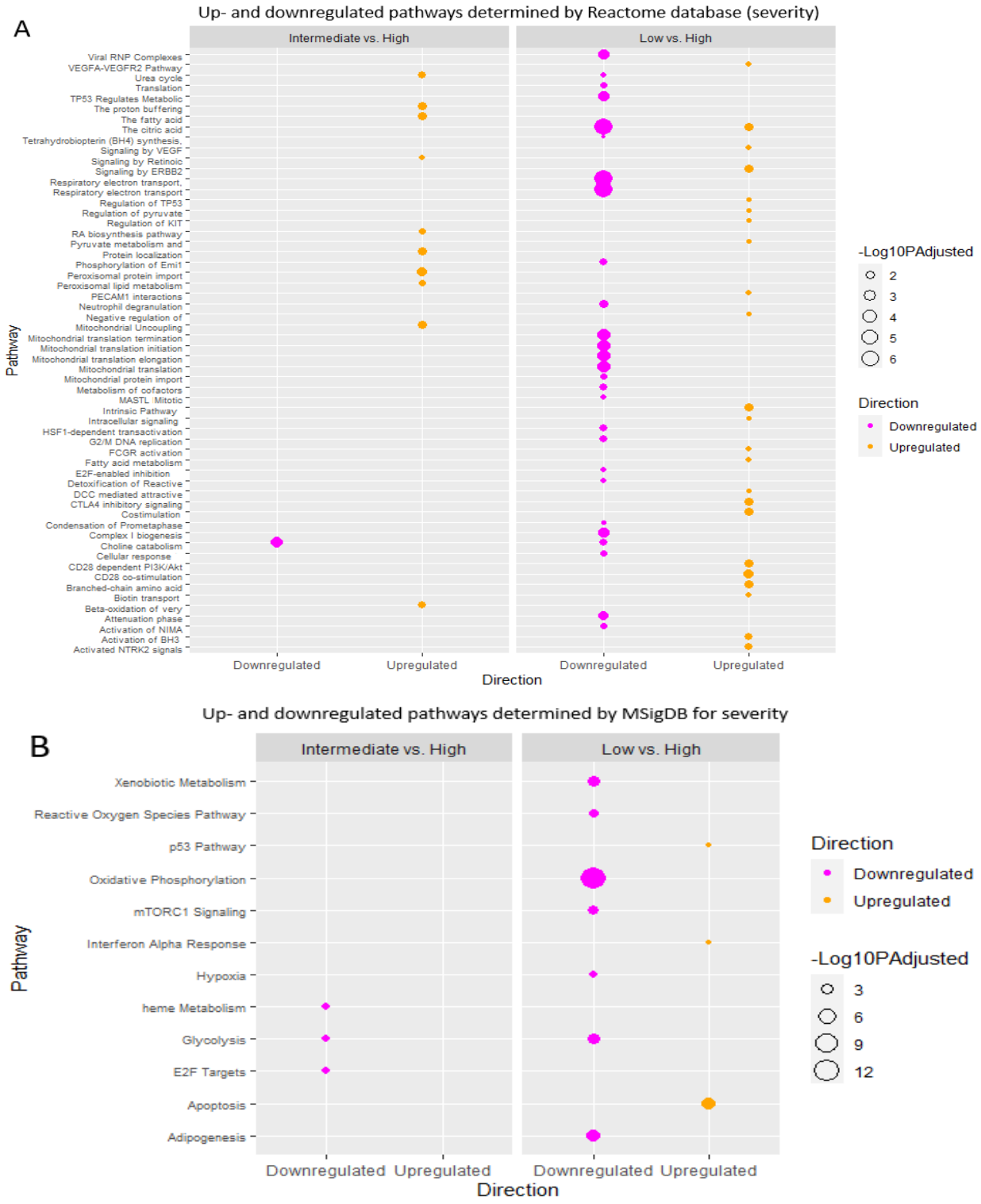


Figure 3– Identification of up and downregulated pathways based on severity. A) The Reactome database showed many significant up- and downregulated pathways (adjusted *p-value* of 0.05) for every comparison between sepsis severity groups. B) The MSigDb Hallmark 2020 database identified up- and downregulated pathways. Only the comparisons involving the High sepsis severity group exhibited significantly expressed pathways.

Establishing severity-based clusters

The found DEGs acted as molecular markers to distinguish different levels of sepsis severity based on gene expression patterns. We proceeded to cluster patients into distinct severity endotypes based on these DEGs.

We considered three different clustering methods: PAM, K-means, and hierarchical clustering. After testing on differently sized gene sets based on median absolute deviation (MAD) on the ER cohorts (n = 266), we decided that PAM was the best clustering method out of the three used (Figure 4B, Table 2). PAM with Manhattan distance provided the most stable clusters for the training and test sets, with average silhouette scores of 0,18 and 0,23 (on k = 2), respectively (Figure 4C, D).

After testing different gene sets based on MAD and three optimization metrics (Table 3), two clusters proved to be the optimal number of clusters. In addition, consensus clustering on PAM revealed that 2 clusters were preferred (Figure 4A). When validating on the ICU cohort, we observed comparable results and determined that two clusters were also the most optimal number of clusters here (see Appendix I). Beyond clustering primarily on DEGs, we also performed clustering on the entire population for unbiased validation. Clustering on the entire population of mito-genes (1.488) revealed the existence of two clusters and provided the most stable results (see Appendix I).

However, PAM was not the preferred clustering method for all mito-genes. In contrast, K-means with Euclidean distance exhibited stable clusters, achieving an average silhouette score of 0,42 on two clusters (see Supplemental S3). Nevertheless, when performing on the same parameter settings on DEGs, results were lackluster in comparison with a silhouette score of 0,15 (see Supplemental S3). Therefore, PAM was still selected as the preferred clustering method.

Clustering method	*ER*	*ICU*
K-means (Euclidean)	0,18	0,19
Hierarchical (Euclidean + Ward.D2)	0,15	0,16
PAM (Euclidean)	0,16	0,20
PAM (Manhattan)	0,18	0,23

Table 2—All clustering methods were used with scoring based on average silhouette scores for two clusters regarding ER (training) and ICU (test/validation). PAM had slightly better average silhouette scores for both the training and test/validation and was subsequently chosen as the preferred clustering method.

% of total genes based on MAD	*WSS*	*Silhouette score*	*Gap score*
25%	2-3	2	6
50%	2-3	2	7
75%	2-3	2	7
100%	2-4	2	2

Table 3—Optimal cluster numbers were determined based on the within-cluster sum of squares (WSS), silhouette scores, and gap scores for different ranked-based sets of genes, using median absolute deviation (MAD) in batches of 25% for PAM.

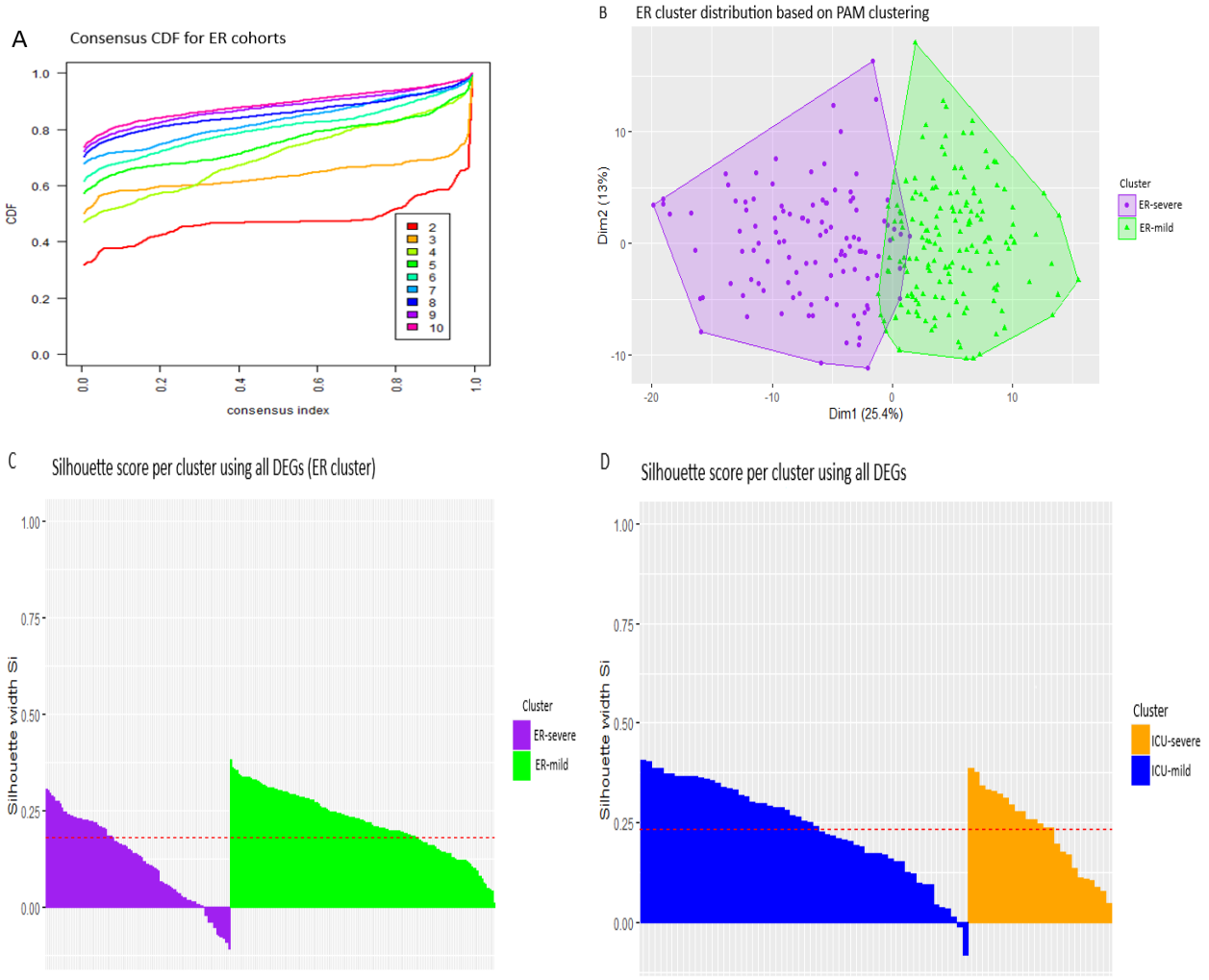


Figure 4–Clusters established based on PAM. A) The Consensus cumulative distribution function (CDF) plot indicates two as the most stable number of clusters (red line) for ER cohorts. B) PAM clusters with 'ER-severe' in purple and 'ER-mild' in green. C) Cluster stability (PAM) based on silhouette scores with 'ER-severe' in purple and 'ER-mild' in green. D) Cluster stability (PAM) based on silhouette scores with 'ICU-severe' in orange and 'ICU-mild' in blue.

When zooming in on the samples that exhibited the lowest silhouette scores, most belonged to either the Intermediate or Low severity subgroup. This indicated that High severity was more stratified into one of the cluster groups and hinted at a stable distinction in severity between clusters (see Figure 5).

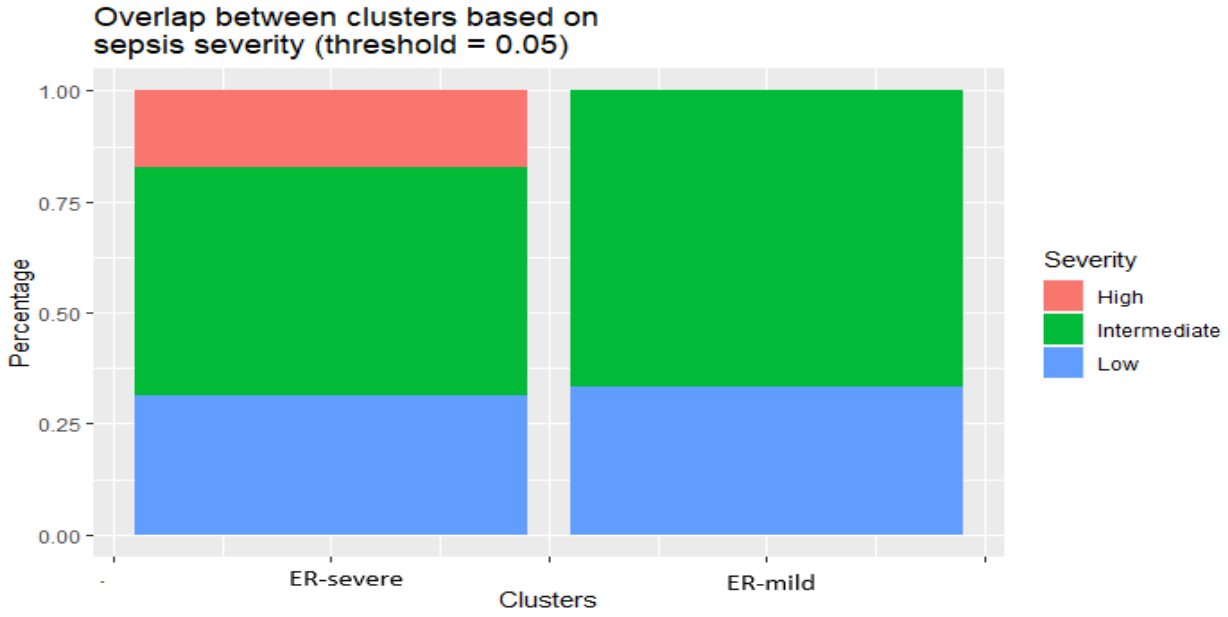


Figure 5—Overlap between ER-based clusters was observed based on low silhouette scores (≤ 0.05). ER-severe exhibited a more diverse set of severity levels at the border between both clusters, whereas ER-mild contained only Intermediate severity samples.

Clusters show clinically relevant differences

Establishing two endotypes provided a more meaningful way to categorize sepsis based on mitochondria-related genes. We investigated whether these clusters differed significantly in severity by comparing them with each other and with healthy controls.

When comparing the established severity-based clusters, we found that 44 variables for the entire ER population are considered significant (Table 4). Some significant clinical variables were related to SOFA, such as "ED SOFA (first)" (adj. p-value of 2.427e-05), "Sepsis Severity" (adj. p-value of 1.012e-04), and "SOFA (W72h)" (adj. p-value of 0.00261e-03). In Table 4, we also observed some crucial indicators of mitochondrial dysfunction due to oxidative stress, including "ED Lactate (first)" (adj. p-value of 3.183e-04), "Treatment Oxygen Therapy" (adj. p-value of 0.00251e-03), and "Lungs Fio2 (W72h)" (adj. p-value of 1.012e-04), the latter indicating an estimation of the oxygen content a patient inhaled. Additionally, important molecular metrics indicating the severity of sepsis were present, including total cell count ("Total Cell Count (W72h)" with adj. p-value of 2.097e-04) and specific classes of immune cells such as neutrophils ("ED Neutrophil count (first)" with adj. p-value of less than 4.693e-06).

Parameter	*ER-severe (n = 109)*	*ER-mild (n = 151)*	*p-value (adj.)*
*Respiratory Rate (W72h)**	21.98 ± 5.78 (91)	20.03 ± 4.54 (115)	0.0488
*Diastolic (W72h)**	58.31 ± 11.8 (80)	62.79 ± 10.33 (77)	0.0358
At Ed Altered Mental State	100% (109/109)	100% (157/157)	0.0310
Outcome ICU Admission	100% (109/109)	100% (157/157)	0.0302
*Heart Map (W72h)**	82.49 ± 14.47 (85)	87.53 ± 12.22 (109)	0.0294
Treatment Hospital Admission	100% (109/109)	100% (157/157)	0.0294
*Systolic (W72h)**	102.18 ± 18.32 (79)	109.65 ± 15.48 (77)	0.0251
Sample Location Colombia	100% (109/109)	100% (157/157)	0.0184
Bands Greater 5 Percent (first)	76.15% (83/109)	85.99% (135/157)	0.0184
Coagulation Pt (first)	18.16 ± 18.97 (30)	13.08 ± 5.51 (37)	0.0184
*Heart Rate (W72h)**	101.89 ± 20.85 (90)	93.86 ± 19.7 (116)	0.0156
Eosinophil Count (first)	0.04 ± 0.08 (86)	0.09 ± 0.16 (129)	0.0123
Coagulation (first)	1.62 ± 1.56 (35)	1.12 ± 0.41 (40)	0.0123
Next 72 Sepsis	100% (109/109)	100% (157/157)	0.0123
qSOFA	1.15 ± 0.83 (109)	0.82 ± 0.76 (157)	0.00760
Outcome Organ Failure	24.82 ± 8.05 (109)	27.13 ± 3.72 (157)	0.00760
Additional Test Procalcitonin	11.54 ± 20.14 (27)	4.09 ± 14.61 (35)	0.00760
*Body Temperature (W72h)**	37.79 ± 1.01 (85)	37.13 ± 3.04 (108)	0.00719
*Lactate (W72h)**	2.61 ± 1.78 (34)	1.34 ± 1.12 (14)	0.00408
Basophil Count (first)	0.14 ± 0.97 (86)	0.09 ± 0.16 (128)	0.00361
Micro Blood Culture Pathogen	98.17% (107/109)	97.45% (153/157)	0.00303
*SOFA (W72h)**	1.98 ± 2.85 (109)	0.84 ± 1.58 (157)	0.00261
Treatment Oxygen Therapy	99.08% (108/109)	100% (157/157)	0.00251
Treatment ICU Admission	100% (109/109)	100% (157/157)	0.00244
Sepsis All 72	100% (109/109)	100% (157/157)	0.00148
Outcome Hospital Stay Days	9.69 ± 10.57 (107)	6.02 ± 6.76 (153)	0.00109
Total Cell Count (first)	11.71 ± 7.73 (109)	8.32 ± 4.58 (156)	8.969e-04
Treatment Antibiotics Given	100% (109/109)	100% (157/157)	7.750e-04
Liver Bilirubin (first)	19.75 ± 33.06 (96)	13.1 ± 14.87 (107)	3.319e-04
Lactate (first)	2.04 ± 1.49 (82)	1.25 ± 0.55 (79)	3.183e-04
Sepsis 3 Excl 1	100% (109/109)	100% (157/157)	2.945e-04
*Total Cell Count (W72h)**	11.17 ± 6.11 (98)	7.88 ± 5.01 (116)	2.097e-04
*Neutrophil Count (W72h)**	8.6 ± 4.98 (60)	5.23 ± 4.49 (89)	1.841e-04
Sepsis Org Dys BC	98.17% (107/109)	97.45% (153/157)	1.606e-04
Sepsis 3 Excl 1 2	100% (109/109)	100% (157/157)	1.469e-04
Sepsis (first)	100% (109/109)	100% (157/157)	1.434e-04
Sepsis 3	100% (109/109)	100% (157/157)	1.434e-04
Sepsis 3 BC	98.17% (107/109)	97.45% (153/157)	1.434e-04
*Lungs Fio2 (W72h)**	31.24 ± 16.76 (37)	21.92 ± 2.76 (53)	1.012e-04
Sepsis Severity	100% (109/109)	100% (157/157)	1.012e-04
SOFA (first)	2.69 ± 2.15 (109)	1.48 ± 1.67 (157)	2.427e-05
Neutrophil Count (first)	10.33 ± 7.22 (90)	5.98 ± 3.96 (138)	4.693e-06
Lymphocyte Count (first)	0.9 ± 0.88 (87)	1.42 ± 1.53 (134)	4.693e-06
Endotype	100% (109/109)	100% (157/157)	6.118e-26

*W72h = Worst within measurement 72h after admission

Table 4–Variables related to ER patients, which were tested on significance between clusters using Wilcoxon rank-sum test and Chi-Squared test for numeric and categorical (*p-value adjusted* 0.05). After that, a post-hoc analysis using the Benjamin-Hochberg method was conducted. For both clusters, the mean value, standard error, and total available values were included for numeric variables, with the share of available data in percentage.

Cluster ER-severe contained the most severe sepsis cases, as the SOFA scores were higher for qSOFA and the highest measurement of SOFA within the first 72 hours after admission compared to those observed in cluster ER-mild (Figure 6A, D). In addition, biomarkers indicative of mitochondrial dysfunction due to oxidative stress were also higher in cluster ER-severe (Figure 6B, C). These included metrics such as lactate and neutrophil counts. There was a difference in mean SOFA scores between both clusters: ER-mild exhibited a mean SOFA score of 0,768 and ER-severe 2,017.

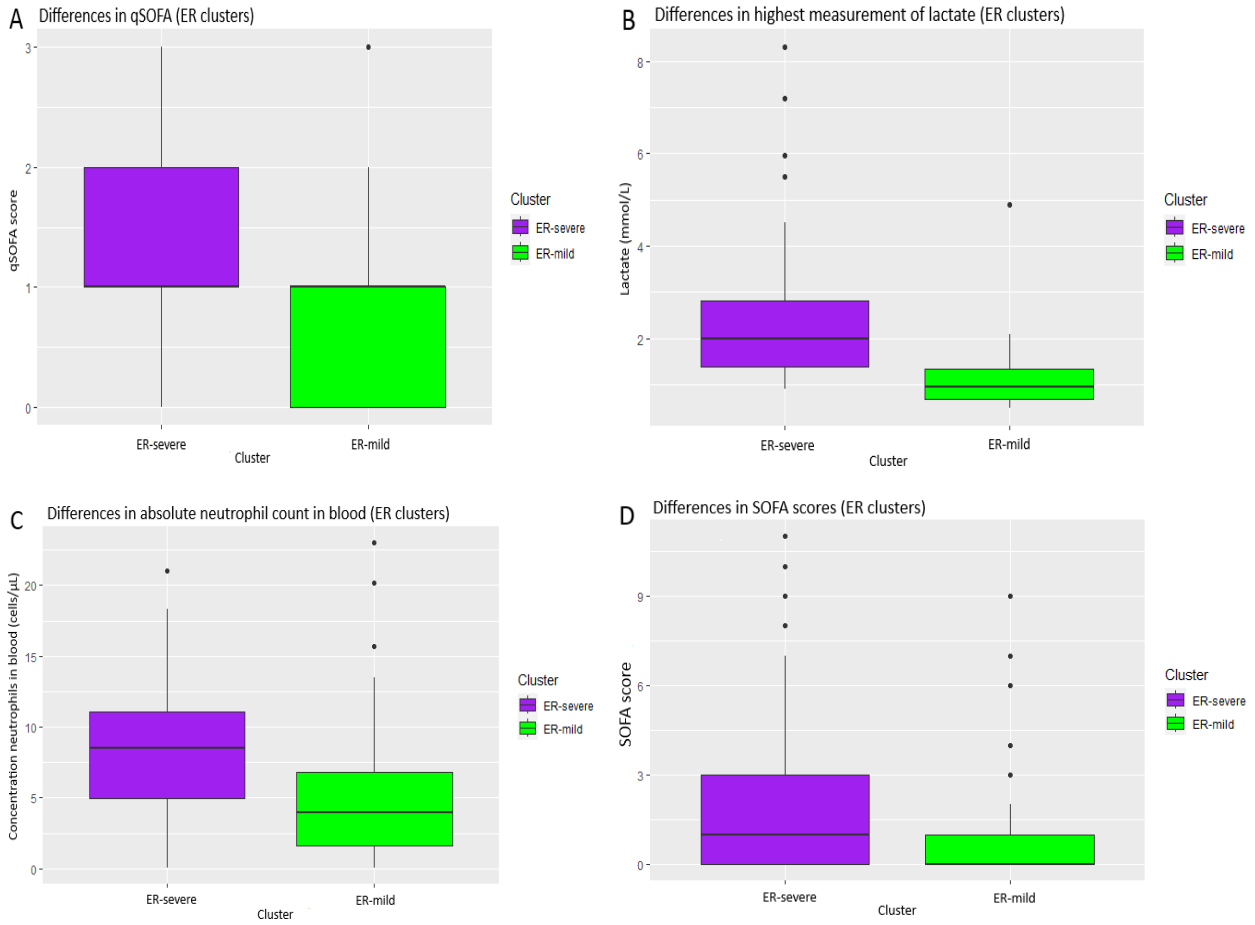


Figure 6—Clinical differences between cluster groups (ER). Cluster 'ER-severe' depicted in purple, and cluster 'ER-mild' in green. A) Cluster ER-severe showed higher values of qSOFA scores than observed in cluster ER-mild. B) Higher rates of lactate were observed in cluster ER-severe. C) The highest neutrophil count in the first 72 hours after admission was worse in cluster ER-severe. D) Cluster ER-severe exhibited higher SOFA scores than cluster ER-mild.

When comparing up and downregulated pathways based on FC changes between the two established clusters with a healthy control population (n = 44), it was observed that pathways related to mitochondria were enriched. These include pathways related to cell electron transportation, mitochondrial translation, elongation, and termination factors. Also, various immune-related (e.g., neutrophil degranulation and CD28) were upregulated for both cluster ER-severe and ER-mild. Cluster ER-severe exhibited various downregulation in pathways related to apoptosis, regulation of TP53, and respiratory electron transportation. These observations suggested a differential regulatory response when compared to the healthy control population (Figure 7A). When comparing gene expression patterns between both clusters, cluster ER-severe contained a larger share of High severity and had higher SOFA scores than cluster ER-mild (Figure 7B).

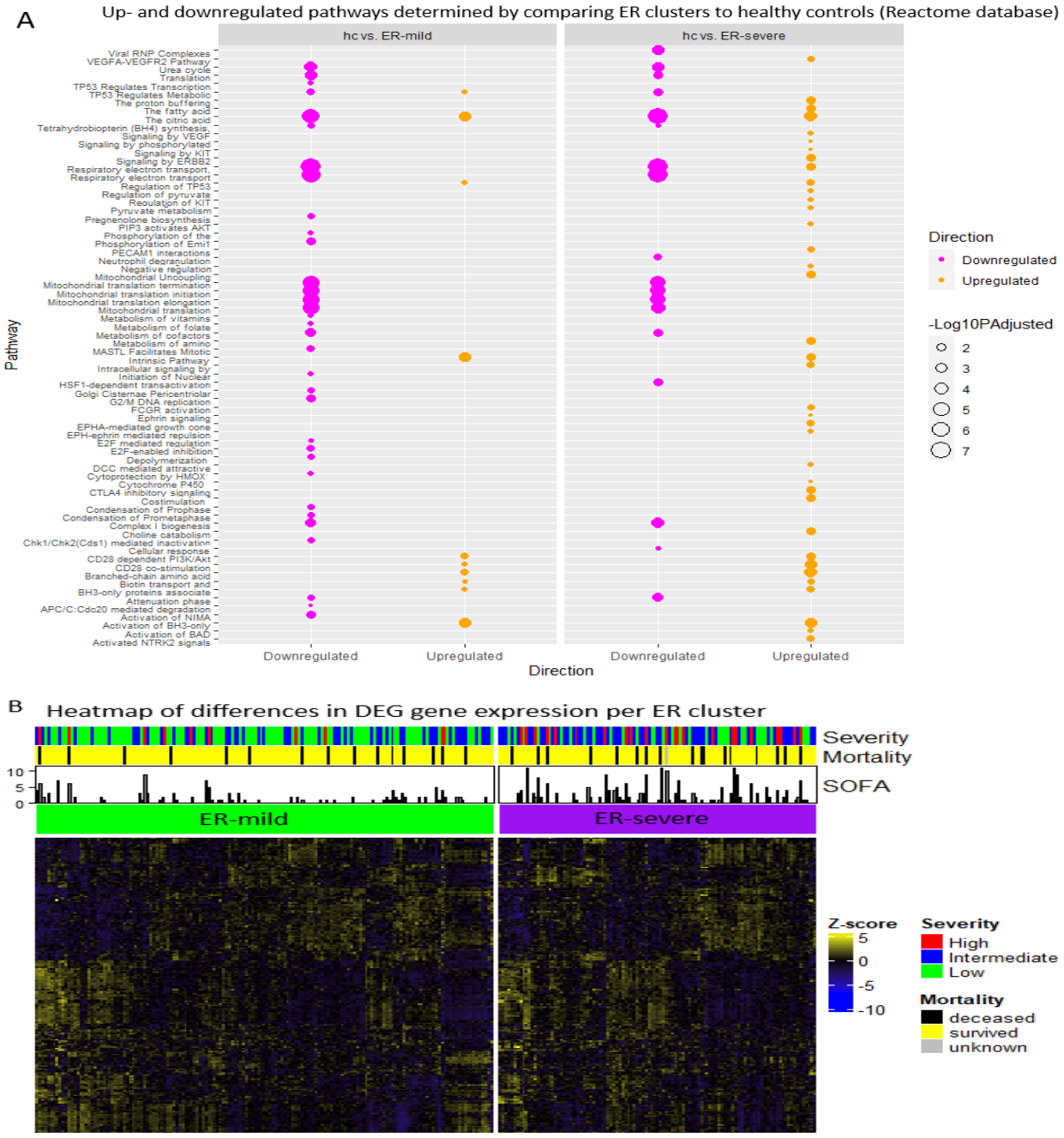


Figure 7—Differences between cluster groups for ER. A) Pathway analysis on cluster groups against a healthy control population of 44 patients. Only significant pathways with a *p*-value of less than 0.05 were incorporated. These were related to apoptosis and TP53. In both clusters, we observed various upregulated pathways, some of which were related to immune cells (neutrophil degranulation), mitochondria-related transcriptional factors, and the respiratory system. B) Heatmap of the two ER-based clusters with gene expression patterns from all 201 DEGs. Cluster ER-severe exhibited higher amounts of High severity (red) than we observed in cluster ER-mild. In addition, the highest measured SOFA scores in the first 72 hours after admission were higher in cluster ER-severe.

When validating the results from clustering on the ICU cohort, we observed that similar clinically relevant variables were significant compared to the ER cohort (Table 5). For instance, multiple parameters regarding SOFA scores proved to be of significance. These included "ICU Outcome SOFA (first)", which highlighted the first measurement of SOFA in an ICU setting (adj. p-value of 0.0022), and "ICU SOFA" (adj. p-value of 0.0370) regarding the highest measured SOFA score during admission. Similarly to the ER cohorts, we found that neutrophil count and white blood cell measurements ("ICU ANC" and "ICU WBC" with an adj. p-value of <0.0001) were significant. Additionally, some of these variables were clearly different between the two clusters, as shown in Figure 8A-C. Here, the first ICU-severe exhibited higher ANC, WBC, and SOFA measures in contrast to cluster ICU-mild.

Parameter	*ICU-mild (n = 57)*	*ICU-severe (n = 25)*	*p-value (adj.)*
RNA Quantity	108 ± 115.88 (57)	148.16 ± 84.24 (25)	0.0370
ED SOFA (first)	6.37 ± 4.55 (57)	9.56 ± 5.33 (25)	0.0370
ICU RNA Quantity	108 ± 115.88 (57)	148.16 ± 84.24 (25)	0.0370
ICU Outcome Survival 28 Day	85.96% (49/57)	96% (24/25)	0.0370
ICU SOFA	6.37 ± 4.55 (57)	9.56 ± 5.33 (25)	0.0370
ICU Outcome SOFA (first)	6.56 ± 4.27 (57)	9.96 ± 4.99 (25)	0.0220
ICU Apache	20.17 ± 7.3 (52)	27.21 ± 10.6 (24)	0.0166
ICU Outcome Survival 28 Day	24.91 ± 7.35 (56)	18.08 ± 10.6 (24)	0.00945
ICU Outcome Survival	96.49% (55/57)	100% (25/25)	0.00515
ICU Outcome Mortality	96.49% (55/57)	100% (25/25)	0.00515
ICU WBC	9.75 ± 7.09 (57)	16.12 ± 6.99 (25)	0.000200
ICU ANC	7.46 ± 4.85 (56)	14 ± 6.07 (25)	2.534e-05

Table 5–Variables related to ICU patients, which were tested on significance between clusters) using Wilcoxon rank sum test and Chi-Squared test for numeric and categorical (*p-value adjusted* 0.05). After that, a post-hoc analysis using the Benjamini-Hochberg method was conducted. For both clusters, the mean value, standard error, and total available values were included for numeric variables, with the share of available data in percentage.

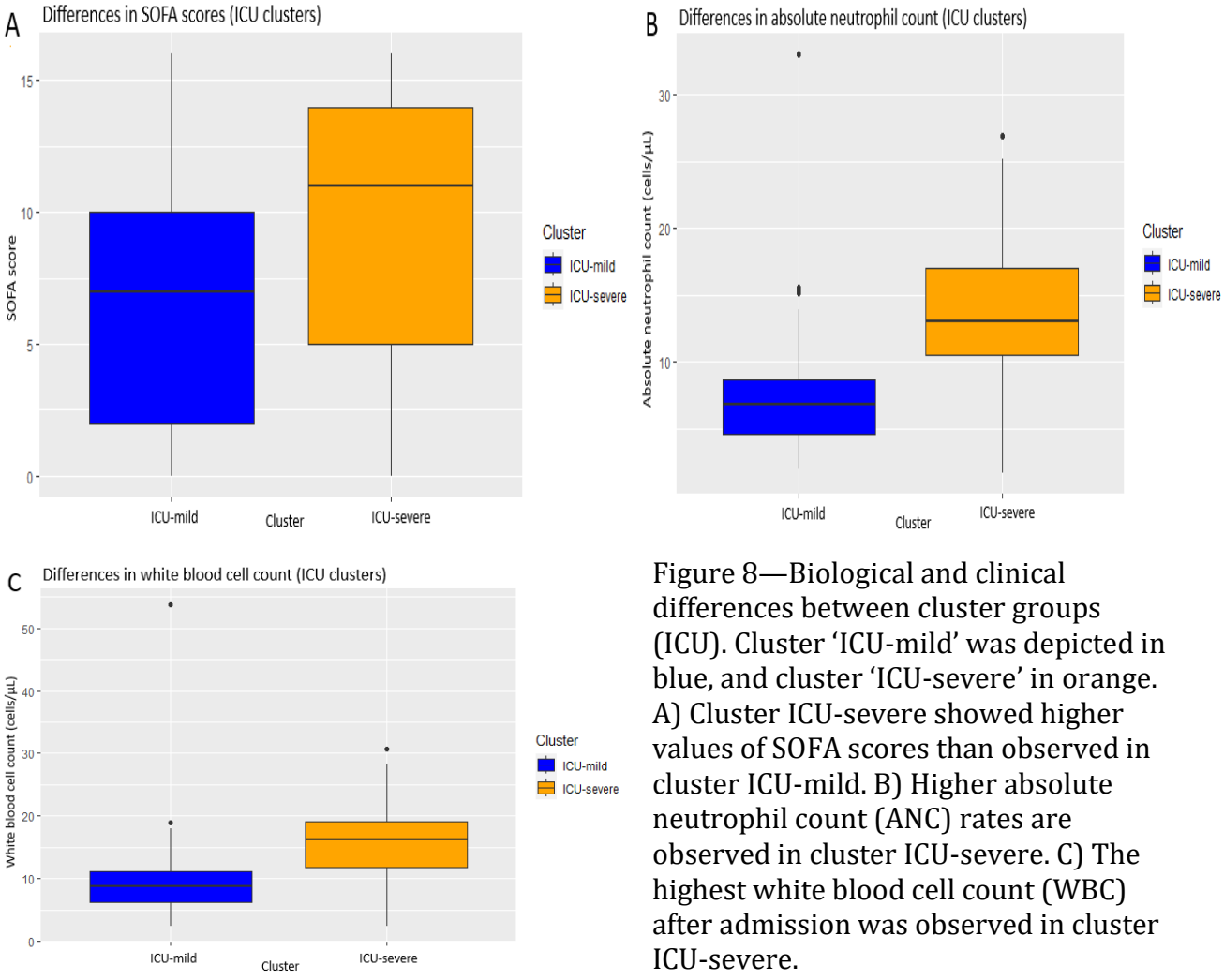


Figure 8—Biological and clinical differences between cluster groups (ICU). Cluster 'ICU-mild' was depicted in blue, and cluster 'ICU-severe' in orange. A) Cluster ICU-severe showed higher values of SOFA scores than observed in cluster ICU-mild. B) Higher absolute neutrophil count (ANC) rates are observed in cluster ICU-severe. C) The highest white blood cell count (WBC) after admission was observed in cluster ICU-severe.

Similar to the observations in Figure 7, we noticed a similar pattern in the ICU cohort: the ICU-severe cluster exhibited higher SOFA scores than the other (see Appendix II). Specifically, the ICU-severe cohort had an average SOFA score of 9,56, and the ICU-mild cohort had 6,53. Additionally, the pathways upregulated in the ER clusters were also observed, sometimes to a greater extent, in the ICU-mild and ICU-severe clusters compared to the healthy controls (see Appendix II).

Feature selection

The clusters proved to be distinct from each other both in severity and mitochondrial function deterioration (via variables such as lactate). For the development of the machine learning model, we exclusively used data from septic patients. When we used all 201 DEGs to predict a patient's cluster number and severity, we encountered overfitting in all five models (see Supplementary S4). Performing various feature selection methods resulted in three gene sets with diverse sizes. Feature selection was based on the High + Intermediate vs. Low group. The threshold for feature selection based on mutual information proved to be arbitrary, but as seen in Figure 9C, recall scores stayed consistent between the 9th and 27th highest-ranked genes. In addition, recall scores between the training and test sets were close enough to determine that overfitting was no longer a problem. Therefore, we evaluated this independent feature selection method on all used models for the nine highest-scoring genes. Feature selection by using coefficients from logistic regression, which demonstrated acceptable accuracy and recall scores based on RFECV, was used as model-based feature selection and as a good indicator of gene importance. A logistic regression model using standard parameter settings (L1 or LASSO regularization) found a gene set of eleven. Additionally, a logistic regression with L2 or ridge regularization concluded that five genes yielded the best recall scores (Figures 9A, B). Despite L2 not being a preferred method for feature selection due to the fact that coefficients cannot be set to zero, its findings still held value. An ElasticNet model was also employed to validate our findings with L1/L2 regularization, showing similar results (see Supplementary S4). We stopped at gene set sizes when recall scores began to dwindle, suggesting overfitting if we added more genes. Gene sets exhibited overlap, especially between five and eleven. However, the nine-gene set only shared one overlapping gene with the eleven-gene set; this was expected as the ranking originated from mutual information, suggesting inherent differences (Figure 9D).

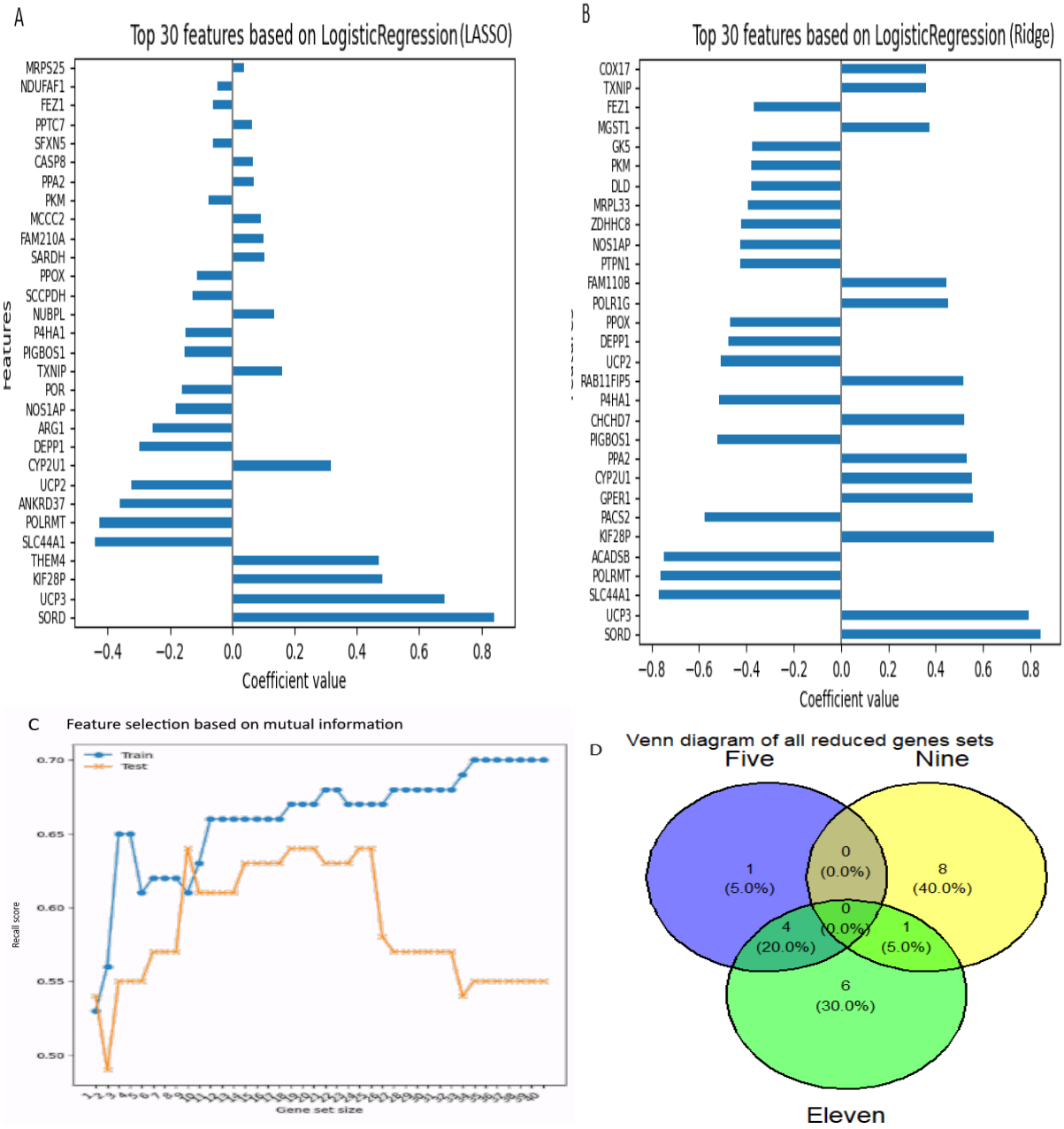


Figure 9—Different methods of feature selection. A) Logistic regression, using L1 regularization or 'LASSO', indicated through coefficients that eleven genes proved to be the best-performing gene set. B) Logistic regression using L2 regularization or "Ridge" and resulted in a gene set consisting of five genes. C) Feature selection through mutual information. The difference between the training (in blue) and test (in orange) scores stayed consistent between 0.60 and 0.65, based on recall scores between a gene set of 9 and 27. Therefore, we selected these nine genes and evaluated all five models for performance. D) Venn diagram depicting the overlap among each gene derived from the different selection methods.

Training and test of reduced gene set on ER cohort

In our analysis, the eleven-gene set demonstrated the best predictive capabilities (see Table 5). The other gene sets also exhibited good sensitivity and specificity (see supplementary S4 for a comparative analysis). We considered five models: random forest, SVM, naïve Bayes, and two forms of logistic regression, namely L1 and L2 regularization. After optimization on the reduced gene sets, the SVM and both logistic regression models showed comparable results (see Appendix III for used parameter settings). All models were capable of predicting the endotype group in the ER cohort with AUC and specificity above 0.95, while sensitivity was around 0,80 for the logistic regression models. Prediction scores for severity were lower across all models, with AUC always above 0,70. The sensitivity and specificity, however, varied. These scores, especially for endotype classification, suggested the potential for generalization in an ER setting. The L2 models had the highest AUC (0,97) but exhibited lower sensitivity compared to both forms of logistic regression. In contrast, SVM performed the best in predicting severity with an AUC of 0,72.

Comparison	*Model*	*AUC*	*Sensitivity*	*Specificity*
Endotype	Logistic regression (L1)	0,96	0,86	0,94
	Logistic regression (L2)	0,97	0,83	0,88
	SVM	0,96	0,83	0,88
High + Intermediate vs. Low	Logistic regression (L1)	0,70	0,72	0,69
	Logistic regression (L2)	0,72	0,71	0,63
	SVM	0,71	0,63	0,72

Table 5—Comparative performance of optimized models in endotype and severity classification. The performance metrics (AUC, sensitivity, and specificity) of three optimized models based on the eleven-gene set were applied to a test set comprising 25% of the ER population.

Validation on ICU cohort

We validated the three models on the ICU cohort to test their generalizability and to assess their effectiveness in a population with more severe sepsis cases. The results indicated that endotype classification remained consistent with the ER test set, with AUC and specificity staying above 0,92 and sensitivity between 0,75 and 0,80 for all models (see Table 6). However, predictive capabilities for severity were notably poorer compared to those in the ER cohort. While AUC and specificity remained comparable, sensitivity was considerably lower and ranged between 0.55 and 0.58 across the models. The L2 model had the highest sensitivity for predicting endotype (0,93) and specificity (0,79). In predicting severity, all models had similar AUCs, but L2 exhibited higher sensitivity. These findings suggested that although the models were effective in classifying endotypes, their capability to predict severity, particularly in high-severity cases, was less robust.

Comparison	*Model*	*AUC*	*Sensitivity*	*Specificity*
Endotype	Logistic regression (L1)	0,93	0,92	0,78
	Logistic regression (L2)	0,93	0,93	0,79
	SVM	0,93	0,92	0,78
High + Intermediate vs. Low	Logistic regression (L1)	0,67	0,57	0,62
	Logistic regression (L2)	0,66	0,58	0,69
	SVM	0,66	0,55	0,61

Table 6—Comparative performance of optimized models in endotype and severity classification. The performance metrics (AUC, sensitivity, and specificity) of three optimized models based on the eleven-gene set are shown. Results are based on the ICU cohort, serving as a validation set.

Final gene set: a zoom-in

After establishing the three reduced gene sets through feature selection and evaluating their prediction capabilities, it was crucial to observe how these genes were expressed across clusters and severity levels. Figure 10A-D depicted the gene expressions from the eleven-gene set for both ER and ICU cohorts. Here we observed that some genes are expressed somewhat differently for the ER cohorts (Figure 10A, B). For instance, *ARG1* was more upregulated for ER-severe compared to ER-mild. The opposite can be said for *THEM4* and *UCP3*; these two genes are more downregulated when comparing the two clusters (Figure 10A). Other genes such as *KIF28P* did not exhibit significant changes in expression. Genes that were upregulated tended to show higher expression in cases of greater severity; for example, ARG1 was more upregulated in High severity cases compared to Low severity ones (Figure 10B). And the reverse was also true. The onset of higher severity in the ICU cohort made differences in gene expressions more pronounced. For example, the differences in ARG1 expression between ICU-severe and ICU-mild were higher than in the ER cohort (Figure 10C). This is true for every gene. For instance, no significant differences in the expression of PPTC7 were observed in the ER cohort, but these differences became apparent in the ICU cohort. The observations made for this gene set were consistent with other gene sets included in the study. Whenever a gene was upregulated or downregulated between clusters in the ER, these differences were more pronounced in the ICU cohort for most cases.

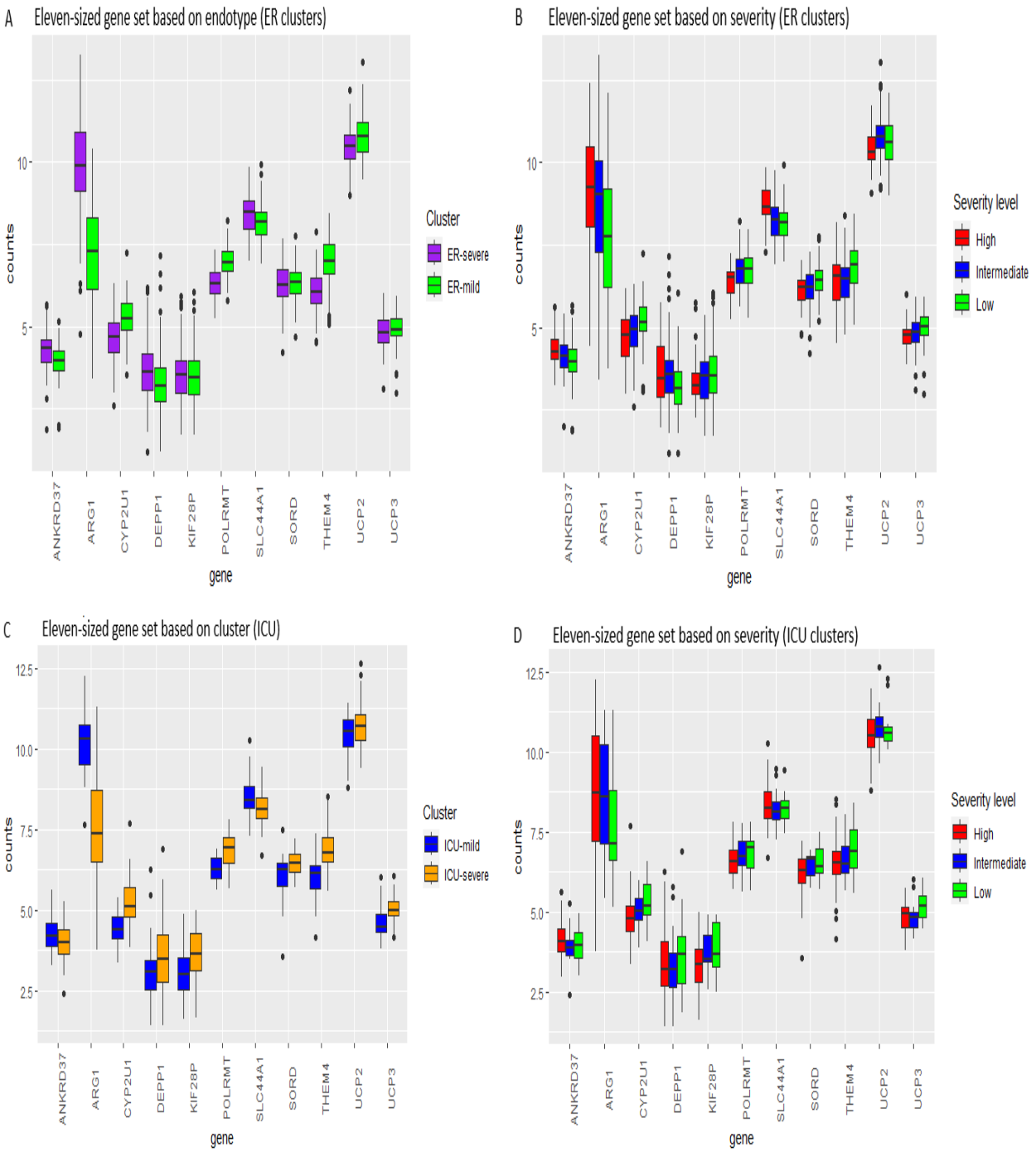


Figure 11—Gene expression patterns of the eleven-gene set across ER and ICU cohorts, with counts normalized using VST. A) Gene expression in the ER cohort is shown per cluster, with ER-severe depicted in green and ER-mild in purple. B) Gene expression in the ER cohort is presented per severity level. C) Gene expression in the ICU cohort is displayed per cluster, with ICU-mild in orange and ICU-severe in blue. D) Gene expression in the ICU cohort is shown per severity level.

Discussion

In this study, we focused on a subset of mitochondrial-related genes to identify potential biomarkers for stratification and aim to build hypotheses for a broader follow-up study. The two established clusters demonstrated distinct gene expression patterns in various biological pathways, particularly those associated with mitochondrial function. Some pathways were altered due to mitochondrial dysfunction induced by oxidative stress. Such as apoptosis and respiratory electron transport. These are commonly observed in septic patients. [13, 46] Although we differentiated between ER and ICU, similar patterns were evident in both settings. Notably, ER-severe and ICU-severe exhibited higher lactate levels, increased white blood cell counts, higher average SOFA scores, longer admission durations, and had greater proportion of High and Intermediate severity cases. The role of neutrophils in sepsis may be a potential therapeutic target and was significantly different between clusters. [47] These findings suggest biological and mechanistic distinctions between the clusters, which lead us to term them as mitochondria-driven endotypes.

However, the presence of missing values in several significant variables resulted in a limited sample size. For instance, crucial lactate measurements were available for only 48 out of 266 cases. [15] This limitation might have caused a misrepresentation of the broader population and underscored the need for careful interpretation of our findings.

We opted for an FC cutoff value of 1,2 to assess changes in mitochondrial gene expression. This decision differs from the more standard practice in similar studies, which often use a cutoff of 1,5 or even 2. [48] This decision is reasonable considering that changes in mitochondrial gene expression due to oxidative stress can be quite noticeable, as highlighted in previous research. [49] However, this way of loose selection criteria was a tradeoff between including DEGs and adding noise within the data and could have affected the stability of clustering. Also peculiar is the difference in DEG identification between DESeq2 and edgeR. edgeR identified 183 and DESeq2 201 DEGs, with an overlap of 180. This difference meant that 10,30% of the total DEGs were only identified by DESeq2, a relatively large proportion. One possible explanation for this difference could be the variation in statistical methods used by both tools. Nonetheless, we decided to proceed with all DEGs identified by DESeq2. While this decision might have introduced some variability, we did not expect it to alter our results in a drastic manner since most unique DEGs to DESeq2 just fell below the selection criteria used with edgeR (see Supplemental S2).

The highest average silhouette score we observed in clustering on all DEGs was 0,18, which was achieved with PAM using Manhattan distance on the ER set. While not exceptionally high, it reflected the known challenges when clustering based on gene expression data derived from a heterogeneous disease like sepsis. [50, 59] Additionally, when applying different methods to improve and validate findings, we observed that the results were not drastically different. These measures included performing clustering on differently sized gene sets via MAD and other clustering methods, validating on the ICU cohort, dimension reduction via PCA, and including all mitochondria-related genes.

Interestingly, most samples with low silhouette scores were from the Low and Intermediate severity subgroups, suggesting that severe septic cases were more stably stratified.

Moreover, the samples indicated from the Low subclass had higher SOFA scores, being identified as Intermediate. These findings suggest a degree of discrimination between the two clusters. However, the mentioned shortcomings may limit the scope of septic patients being accurately stratified into one of the established endotypes.

Another limitation of our study is the way we identified DEGs. We performed DE analysis before clustering on the basis of establishing severity-based gene signatures. The assumption was to extract genes showing the most significant expression changes. This result was also performed on the whole transcriptome, resulting in most mitochondria-related DEGs being recognized in a DEG set of over 2.000 genes (see Supplementary 2). While this approach may have identified the most relevant genes. We recognize that this approach may have introduced selection bias by potentially overlooking genes that did not meet threshold criteria. These genes might have contributed to more stable clusters and may have contributed to the onset of mitochondria dysfunction since mitochondria are tightly regulated. Therefore, while our approach based clusters on more significantly altered genes, we may have "missed" other biologically relevant genes. This selection bias has the potential to reduce the generalization of our results, and interpretations may need to be reassessed in a validation study. Therefore, we cannot foresee the implementation of these endotypes as a prognostic or diagnostic tool in their current form.

We have optimized three models that are capable of accurately predicting the endotype a septic patient belonged to, regardless of the size of the gene set. These outcomes were anticipated since the clusters are based on the expression patterns of the DEGs. However, feature selection did not yield any optimal subset of mitochondria-related genes that could effectively differentiate between the extreme phenotypes (High + Intermediate vs. Low) in the same manner. This is the challenge regarding trying to model the heterogeneous nature of sepsis. The large set of 201 DEGs made feature selection difficult; evaluating all possible gene combinations with different sizes is not only computably intensive but also has the potential for establishing gene set based redundant genes. This complicated identification of truly significant genes. For example, we observed that RFECV continuously switched the optimal number of genes, indicating the predictive capability of many DEG combinations. Therefore, the main objective was to identify the fewest gene set with the greatest accuracy score, not necessarily the best-performing gene set. For this reason, we employed several feature selection methods to identify overlaps, which led to the identification of three gene sets of varying sizes. The set comprising eleven genes out of the three subsets achieved the highest AUC scores in endotype classification, resulting in AUCs of 0,96 and 0,93 for ER and ICU cohorts. In clinical studies, AUC scores between 0,90 and 1,00 are considered excellent, indicating a high level of accuracy in distinguishing between two groups. [51]

However, there are some limitations. Even though the severe endotypes represent more severe sepsis cases, classification of severity remained challenging. When we validated our models on the ICU cohort, we saw that a large number of the combined class of High and Intermediate severity was misclassified as Low. This observation raised concerns about the all models' generalizability in accurately classifying patients with moderately high SOFA scores, especially in an ICU setting. This distinction is crucial because patients with moderately high SOFA scores still require significant medical attention. The difference in

performance between the ER and ICU validation cohorts could be attributed to the fact that the ER cohort had a significant class imbalance; only 32 out of 266 patients were considered to have High severity. This might mean that the model has been trained on distinction moderately high and lower SOFA scores, as they are presented in the ER. This leads to the fact that these models might be better suited for endotype classification in an ER setting. Nonetheless, these results show that mitochondria-related genes play an essential role in further deterioration. Therefore, further research to establish clinical diagnostic tools based on mitochondrial gene expression might be valuable to consider.

While we did not necessarily find an optimal model for endotype classification, the optimized models we delivered have shown excellent predictive capabilities. However, there are some considerations to be weighed in naming a model that can be used for further research. SVM has the advantage of being great at handling non-linear relationships and managing diverse and complex datasets. In contrast, its results may be difficult to interpret. On the other hand, L2 is great at capturing collinearity, but it operates on the assumption of linearity (our data is nonlinear in nature). The same is evident for L1; however, it has the advantage of model simplicity and is scalable on large datasets. It also has the advantage of being the model that selected the gene set. All models were effective with high-dimensional data and are widely used in identifying biomarkers with heterogeneous diseases such as sepsis. Due to its simplicity, widespread usage in medical research, adaptability to handling complex data, and great accuracy in endotype classification, we considered logistic regression with L1 as our final model. [61] These qualities make L1 a great candidate for further research on this dataset.

Our results were similar or better in predicting severity compared to other, more traditional biomarkers. For instance, one exploratory study tried to establish 17 different cytokines as biomarkers. [52] IL-6 and IL-8 had good-scoring AUCs of 0,685 and 0,7000, respectively. Another study looked at procalcitonin, IL-6, and IL-8. These established AUCs of 0,92, 0,75, and 0,71, respectively. [54] However, omics data tends to be better at capturing sepsis heterogeneity and has the potential to be a better decision-making tool. [54] In addition, omics can pave the way for a more personalized approach to medicine.

Our study focused on a subset of genes to discover mitochondria-related biomarkers and even though we validated the found DEGs with the entire transcriptome, we did not account for cell proportions. This comes with some potential limitations. For instance, overlooking cellular heterogeneity since sepsis can affect cell types differently due to differences in the epigenetics of mitochondria, especially in whole-blood genetics. [55, 7] One such method with promising capabilities is CIBERSORT, a widely used deconvolution tool to estimate cell composition based on RNA-Seq data. Its practices have been attributed to the discovery of many new therapeutics. [56] Future research should extend the scope established in this study by incorporating cell proportions and expand upon our findings. This study laid the groundwork for such an approach while highlighting the importance of mitochondrial dysfunction in discriminating septic patients based on severity.

Only a few studies have been conducted that have tried to establish specific mitochondrial-driven endotypes. One study tried identifying mitochondria-related biomarkers and determined that two clusters were also the most appropriate number of clusters. [49]

However, ER and ICU were not distinguished since it was a case-control study. The study's authors identified three biomarkers, namely *BCKDHB*, *NDUFB3*, and *LETMD1*, none of which we found in the best-performing gene set. The ROC curve for validation had an AUC score of 0,768, whereas we scored 0,78 on the ICU validation cohort. Another study tried to establish gene signatures for sepsis for prognostic via single-cell sequencing and found five mitochondrial-related genes, which were significantly up-or downregulated. After that, the authors managed to identify two distinct cluster groups, each of which had unique expression levels of immune cells. These five genetic biomarkers were *NMR5*, *MAOA*, *COX7b*, *PPM1K*, and *TTC19*. [57] Predicting capabilities vary since the authors constructed an over-time prognosis tool for 7-, 14-, and 28-day periods. Nevertheless, most ROC curves had an AUC score between 0,65 and 0,75. Again, no gene overlaps with our DEG set. The existence of at least two endotypes or subgroups in sepsis based on gene expression patterns of mitochondria is in line with other studies. Also, prediction capabilities were comparable.

Reasons for these observed differences could be the implementation and goals of the named studies. First, our study focused on establishing mitochondria-related biomarkers based on an already established severity measurement, whereas the others did not. These studies conducted unsupervised clustering and extracted DEGs after that. Second, both studies extracted DEGs by comparing healthy controls with septic patients; we compared various levels of sepsis severity. Lastly, both studies used stricter thresholds regarding fold changes (FC of | 2 |), whereas we were more interested in detecting minor gene expression changes. This resulted in 201 DEGs and three and five genes for the other two studies, respectively. These differences underline each study's goals; these authors wanted to establish endotypes and explore biomarkers to distinguish between healthy and septic patients. Our goals were more aligned to explore biomarkers to stratify patients into different severity-based endotypes. Nonetheless, it remains clear that there are at least two distinct endotypes.
Sepsis is a complex condition, with mitochondrial dysfunction only representing one aspect of this complex clinical syndrome. Other heterogeneous characteristics of sepsis come from immune responses, distinct levels of inflammation, divergent speeds of deterioration and organ dysfunction. They should be considered for an accurate clinical implementation, which our model lacks [50] Therefore, focusing solely on RNA-Seq data in a clinical setting would not sufficiently address all challenges of the heterogeneous nature of sepsis. [59, 50] Employing multimodal methods could enhance the assessment of septic patients by integrating clinical data relatable to mitochondria dysfunction like lactate levels, other types of omics data, and cardiolipin levels. [58] Other studies followed a more immune-related gene approach and found, in some cases, better prediction scores than solely focusing on mitochondria-related biomarkers. The study conducted by Baquir and researchers is an ideal example; based on a 40-gene set, the prediction capabilities are better than our model. Predicting sepsis severity signatures achieved an AUC score of 0,75 compared to 0,74 on validation cohorts. [9] Thus, a multimodal prediction model that includes mitochondrial-related genes, clinical attributes, and immune-related factors could be a valuable next step in refining our model for predictive purposes.

Conclusion and future work

In conclusion, this project aimed to establish severity-driven endotypes based on the gene expressions of mitochondria-related genes. We can conclude that this goal has been achieved. We developed an optimized logistic regression with L1 regularization model, which was capable of predicting the extreme phenotypes of SOFA-based sepsis severity with acceptable scores. Additionally, it accurately predicted the established mitochondria-driven endotype cluster by using a novel eleven-sized gene set. However, several ongoing challenges remain. The heterogeneity of sepsis, the imbalance in High severity classes, and the difficulties in predicting outcomes based on mitochondria-related genes all suggest to interpret our findings with some caution. These factors indicate that utilizing these findings to assist in clinical decision-making is still premature. Further research is required to address these complexities effectively.

Future work should focus on understanding how mitochondria-related genes function in the entire cell population. This could involve a validation study or an extension of current research. Efforts in model-building could benefit from integrating related to mitochondria dysfunction. Additionally, incorporating other nonmitochondrial-related biomarkers could address the heterogeneity of sepsis more effectively. Such an approach would likely improve the prediction capabilities of early sepsis recognition which contributes to the advancement of personalized medicine.

References

[1] M. Singer *et al.*, "The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801–810, Feb. 2016, doi: 10.1001/jama.2016.0287.

[2] R. Luhr, Y. Cao, B. Söderquist, and S. Cajander, "Trends in sepsis mortality over time in randomized sepsis trials: A systematic literature review and meta-analysis of mortality in the control arm, 2002–2016," *Critical Care*, vol. 23, no. 1, p. 241, 2019, doi: 10.1186/s13054-019-2528-0.

[3] E. K. Stevenson, A. R. Rubenstein, G. T. Radin, R. S. Wiener, and A. J. Walkey, "Two decades of mortality trends among patients with severe sepsis: A comparative meta-analysis*," *Critical Care Medicine*, vol. 42, no. 3, pp. 625–631, 2014, doi: 10.1097/CCM.0000000000000026.

[4] K. E. Rudd *et al.*, "Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the global burden of disease study," *The Lancet*, vol. 395, no. 10219, pp. 200–211, Jan. 2020, doi: 10.1016/s0140-6736(19)32989-7.

[5] J. Hajj, N. Blaine, J. Salavaci, and D. Jacoby, "The "centrality of sepsis": A review on incidence, mortality, and cost of care," *Healthcare*, vol. 6, no. 3, p. 90, Jul. 2018, doi: https://doi.org/10.3390/healthcare6030090.

[6] C. J. Paoli, "Epidemiology and costs of sepsis in the united states-an analysis based on timing of diagnosis and severity level," vol. 46, no. 12, pp. 1889–1897, Jan. 2018, doi: 10.1097/CCM.0000000000003342.

[7] A. Leligdowicz and M. A. Matthay, "Heterogeneity in sepsis: New biological evidence with clinical applications," *Critical Care*, vol. 23, no. 1, Mar. 2019, doi: 10.1186/s13054-019-2372-2.[

8] T. E. Sweeney *et al.*, "Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters," *Critical Care Medicine*, vol. 46, no. 6, pp. 915–925, Jun. 2018, doi: 10.1097/ccm.0000000000003084.

[9] A. Baghela *et al.*, "Predicting sepsis severity at first clinical presentation: The role of endotypes and mechanistic signatures," *eBioMedicine*, vol. 75, p. 103776, Jan. 2022, doi: 10.1016/j.ebiom.2021.103776.

[10] S. Lambden, P. F. Laterre, M. M. Levy, and B. Francois, "The SOFA score—development, utility and challenges of accurate assessment in clinical trials," *Critical Care*, vol. 23, no. 1, Nov. 2019, doi: 10.1186/s13054-019-2663-7.

[11] S. Maitra, A. Som, and S. Bhattacharjee, "Accuracy of quick sequential organ failure assessment (qSOFA) score and systemic inflammatory response syndrome (SIRS) criteria for predicting mortality in hospitalized patients with suspected infection: A meta-analysis of observational studies," *Clinical Microbiology and Infection*, vol. 24, no. 11, pp. 1123–1129, Nov. 2018, doi: 10.1016/j.cmi.2018.03.032.

[12] M. P. Murphy, "How mitochondria produce reactive oxygen species," *Biochemical Journal*, vol. 417, no. 1, pp. 1–13, Dec. 2008, doi: 10.1042/bj20081386.

[13] M. Ott, V. Gogvadze, S. Orrenius, and B. Zhivotovsky, "Mitochondria, oxidative stress and cell death," *Apoptosis*, vol. 12, no. 5, pp. 913–922, Feb. 2007, doi: 10.1007/s10495-007-0756-2.

[14] K. Nakahira, S. Hisata, and A. M. K. Choi, "The roles of mitochondrial damage-associated molecular patterns in diseases," *Antioxidants & Redox Signaling*, vol. 23, no.

17, pp. 1329–1350, Dec. 2015, doi: 10.1089/ars.2015.6407.
[15] M. Singer, "The role of mitochondrial dysfunction in sepsis-induced multi-organ failure," *Virulence*, vol. 5, no. 1, pp. 66–72, Nov. 2013, doi: 10.4161/viru.26907.
[16] B. S. Star, E. C. van der Slikke, A. van Buiten, R. H. Henning, and H. R. Bouma, "The novel compound SUL-138 counteracts endothelial cell and kidney dysfunction in sepsis by preserving mitochondrial function," *International Journal of Molecular Sciences*, vol. 24, no. 7, p. 6330, Mar. 2023, doi: 10.3390/ijms24076330.
[17] S. Andrews, "FastQC: A quality control tool for high throughput sequence data." http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2010.
[18] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, "MultiQC: Summarize analysis results for multiple tools and samples in a single report," *Bioinformatics*, vol. 32, no. 19, pp. 3047–3048, Jun. 2016, doi: 10.1093/bioinformatics/btw354.
[19] A. Dobin and T. R. Gingeras, "Mapping RNA-seq reads with STAR," *Current Protocols in Bioinformatics*, vol. 51, no. 1, Sep. 2015, doi: 10.1002/0471250953.bi1114s51.
[20] S. Anders, P. T. Pyl, and W. Huber, "HTSeq—a python framework to work with high-throughput sequencing data," *Bioinformatics*, vol. 31, no. 2, pp. 166–169, Sep. 2014, doi: 10.1093/bioinformatics/btu638.
[21] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2023. Available: https://www.R-project.org
[22] H. Wickham, R. François, L. Henry, K. Müller, and D. Vaughan, *Dplyr: A grammar of data manipulation*. 2023. Available: https://dplyr.tidyverse.org
[23] H. Wickham *et al.*, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, Nov. 2019, doi: 10.21105/joss.01686.
[24] M. Kuhn, "Building predictive models inRUsing thecaretPackage," *Journal of Statistical Software*, vol. 28, no. 5, 2008, doi: 10.18637/jss.v028.i05.
[25] H. Wickham, *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York, 2016. Available: https://ggplot2.tidyverse.org
[26] D. Scheper, "Unraveling the role of mitochondria dysfunction in early sepsis." https://bitbucket.org/jscscheper/internship/src/main/ , 2023.
[27] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology*, vol. 15, no. 12, Dec. 2014, doi: 10.1186/s13059-014-0550-8.
[28] Y. Chen, A. T. L. Lun, and G. K. Smyth, "From reads to genes to pathways: Differential expression analysis of RNA-seq experiments using rsubread and the edgeR quasi-likelihood pipeline," *F1000Research*, vol. 5, p. 1438, Aug. 2016, doi: 10.12688/f1000research.8987.2.
[29] G. Yu and Q.-Y. He, "ReactomePA: An r/bioconductor package for reactome pathway analysis and visualization," *Molecular BioSystems*, vol. 12, no. 2, pp. 477–479, 2016, doi: 10.1039/c5mb00663e.
[30] M. V. Kuleshov *et al.*, "Enrichr: A comprehensive gene set enrichment analysis web server 2016 update," *Nucleic Acids Research*, vol. 44, no. W1, pp. W90–W97, May 2016, doi: 10.1093/nar/gkw377.
[31] Jeffrey T. Leek , W. Evan Johnson>, Hilary S. Parker *et all.*, "Sva." Bioconductor, 2017. doi: 10.18129/B9.BIOC.SVA.
[32] T. Galili, "Dendextend: An r package for visualizing, adjusting and comparing trees of hierarchical clustering," *Bioinformatics*, vol. 31, no. 22, pp. 3718–3720, Jul. 2015, doi:

10.1093/bioinformatics/btv428.
[33] D. M. Wilkerson and N. D. Hayes, "ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking," *Bioinformatics*, vol. 26, no. 12, pp. 1572–1573, 2010.
[34] A. Kassambara and F. Mundt, *Factoextra: Extract and visualize the results of multivariate data analyses*. 2020. Available: https://CRAN.R-project.org/package=factoextra
[35] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, "NbClust: An r package for determining the relevant number of clusters in a data set," *Journal of Statistical Software*, vol. 61, no. 6, pp. 1–36, 2014, Available: https://www.jstatsoft.org/v61/i06/
[36] Z. Gu, R. Eils, and M. Schlesner, "Complex heatmaps reveal patterns and correlations in multidimensional genomic data," *Bioinformatics*, vol. 32, no. 18, pp. 2847–2849, May 2016, doi: 10.1093/bioinformatics/btw313.
[37] Python Software Foundation, "Python (version 3.11.12)." https://www.python.org/, 2023.
[38] T. Kluyver *et al.*, "Jupyter notebooks – a publishing format for reproducible computational workflows." IOS Press, pp. 87–90, 2016.
[39] W. McKinney *et al.*, "Data structures for statistical computing in python," in *Proceedings of the 9th python in science conference*, Austin, TX, 2010, pp. 51–56.
[40] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, doi: 10.1038/s41586-020-2649-2.
[41] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
[42] M. Waskom *et al.*, "Mwaskom/seaborn: v0.8.1 (september 2017)." Zenodo, Sep. 2017. doi: 10.5281/zenodo.883859.
[43] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/mcse.2007.55.
[44] Reiichiro Nakano, "Reiinakano/scikit-plot: v0.2.1." Zenodo, 2017. doi: 10.5281/ZENODO.293191.
[45] S. Raschka, "MLxtend: Providing machine learning and data science utilities and extensions to python's scientific computing stack," *Journal of Open Source Software*, vol. 3, no. 24, p. 638, Apr. 2018, doi: 10.21105/joss.00638.
[46] H. Cui, Y. Kong, and H. Zhang, "Oxidative stress, mitochondrial dysfunction, and aging," *Journal of Signal Transduction*, vol. 2012, pp. 1–13, Oct. 2012, doi: 10.1155/2012/646354.
[47] X. Shen, K. Cao, Y. Zhao, and J. Du, "Targeting neutrophils in sepsis: From mechanism to translation," *Frontiers in Pharmacology*, vol. 12, Apr. 2021, doi: 10.3389/fphar.2021.644270.
[48] C. Ploumi, I. Daskalaki, and N. Tavernarakis, "Mitochondrial biogenesis and clearance: A balancing act," *The FEBS Journal*, vol. 284, no. 2, pp. 183–195, Aug. 2016, doi: 10.1111/febs.13820.
[49] Q. Shu, H. She, X. Chen, L. Zhong, J. Zhu, and L. Fang, "Identification and experimental validation of mitochondria-related genes biomarkers associated with immune infiltration for sepsis," *Frontiers in Immunology*, vol. 14, May 2023, doi: 10.3389/fimmu.2023.1184126.
[50] A. A. Woodward, R. J. Urbanowicz, A. C. Naj, and J. H. Moore, "Genetic heterogeneity: Challenges, impacts, and methods through an associative lens," *Genetic Epidemiology*, vol.

46, no. 8, pp. 555–571, Aug. 2022, doi: 10.1002/gepi.22497.
[51] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, Oct. 1978, doi: 10.1016/s0001-2998(78)80014-2.
[52] F. A. Bozza *et al.*, "Cytokine profiles as markers of disease severity in sepsis: A multiplex analysis," *Critical Care*, vol. 11, no. 2, p. R49, 2007, doi: 10.1186/cc5783.
[53] S. HARBARTH *et al.*, "Diagnostic value of procalcitonin, interleukin-6, and interleukin-8 in critically ill patients admitted with suspected sepsis," *American Journal of Respiratory and Critical Care Medicine*, vol. 164, no. 3, pp. 396–402, Aug. 2001, doi: 10.1164/ajrccm.164.3.2009052.
[54] T. Itenov, D. Murray, and J. Jensen, "Sepsis: Personalized medicine utilizing 'omic' technologies—a paradigm shift?" *Healthcare*, vol. 6, no. 3, p. 111, Sep. 2018, doi: 10.3390/healthcare6030111.
[55] A. E. Jaffe and R. A. Irizarry, "Accounting for cellular heterogeneity is critical in epigenome-wide association studies," *Genome Biology*, vol. 15, no. 2, p. R31, 2014, doi: 10.1186/gb-2014-15-2-r31.
[56] A. M. Newman *et al.*, "Robust enumeration of cell subsets from tissue expression profiles," *Nature Methods*, vol. 12, no. 5, pp. 453–457, Mar. 2015, doi: 10.1038/nmeth.3337.
[57] Q. Shu *et al.*, "Construction and validation of a mitochondria-associated genes prognostic signature and immune microenvironment characteristic of sepsis," *International Immunopharmacology*, vol. 126, p. 111275, Jan. 2024, doi: 10.1016/j.intimp.2023.111275.
[58] J. N. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol, "Multimodal biomedical AI," *Nature Medicine*, vol. 28, no. 9, pp. 1773–1784, Sep. 2022, doi: 10.1038/s41591-022-01981-2.
[59] Tsimberidou, A-M., "Initiative for Molecular Profiling and Advanced Cancer Therapy and challenges in implementation of precision medicine," *Elsevier BV*, vol. 41, p. 176-181, May 2017, doi: 10.1016/j.currproblcancer.2017.02.002.
[60] W. A. Knaus et al., "The APACHE III Prognostic System," *Chest*, vol. 100, no. 6, pp. 1619–1636, Dec. 1991, doi: 10.1378/chest.100.6.1619
[61] Lu, Y, Zhao, Q., "Identification of potential diagnostic and prognostic biomarkers for sepsis based on machine learning," *Computational and Structural Biotechnology Journal*, vol. 21, p. 2316-2331, 2023, doi: 10.1016/j.csbj.2023.03.034.

Appendix I: Mito-genes (clustering)

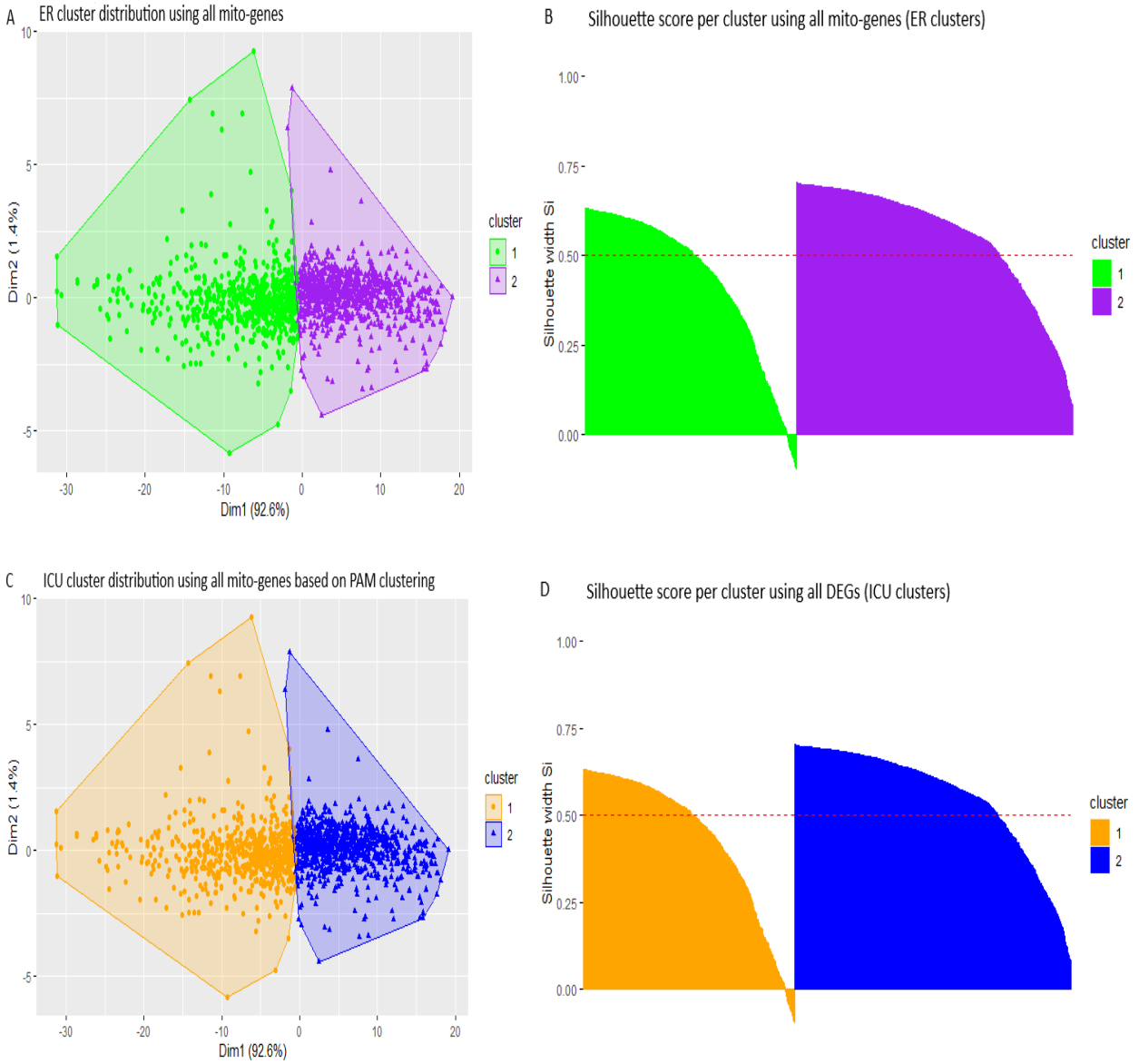


Figure 11—Clustering results regarding all the mitochondria-related genes after removing low-expression genes. A) ER cluster plot displayed two distinct clusters. B) ER cluster silhouette width per cluster; clusters show high stability. C) ICU cluster plot for two distinct clusters. D) ICU cluster silhouette width per cluster; clusters show high stability.

Appendix II: Differences in ICU cohort

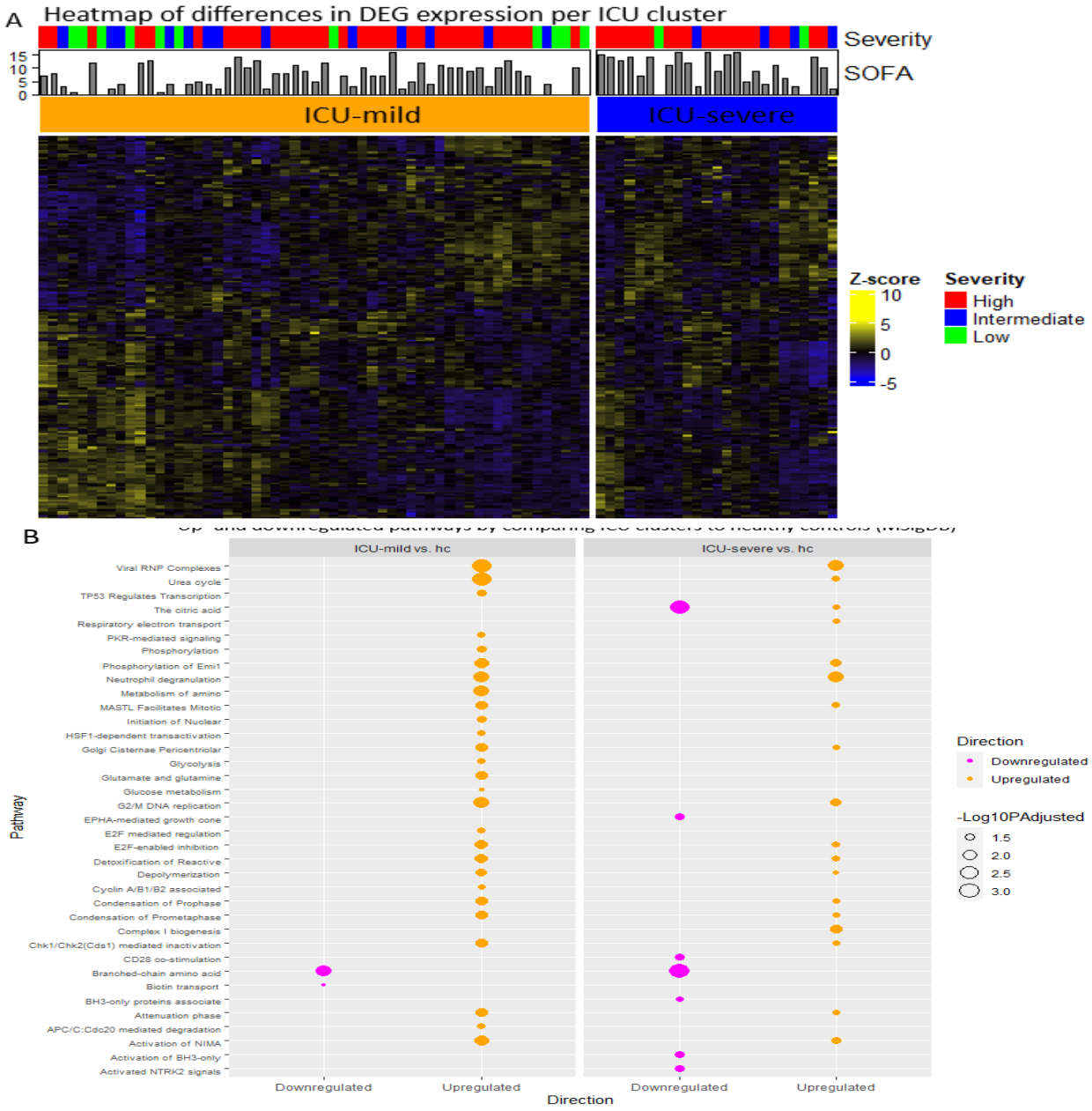


Figure 12—Differences between cluster groups for ICU. A) Heatmap of the two ICU-based clusters with gene expression patterns from all 201 DEGs. The ICU-severe cluster exhibited higher amounts of High severity compared to the ICU-mild cluster. In addition, the highest SOFA scores measured within in the first 72 hours after admission were higher in ICU-severe cluster. B) Pathway analysis was conducted on the cluster groups in comparison with a healthy control population of 44 patients. Only pathways with a p-value of less than 0.05 were considered significant. These were related to apoptosis and TP53. In both clusters, various upregulated pathways were observed like neutrophil degranulation.

Appendix III: Parameter settings of optimized models

Parameter	*SVM*	*Logistic Regression (L1)*	*Logistic Regression (L2)*
C	100	20	1
gamma	0,01	-	-
kernel	*rbf*	-	-
penalty	-	L1	L2
solver	-	*saga*	*saga*

Table 7—optimized hyperparameter settings for three models. A high *C* value was chosen to balance the fit to the training data and avoiding overfitting. Additionally, a low *gamma* was selected, which controls the trade-off between maximizing the margin of separation between classes and minimizing classification errors. These settings showed that models did not require overly strict decision boundaries.

Appendix IV: Dutch translation of abstract

Sepsis is een grote bijdrager aan sterfte in ziekenhuizen. Dit komt voornamelijk doordat sepsis zich op vele manieren kan voordoen. Mitochondriën die beschadigd zijn geraakt door hevige immuunreacties, kunnen sepsis verergeren. Het analyzeren van genen gerelateerd aan mitochondriën kan een vroegtijdige indicatie geven over de gesteldheid van een patiënt en helpen met het opstellen van een behandelplan.

De genprofielen van 348 patiënten met sepsis—afkomstig van vier spoedafdelingen en één intensivecareafdeling–zijn gebruikt en vergeleken met 44 gezonde individuën. Klinisch relevante genprofielen zijn totstandgekomen door het toepassen van verschillende vormen van kunstmatige intelligentie.

Genen die anders tot expressie komen (*DEGs*), zijn geïdentificeerd doormiddel van het vergelijken van groepen patiënten met varierende ernst van sepsis. Deze ernst wordt bepaald door een zogeheten SOFA-score. Er zijn twee unieke groepen totstandgekomen door het gebruik van de *DEGs* op basis van patiënten afkomstig uit de vier spoedafdelingen. Deze groepen worden ook wel endotypen genoemd. De twee endotypen zijn vergeleken met de groep gezonde individuën en gevalideerd op de patiënten van de intensivecare. Doormiddel van selectie op relevante genen, zijn er drie verschillende sets van genen ontdekt die op een accurate wijze de endotypegroep en ernst van sepsis kunnen voorspellen. Een logistische regressie algoritme met L1 regularisatie voorspelde endotypegroepen met een accuratie van 96% en 93% in ER- en ICU-cohorten op basis van elf genexpressies.

De genprofielen en endotypen die zijn ontdekt, geven de essentiele rol van mitochondriën in het tijdig herkennen van sepsis aan. Toekomstig onderzoek moet gericht zijn op het ontwikkelen van een multimodaal prognostische tool voor een verbetering in stratificatie van patiënten.