

Population scale proteomics enables adaptive digital twin modelling in sepsis

Aaron M. Scott^{1,**}, Lisa Mellhammar^{1,4,**}, Erik Malmström^{1,3,**},
Axel Goch Gustafsson¹, Anahita Bakochi², Marc Isaksson²,
Tirthankar Mohanty¹, Louise Thelaus¹, Fredrik Kahn^{1,4},
Lars Malmström¹, Johan Malmström^{1,2,*}, Adam Linder^{1,4,*}

¹Division of Infection Medicine, Department of Clinical Sciences Lund, Faculty of Medicine, Lund University, Lund, Sweden

²Swedish National Infrastructure for Biological Mass Spectrometry (BioMS), Lund University, Lund, Sweden

³Emergency Medicine, Department of Clinical Sciences, Skåne University Hospital, Lund University, Lund, Sweden

⁴Department of Infectious Diseases, Skåne University Hospital, Lund, Sweden

* Corresponding authors: Adam Linder (adam.linder@med.lu.se), Johan Malmström (johan.malmstrom@med.lu.se) for proteomics and mass spectrometry related aspects of the study.

** Aaron M. Scott, Lisa Mellhammar, and Erik Malmström contributed equally to the manuscript.

Keywords: sepsis, proteomics, explainable artificial intelligence, machine learning, digital twin analysis, subphenotype identification, biomarker identification

Abstract

Sepsis is one of the leading causes of mortality in the world. Currently, the heterogeneity of sepsis makes it challenging to determine the molecular mechanisms that define the syndrome. Here, we leverage population scale proteomics to analyze a well-defined cohort of 1364 blood samples taken at time-of-admission to the emergency department from patients suspected of sepsis. We identified panels of proteins using explainable artificial intelligence that predict clinical outcomes and applied these panels to reduce high-dimensional proteomics data to a low-dimensional interpretable latent space (ILS). Using the ILS, we constructed an adaptive digital twin model that accurately predicted organ dysfunction, mortality, and early-mortality-risk patients using only data available at time-of-admission. In addition to being highly effective for investigating sepsis, this approach supports the flexible incorporation of new data and can generalize to other diseases to aid in translational research and the development of precision medicine.

1 Introduction

Sepsis is a clinical syndrome that has been re-defined as life-threatening organ dysfunction caused by a dysregulated host response to infection (sepsis-3) [1]. Despite its clinical impact, and over 100 clinical studies [2], the mortality of sepsis remains high and the definition imprecise [2]. Sepsis patients are a heterogeneous population in terms of age, underlying comorbidities, infecting pathogen, and the infection foci. In addition, they present complex pathophysiology,

including concomitant inflammation and immunosuppression with rapid disease progression, further augmenting the heterogeneity of the syndrome. Currently, the severity of organ dysfunction in sepsis is measured by the Sequential Organ Failure Assessment (SOFA) scoring system and the standardized treatment strategy for sepsis includes early broad-spectrum antibiotic therapy, source control, and supportive care. Modulating the host response to infection has been proposed as a treatment strategy for decades without efficacy, although previous reports indicate that such treatments may be beneficial for some patients [2].

Previous attempts to stratify sepsis into more homogeneous subphenotypes have focused on subdividing patients based on clinical or molecular data [3–7]. The results of these studies are not always comparable and sometimes contradictory, motivating the need for a more personalized and adaptive approach for stratification. Optimally, subphenotypes are defined by a minimal number of molecular and/or clinical features to facilitate clinical implementation, and such approaches have shown to be useful in the treatment and definition of other diseases [8–12]. However, the use of hard cutoffs and rule-based methods, while minimizing the number of subphenotypes, make them heterogeneous and challenging to define. In addition, updating subphenotype models as new information is acquired is non-trivial, and definitions will need to be repeatedly revised to account for changes in the underlying data.

The concept of a digital twin (DT) historically emerged from the field of engineering to model real-world systems that may or may not currently exist [13]. In healthcare, the digital twin model leverages molecular and clinical data from well-defined patient cohorts as virtual representations of patients with known disease trajectories. This information is subsequently used to predict the current and future states of incoming patients with unknown disease trajectories using real-time data [14]. The emerging DT concept has been empowered by new developments in generative artificial intelligence and has already contributed to understanding disease, biomarker discovery, and drug development in other fields [14]. As a proof-of-concept, this model has been applied in sepsis to predict patient response using prospective clinical observations on 29 patients in the intensive care unit (ICU) [15]. The approach circumvents the need for defining rule-based phenotypes by instead analyzing patient neighborhoods to predict, diagnose, and model outcomes for new patients. As the concept is based on matching new patients with their closest digital families, this model can quickly accommodate new data without the need for redefining rules and provides a framework that will improve as more data is integrated.

Recent developments in liquid chromatography mass spectrometry (LC-MS/MS) proteomics have increased throughput, allowing for 1000s of samples to be run for a single study [16–18]. This increased throughput facilitates the population scale interrogation of proteome profiles in cohorts that closely resemble the pool of patients in a geographic area for a specific disease. This population-scale analysis can reveal hidden patterns specific to homogeneous groups of patients while maintaining statistical power to compare intricate differences between rare groups (based on microbial diversity, comorbidities, or treatment impact). Although population-scale proteomics may provide a more accurate picture of a disease, the increased sample size introduces data analysis issues, making interpretation difficult. Explainable artificial intelligence (XAI) has emerged as a promising feature selection technique to identify groups of proteins most strongly associated with a specific proteome state while remaining interpretable [19–26]. In systemic inflammatory syndromes, such as sepsis, this is particularly useful as many proteins may be found to be differentially abundant between conditions [27] and the complex signal requires simplification for interpretation [28]. XAI can help reduce the high-dimensionality of -omics data to an interpretable scale for downstream analysis.

In this study, we employed population scale proteomics to analyze a well-defined cohort of sepsis blood plasma samples taken at time-of-admission to the hospital. Using new developments in XAI, we provide a framework for the adaptive molecular investigation and digital twin modelling of sepsis to model clinical trajectories, predict mortality, and identify high-mortality-

risk patients in an adaptive and efficient manner.

2 Results

The overall goal of this study was to investigate if population scale proteomics could enable the prediction of clinical trajectories for sepsis patients using only data available at time-of-admission. In total, 3887 patients with suspected infection and the highest triage level (sepsis alert) [29] were admitted to the emergency department (ED) during the study period, with 3732 patients eligible for inclusion. In this study, 1264 were randomly selected and 100 additional patients were randomly chosen among the blood culture positive fraction to increase the likelihood of including patients with sepsis for a total of 1364 patients (**Figure 1a**). The patients were randomly split into two groups (80% train, 20% test) that retained the ratios of group labels in the overall cohort to generate a training set ($n = 1091$) for analysis and a held-out test set ($n = 273$) for unbiased evaluation throughout the study (**Figure 1b**).

A manual in-depth clinical chart review was conducted by infectious disease clinicians to create a validated clinical database using a structured protocol to generate detailed clinical information and to ensure the accuracy of the cohort. Over 370 clinical parameters per patient were retrieved manually and the likelihood of organ dysfunction and infection was validated to minimize the inclusion of sepsis mimics. For each of the 1364 patients, blood samples were collected at time-of-admission and the plasma was analyzed using high-resolution liquid chromatography and tandem mass spectrometry (LC-MS/MS) quantifying in total 838 proteins. A combination of the extracted clinical parameters and the time-of-admission plasma proteome maps were used for downstream analysis (**Figure 1c**). Additional clinical characteristics and demographics for the cohort are available in **Supplementary Table T1**.

There are 5 groups defined in the cohort: no infection and no organ dysfunction (1), no infection with organ dysfunction (2), infection without organ dysfunction (3), infection with organ dysfunction (sepsis-3) (4), and septic shock (5) (**Figure 1d**). These diagnoses are based on an increase in SOFA score and the presence of infection, determined using the Linder-Mellhammar Criteria for Infection (LMCI) [30] (**Figure 1e**). There is substantial overlap between non-sepsis (1-3) and sepsis (4-5) groups for the SOFA increase (**Figure 1f**) and the LMCI score (**Figure 1g**), highlighting the need to utilize both metrics simultaneously. Both SOFA and LMCI scores are clinical outcomes and are not available at time-of-admission.

2.1 Predicting sepsis patients from non-sepsis mimics

In the cohort, around 70% of the patients had sepsis or septic shock while around 10% of the patients had organ dysfunction unrelated to sepsis (**Figure 1d**). A uniform manifold approximation and projection (UMAP) of all quantified proteins in the 1364 proteome maps showed little separation between non-sepsis and sepsis patients (**Figure 1h**). Statistical comparison between the two groups revealed several differentially abundant proteins that were used to train a classifier to predict sepsis with a ROC-AUC of 0.71 based on the test set (**Supplementary Figure S1** and **Figure 1i**). Repeating the same process for septic shock, generated a protein panel that performed 33.8% better (0.95 ROC-AUC) than the original sepsis classifier on a filtered test set (non-sepsis vs. septic shock), but worse on the complete test set (**Supplementary Figure S1** and **Figure 1i**). A possible explanation for the relatively low performance of the classifiers is the heterogeneity within the cohort groups (1-5). To test this, we rescored all sepsis patients using the septic shock protein panel from the training set and separated them into groups of high probability severe sepsis (>0.5 probability) and low probability severe sepsis (≤ 0.5 probability) (**Figure 1i**). The more severe sepsis patients correspond to lower survival than the less severe sepsis group (**Figure 1k**), indicating that this strategy could be useful

in identifying patients with potentially worse survival using proteomic markers from time-of-admission samples. Collectively, these results demonstrate that the use of all quantified proteins generates relatively poor separation between the cohort groups. One possible explanation for this is that the predominant signal picked up from the high-dimensional proteomics data is the unspecific inflammatory response that does not separate the patients into discrete subgroups.

2.2 Organ dysfunction panels to construct a low-dimensional latent space

To stratify the cohort more effectively, we speculated that reducing the high-dimensional proteomics data into a low-dimensional interpretable embedding could reveal actionable patterns within the data. To that end, we mirrored the diagnosis of sepsis computationally by investigating the proteomic signatures derived from different SOFA categories (i.e. respiratory, coagulation, liver, renal, cardiovascular, and central nervous system (CNS)), the presence of a verified infection, and sepsis from non-sepsis. For these categories, we identified protein panels using differential analysis and XAI to select for proteins specific to the clinical outcome they were predicting (**Figure 2a**). The selected panels of proteins could predict future clinical diagnosis from time-of-admission samples with areas under the curve (AUC) ranging from 0.57 to 0.90 (**Figure 2b**). Aggregating all proteins from the individual protein panels (**Figure 2c**) generated a combined protein-organ dysfunction network comprised of 65 unique proteins (**Figure 2d**). Most proteins were singularly identified for a panel, indicating a strong association with their specific category. Two notable examples are the highly lung-specific pulmonary surfactant-associated protein B (PSPB) that was found most important for predicting respiratory dysfunction, and cystatin-c (CYTC), important in renal dysfunction, which has previously been identified as a blood-based biomarker for acute kidney injury [31]. Other examples, such as C-reactive protein (CRP), a common biomarker for infection and inflammation, and lipopolysaccharide-binding protein (LBP), a biomarker for infection [32], were both identified as important in predicting infection. Functional analysis with STRING-DB [33] revealed several significant interactions and biological processes statistically associated with the selected protein panels (**Supplementary Figure 2**). For example, pathways involved in lipid metabolism were found enriched for the liver panel, coagulation and platelet regulatory pathways were found enriched for the coagulation panel, and opsonization and the response to bacteria were found enriched for the infection panel. The biological relevance between the selected proteins and their corresponding categories emphasizes the robustness of our feature selection method, as the selected proteins are tied to biological functions of the different categories rather than just common inflammatory pathways. In the final step, we used these individual protein panels to predict probabilities for every patient in the training data to create an interpretable probabilistic latent space (ILS) that incorporates probabilities of organ dysfunction, infection, and sepsis to be used for further investigation into the molecular mechanisms of the syndrome (**Figure 2e**).

2.3 Identifying distinct subphenotypes using the interpretable probabilistic latent space

To determine if the interpretable latent space could reveal distinct subphenotypes, we visualized a UMAP of the ILS using only sepsis and septic shock patients (groups 4-5) from the cohort. In contrast to the full proteome, this UMAP revealed marked substructures of patients. We applied the k-means algorithm directly to the ILS to identify the 5 distinct subphenotypes within the data (**Figure 3a**). Each cluster was associated with distinct differences in predicted organ dysfunction (**Figure 3b**) and mortality rates (**Figure 3c**). Cluster 4 exhibited the best survival and was composed of high probability infection and sepsis but low probability severe organ dysfunction. Cluster 2 had high survival and was composed of high probability sepsis,

infection, and respiratory dysfunction, but low probability severe organ dysfunction in the other categories. In contrast, clusters 0, 1, and 3 had worse survival and were all composed of high probability sepsis, infection, and multiple severe organ dysfunction (**Figure 3b**).

To understand the molecular and clinical differences between the 5 subphenotypes, we used the previously described XAI methods to identify the five most prominent proteomic and clinical features for each cluster using the data available on time-of-admission (**Figure 3d**). Interestingly, the patients in cluster 1, associated with high probability liver dysfunction, had higher levels of bilirubin, which is a biomarker for liver dysfunction, than the other clusters as well as higher levels of lactate. The proteins found important for cluster 1 were also found in the liver dysfunction panel (**Figure 2c**). Cluster 0, the cluster with the highest mortality and high probability of renal dysfunction, was associated with high creatinine levels and lower systolic blood pressure. Although there were distinct differences between the identified subphenotypes, it is clear in **Figure 3d** that there is considerable overlap of the clinical parameters and sharedness of selected proteins between the clusters. These results indicate that the subphenotypes are heterogeneous and their associated features are not sufficient by themselves to accurately stratify sepsis patients.

To further investigate the degree of heterogeneity within the clusters, we performed hierarchical clustering of the ILS for each patient in each subphenotype. The analysis revealed that each cluster was comprised of distinct subclusters of patients revealing an additional 30 subphenotypes from the original 5 subphenotypes (**Figure 3e** and **Supplementary Figure 3**). For example, hierarchical clustering revealed 4 subclusters within cluster 1 that form separate structures in the UMAP (**Figure 3f**). All 4 subclusters consist of predicted liver dysfunction but have different frequencies of other organ dysfunctions (**Figure 3g**) and mortality rates (**Figure 3h**). These results show that the ILS can be used to identify potential subgroups within the sepsis cohort. However, these subgroups are heterogeneous themselves and the number of distinctly unique subgroups could be set arbitrarily high. From these findings, we conclude that the use of subphenotypes for the precision treatment of sepsis patients is not optimal and a more dynamic method that can adapt as new data is acquired could lead to more actionable translational results.

2.4 Modelling new patient outcome with adaptive digital family analysis

As digital twin modelling has previously been shown to enable predictions of future clinical trajectories, we first leverage the population scale proteomics data to predict the ILS for the entire training cohort, building a database of 1091 virtual patient representations. In the second step, probabilities for the 8 panels used to create the ILS were predicted for the 273 previously unseen patients from the held-out test set to provide an unbiased evaluation of the approach. Using these probabilities, we selected 273 digital families each consisting of 2.5% of the nearest neighbors in the database (**Figure 4a**). These digital families enable the dynamic modelling of all 370 clinical parameters extracted from the cohort using time-of-admission proteomics data.

The SOFA score increase is an important measure for poor prognosis but requires repeated measurements over multiple days. Here, we used the digital families to predict the increase in SOFA score after 24 hours for all 273 patients in the held-out test set and evaluated the accuracy. The digital families could be used to accurately predict the increase in SOFA score with a median error of 1 from time-of-admission samples, showcasing the ability to accurately predict future clinical outcomes for a patient (**Figure 4b**). Similar results were obtained when predicting the 30-day mortality for each patient within the held-out test set. Here, we calculated the 30-day mortality rate of each digital family and compared that rate to the bootstrapped (n=1000) mortality of test patients for each predicted mortality bin (**Figure 4c**). This analysis revealed a linear relationship between the predicted and observed mortality rate, although the

model overestimated mortality rates in high mortality families. Additionally, by calculating the mortality for each patient in the held-out test set based on their digital families, we can effectively create a mortality map of sepsis (**Figure 4d**). The map shows that the high mortality rates are concentrated in the top of the high-mortality clusters we identified above, but that the mortality landscapes within each cluster are not evenly distributed. If a new patient is placed in an area of high mortality, it is possible to determine the mortality risk of a patient based on their digital family without placing them in a subphenotype first.

Interestingly, the UMAP in **Figure 4d** shows that many of the digital families have mortality rates below 10%. To investigate if these predictions could be further subdivided, we modelled the survival profiles for each family (**Figure 4f**) to identify families with low overall mortality rates, but with a high proportion of mortality within the first 7 days. For each family, the ratio of 7-day mortality over 30-day mortality (mortality ratio) was calculated and plotted against the 30-day mortality of the digital family (**Figure 4e**), revealing four distinct subgroups of digital families. One group, comprised of high 30-day mortality rates with over 60% mortality ratio, was populated mostly by patients from the high mortality cluster 4 subphenotype. Conversely, there were two groups with low 30-day mortality rate but with drastically different mortality ratios. We highlight a patient in **Figure 4d-e** who died in less than 3 days in the hospital that was associated with a digital family with an overall predicted mortality of 20% (**Figure 4f**). However, the family had a proportionally high mortality ratio where more than 80% of the mortality occurred within 7-days. Although future evaluation of these patterns is required to determine the most useful parameters for clinical decision making, these results demonstrate that the digital family model generates new opportunities to analyze and stratify patients previously unavailable using conventional approaches. In conclusion, the methods used here showcase how the complexity of sepsis can be modelled with the ILS and how time-of-admission samples can be used to predict future clinical trajectories with high accuracy using the adaptive digital family approach. The results here suggest that digital family modelling represents an alternative strategy to rule-based subphenotyping and can provide a more dynamic approach to patient stratification. This adaptive approach could prove highly effective in the clinic for future implementations of precision medicine strategies in sepsis.

3 Discussion

In this study, we integrated population scale proteomics with clinical data and apply XAI to explore the molecular landscape of patients admitted to the ED with suspected sepsis. Using more than 1,300 time-of-admission patient samples, we demonstrate the heterogeneity of sepsis and advocate for a more adaptive personalized approach for the identification and treatment of sepsis patients using digital families. The population scale cohort of time-of-admission samples enabled us to detect early patterns of heterogeneity that may be lost in smaller cohorts and adequately visualize how heterogeneous the molecular landscape of sepsis is.

Digital twin models have been used successfully in several fields to understand disease, identify patients with similar response to different treatments, to predict outcome, and in drug development [14, 34]. However, to our knowledge, digital twin modelling has only been applied in one sepsis study of 29 patients to predict response to specific treatment [15]. We demonstrate how digital twin modelling can be applied on a large scale to accurately predict patient outcomes and identify high-risk patients that may require early intervention or altered care. Currently, the available data in our study is around 30-fold larger than existing digital twin sepsis studies. As the performance of digital twin models increase with data, the population-scale size of the cohort provides a substantial benefit. In addition, the model supports the iterative incorporation of data where new samples and other data modalities (i.e. metabolomics, lipidomics, genomics, transcriptomics) can be added to the database as they are acquired. With more diverse data, new patients will be more accurately linked to their digital families which will further improve

the modelling of their clinical disease trajectories. In our study, another area of potential improvement is the accuracy of the ILS. As the accuracy of the ILS depends on the performance of the related classifiers, an improvement in the underlying clinical labels has the potential to further benefit digital twin modelling. The fluid nature of the model will also allow the instant accommodation of new variations of sepsis when emerging diseases (i.e. COVID-19, invasive group A streptococcal disease) surge. The performance of these models would consistently and rapidly increase while providing a continuously evolving map of the sepsis disease landscape. This is in stark contrast to the use of subphenotypes for patient stratification, as subphenotype definitions will change constantly as new data is acquired and will require repeated updating. In complex and heterogeneous syndromes, such as sepsis, dynamic digital twin modelling provides the potential to unlock multi-omic and population scale studies for translation research and precision medicine.

An intriguing extension of the digital twin approach is the repurposing and identification of treatments for new patients. Here, a digital family could be used to guide treatment for new patients if a high proportion of the digital family was responsive to the treatment in the past. Digital families with similar molecular and clinical profiles could also be used as a tool in the future for adaptive clinical trials, where more homogeneous patient groups are selected at the time of admission and monitored over the course of the treatment. This is particularly relevant for sepsis, where over 100 clinical intervention trials have failed in the past four decades [2]. Potentially, some of these treatments could be rescued and applied to patients who present a matching receptive profile for that treatment. The digital twin model could accommodate this type of analysis by expanding to include time-course and treatment response data for individual patients and to extend the modality of the acquired data to analyze the plasma metabolome and lipidome.

Overall, we provide a robust framework for the comprehensive investigation of sepsis. We leverage population scale proteomics with XAI methods to model digital families for new patients that can predict survival, clinical parameters, and identify high risk patients. In addition to being highly effective for the study of sepsis, we believe this methodology could easily be applied to other complex diseases to aid in translational research and precision medicine.

4 Methods

The standards for reporting of diagnostic accuracy (STARD) guidelines were followed [35].

4.1 Settings and patient population

Patients were eligible for inclusion if ≥ 18 years and admitted to the emergency department (ED) at Skåne University Hospital, a tertiary hospital in Lund, Sweden as sepsis alert. Sepsis alert is a modified triage system to detect patients with suspected sepsis in the ED. The criteria for sepsis alert are patients with any of the following: active seizures, unconsciousness, a respiratory rate higher than 30 breaths/min or lower than 8 breaths/min, oxygen saturation below 90%, a regular heart rate over 130 beats/min, or an irregular heart rate over 150 beats/min, a systolic blood pressure (SBP) below 90 mmHg, p-lactate >3.5 mmol in combination with fever, $>38^\circ\text{C}$ or a history of fever or chills [29]. Patients meeting the criteria for sepsis alert are cared for immediately by emergency physician, infection consultant, nurse, and assistant nurse and undergo immediate control of vital parameters, physical examination, blood sampling for biochemical analyses and culture and other microbiological sampling. Sepsis alert patients are also considered for other diagnostic procedures, treatment, level of care and surveillance.

Patients were included prospectively, consecutively from 1st September 2016 to 31st of March 2023 and venous blood samples were drawn in citrate tubes at admission. Samples were centrifuged and stored at -70°C within 2 hours of collection.

4.2 Ethical permission

Informed consents were collected through an opt-out procedure. Patients received a mail following the ED admission and were given the opportunity to opt-out of inclusion in the study by sending back a form in a pre-franked envelope. Ethical approval for the study was obtained from the regional ethics board (file numbers 2022-01454-01, 2014/741 and 2016/271).

4.3 Clinical data collection

Clinical data was collected retrospectively through clinical chart review. Chart reviews were conducted by medical researchers using a structured protocol and validated by infectious disease clinicians.

Data were collected on demographics, vital signs, laboratory testing, microbiological investigations, medical history, diagnostic and therapeutic procedures, concomitant or new medications and level of care.

There are 5 groups defined in the cohort overall: no organ dysfunction under the sepsis-3 definition [1] (< 2 increase in sequential organ failure (SOFA)) and no infection based on the LMCI score (1), organ dysfunction under the sepsis-3 definition (≥ 2 increase in SOFA) with no infection according to the LMCI score (2), no organ dysfunction under the sepsis-3 definition with infection according to the LMCI score (3), and sepsis (4) and septic shock (5) defined according to Sepsis-3 definitions with organ dysfunction defined as an increase in sequential organ failure assessment (SOFA) ≥ 2 and verified infection was determined using the LMCI score. SOFA was modified to be compatible with use outside intensive care units (ICU) (**Supplementary Table T3**).

The following comorbidities were recorded: cardiovascular disease, liver disease, malignancy, respiratory disease, diabetes mellitus, chronic kidney disease, immunodeficiency and Charlson comorbidity index.

Microbiological diagnostics were performed at the clinical microbiology department (Laboratory Medicine Skåne, Lund, Sweden) per standard clinical practice.

4.4 Sample preparation for LC-MS/MS analysis

Plasma samples were diluted 1:10 with 100 mM ammonium bicarbonate (Sigma-Aldrich), and a total of 10 μ l of diluted sample (corresponding to 1 μ l plasma) was used for protein digestion. Samples were incubated for 60 minutes at 37°C in 4 M urea (Sigma-Aldrich) and 60 mM dithiothreitol (DTT, Sigma-Aldrich) for denaturation and reduction. Addition of DTT to the samples, as well as the following additions, were performed with an Agilent Bravo liquid handler. The samples were then alkylated using 80 mM 2-iodoacetamide (Sigma-Aldrich) for 30 minutes at room temperature in the dark, followed by digestion with 2 μ g LysC (Lysyl Endopeptidase, Mass Spectrometry Grade, Wako) for two hours at room temperature. The samples were further digested using 2 μ g trypsin (sequence-grade modified porcine trypsin, Promega) for 16 hours at room temperature. The digestion was stopped with 10% trifluoroacetic acid until pH ~ 2 . The samples were stored at -80 °C until LC-MS/MS analysis.

For peptide clean-up, digested samples were diluted 1:12 in water to generate suitable concentrations and subsequently loaded onto disposable Evotip Pure C18 trap columns (Evosep Biosystems, Odense, Denmark), which were prepared according to the manufacturer's instructions. Briefly, the Evotips were activated in 0.1% formic acid in acetonitrile, conditioned by wetting the tips in 2-propanol and equilibrated in 0.1% formic acid. Twenty microliters of each diluted sample were transferred to tips, followed by washing with 0.1% formic acid. The Evotips were stored in 0.1% formic acid until LC-MS/MS analysis.

4.5 LC-MS/MS analysis

Evosep One LC system (Evosep Biosystems) was used for separation with nanoflow reversed-phased chromatography. All samples were run with the 60 samples per day (SPD) method (gradient length of 21 min, vendor standard settings) with a flow rate of 1 $\mu\text{l}/\text{min}$. Samples were analyzed with timsTOF Pro 2 ion mobility mass spectrometer (Bruker Daltonics) using an 8 cm x 150 μm Evosep column (Evosep Biosystems) packed with 1.5 μm ReproSil-Pur C18-AQ particles. MS data was acquired using data-independent acquisition serial fragmentation (diaPASEF) method. The accumulation and ramp times were set to 100 ms. Variable isolation windows were used for the diaPASEF method and specified in **Supplementary Table T2**, with an estimated cycle time of 2.76 s. The collision energy was ramped linearly as a function of the mobility from 59 eV at $1/K0 = 1.5 \text{ Vs cm}^{-2}$ to 20 eV at $1/K0 = 0.6 \text{ Vs cm}^{-2}$.

4.6 LC-MS/MS data analysis

Data was analyzed using DIA-NN (1.8.1) [36] using a spectral library created from human tissue and plasma samples and match-between-runs enabled. Downstream analysis was performed using DPKS [21]. Samples were randomly split into a training (80%) and test (20%) set keeping the distributions of the included groups (1-5) intact and were processed separately using the following procedure to obtain protein quantities. Results were filtered for 1.0% false discovery rate at the precursor and library level. Precursors were normalized using a retention time sliding window mean normalization method [37] using DPKS. Proteins were quantified from precursors using the relative quantification *iq*[38] algorithm in the DPKS package.

4.7 Statistical analysis and explainable machine learning

All statistical analysis, explainable machine learning, and feature selection was performed using DPKS [21]. Statistical differential tests were performed between 2 groups using linear regression models for proteomics data, and ANOVA for clinical data. Multiple testing correction was performed using the Storey-Tibshirani method [39]. Features (proteins or clinical features) below a 0.1 corrected p-value were selected for downstream machine learning analysis. Feature importance was estimated using SHAP [40] with 100 bootstrapping iterations ($n=100$) to account for stochasticity in explainable machine learning methods caused by the background data using an XGBoost classifier [41]. The background data was resampled to correct for any class imbalances between groups during interpretation.

4.8 Sepsis prediction

Comparisons between groups of non-sepsis (1, 2, 3) ($n=346$) and sepsis (4, 5) ($n=745$) were performed using the described statistical and explainable machine learning methods. Protein ranks were calculated for 100 bootstrap iterations and the number of times a protein appeared in the top-20 was calculated. Proteins were selected if they were above 2 standard deviations from the mean number of occurrences for a protein in the top-20. An XGBoost classifier was trained to predict sepsis using the minimized protein panel. This classifier was validated using the unbiased held-out test set.

Comparisons between non-sepsis (1, 2, 3) ($n=346$) and septic shock (5) ($n=67$) were performed using the above described statistical and explainable machine learning methods. Protein ranks were calculated for 100 bootstrap iterations and the number of times a protein appeared in the top-20 was calculated. Proteins were selected if they were above 2 standard deviations from the mean number of occurrences for a protein in the top-20. An XGBoost classifier was trained to predict sepsis using the minimized protein panel. This classifier was validated using

the unbiased held-out test set (n=273) and a filtered test dataset consisting of just the sepsis groups 1, 2, 3, and 5 (non-sepsis and septic shock).

4.9 Septic shock stratification

Probabilities for all patients in the training data were calculated using the septic shock classifier. The non-septic shock sepsis patients (4) were split into 2 groups (≤ 0.5 and > 0.5 probability) and their survival rates over 30-days were plotted.

4.10 Protein panel identification

Eight clinical outcomes were identified and analyzed to investigate the underlying proteins associated with them: 6 Organ dysfunction groups based on SOFA scores (respiratory, renal, liver, cardiovascular, coagulation, and central nervous system), presence of infection, and sepsis-3 (1, 2, 3 vs. 4, 5). This analysis was performed using the training data and validated using the unbiased test set. The following comparisons were made for the 6 organ dysfunction groups:

- Respiratory organ dysfunction: respiratory SOFA 0 (n=92) to respiratory SOFA >2 (n=102)
- Coagulation organ dysfunction: platelets $>200 \times 10^9/L$ (n=502) to platelets below $150 \times 10^9/L$ (n=109)
- Liver organ dysfunction: bilirubin $<20 \mu\text{mol}/L$ (n=556) to bilirubin $>32 \mu\text{mol}/L$ (n=70)
- Cardiovascular organ dysfunction: cardiovascular SOFA 0 (n=618) to cardiovascular SOFA >0 (n=469)
- Central nervous system (CNS) organ dysfunction: CNS SOFA 0 (n=766) to CNS SOFA >0 (n=321)
- Renal organ dysfunction: renal SOFA 0 (n=615) to renal SOFA >0 (n=291)

The cutoffs are preferably chosen to provide a zone of uncertainty of 1 SOFA in-between to separate a more severe presence from the absence of organ dysfunction when possible. For respiratory dysfunction the zone of uncertainty between presence and absence of organ dysfunction is expanded since oxygen saturation is a volatile examination. As there is no definition for infection within the Sepsis-3 definition, we used the Linder-Mellhammar Criteria of Infection [30] (no infection = 139, verified infection = 741). The sepsis panel was determined using groups 1, 2, 3 (n=346) vs. 4, 5 (n=745) in the cohort.

For each of the 8 panels, proteins were selected using the statistical analysis and explainable machine learning methods described above. The percentage a protein is found in the top-10 in a bootstrap iteration (n=100) was calculated, and the top 90th percentile were selected as the representative proteins for a particular comparison.

To verify the biological function of the selected panels, we predicted high-confidence (> 0.9) functional interactions using STRING-DB [33] with up to 20 interactions and performed GO Biological Function overrepresentation tests to infer the biological function associated with the selected protein panels.

4.11 Creation of the probabilistic latent space

The probabilistic interpretable latent space is created by predicting probabilities for every patient in the training data for each of the 8 clinical outcome protein panels. This created a 1091 x 8 matrix that contains no missing values where each column represents the probability that a patient will develop one or several organ dysfunctions, infection, or sepsis based on time-of-admission samples. These probabilities are interpretable and can be associated with the molecular panel profiles identified using the explainable machine learning approach described above.

4.12 Identification and analysis of subphenotypes

The 8-dimensional ILS was clustered using the k-means algorithm ($k = 5$) into 5 different groups. The survival profiles were constructed for each subgroup over 365 days and were compared at 3 days, 30 days, and 365 days. To obtain the qualitative descriptions of the ILS for each of the clinical outcome protein panels, the average rate of panel prediction was calculated by dividing the number of positively predicted instances (> 0.5) by the total number of patients in a particular cluster.

The most important proteins and clinical parameters were identified for each subgroup using the statistical and explainable machine learning methods described above by comparing each cluster to a combination of all other clusters (one-vs-all). The 5 most important clinical parameters were visualized for each cluster against the values from all other clusters.

4.13 Creation of the digital family model and analysis of the unbiased held-out test set

Probabilities were predicted for the held-out test set ($n=273$) and neighborhoods consisting of 2.5% of the ILS were extracted the nearest neighbor algorithm to build digital families consisting of 27 patients per patient in the held-out test set. The predicted SOFA increase was derived for each family by calculating the median SOFA increase for day 1. The SOFA error was calculated as the difference between the predicted SOFA increase and the actual SOFA increase of the test patient. 30-day survival was calculated for 3 bins of predicted 30-day mortality for each digital family ($[0.0-0.1]$, $[0.1-0.3]$, $[0.3-0.6]$ predicted mortality bins). For each predicted 30-day mortality rate, the actual 30-day mortality rate of the test set was calculated for the current bin. This was bootstrapped ($n=1000$) to get the estimated 30-mortality rate in the held-out test set.

The mortality map for sepsis was created by calculating the proportion of patients within the digital family that died within 30-days. Each patient in the test data was colored by this rate and reduced using the same UMAP used to visualize the subphenotypes. The mortality ratio was calculated by modelling the 7-day mortality rate for each family and dividing it by the 30-day mortality rate for the family. These were plotted against each other to visualize how much of the 30-day mortality occurred earlier in the hospital visit.

5 Acknowledgements

L.M. is funded by the Swedish Research Council (grant number VR-2020-02419), the Wallenberg foundation (grant number 2016.0023) and Alfred Österlunds Foundation. J.M. is a Wallenberg academy fellow (KAW 2017.0271) and is also funded by the Swedish Research Council (Vetenskapsrådet, VR) (2019-01646 and 2018-05795), the Wallenberg foundation (KAW2016.0023, KAW2019.0353 and KAW2020.0299), and Alfred Österlunds Foundation. E.M. is funded by Wenner-Gren Foundation (FT2020-0003), the Crafoord Foundation, and the Swedish Society of

Medicine (SLS-985287). F.K. is funded by Region Skåne ALF project and the Crafoord Foundation. A.L. is funded by the Swedish Research Council VR 2023-02707 and Region Skane ALF project 2022-0146.

References

1. Singer, M. *et al.* The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **315**, 801–810. ISSN: 0098-7484 (2016).
2. Marshall, J. C. Why have clinical trials in sepsis failed? *Trends in Molecular Medicine* **20**, 195–203. ISSN: 1471-4914 (2014).
3. Burnham, K. L. *et al.* Shared and Distinct Aspects of the Sepsis Transcriptomic Response to Fecal Peritonitis and Pneumonia. *American Journal of Respiratory and Critical Care Medicine* **196**, 328–339. ISSN: 1073-449X (2017).
4. Bhavani, S. V. *et al.* Identifying Novel Sepsis Subphenotypes Using Temperature Trajectories. *American Journal of Respiratory and Critical Care Medicine* **200**, 327–335. ISSN: 1073-449X (2019).
5. Scicluna, B. P. *et al.* Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *The Lancet Respiratory Medicine* **5**, 816–826. ISSN: 2213-2600 (2017).
6. Seymour, C. W. *et al.* Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA* **321**, 2003–2017 (May 2019).
7. Antcliffe, D. B. *et al.* Transcriptomic Signatures in Sepsis and a Differential Response to Steroids. From the VANISH Randomized Trial. *American Journal of Respiratory and Critical Care Medicine* **199**, 980–986. ISSN: 1073-449X (2018).
8. Reddy, K. *et al.* Subphenotypes in critical care: translation into clinical practice. *The Lancet Respiratory Medicine* **8**, 631–643. ISSN: 2213-2600 (2020).
9. Calfee, C. S. *et al.* Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *The Lancet Respiratory Medicine* **2**, 611–620. ISSN: 2213-2600 (2014).
10. Zhang, H. *et al.* Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. *Nature Medicine*, 1–10. ISSN: 1078-8956 (2022).
11. Kuruvilla, M. E., Lee, F. E.-H. & Lee, G. B. Understanding Asthma Phenotypes, Endotypes, and Mechanisms of Disease. *Clinical Reviews in Allergy & Immunology* **56**, 219–233. ISSN: 1080-0549 (2019).
12. Cho, N. Molecular subtypes and imaging phenotypes of breast cancer. *Ultrasonography* **35**, 281–288. ISSN: 2288-5919 (2016).
13. Grieves, M. & Vickers, J. Transdisciplinary Perspectives on Complex Systems, New Findings and Approaches, 85–113 (2016).
14. Bordukova, M., Makarov, N., Rodriguez-Esteban, R., Schmich, F. & Menden, M. P. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. *Expert Opinion on Drug Discovery* **19**, 33–42. ISSN: 1746-0441 (2024).
15. Lal, A. *et al.* Development and Verification of a Digital Twin Patient Model to Predict Specific Treatment Response During the First 24 Hours of Sepsis. *Critical Care Explorations* **2**, e0249 (2020).

16. Meier, F. *et al.* diaPASEF: parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nature Methods* **17**, 1229–1236. <http://dx.doi.org/10.1038/s41592-020-00998-0> (2020).
17. Meier, F. *et al.* Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Molecular and Cellular Proteomics* **17**, 2534–2545 (2018).
18. Mi, Y. *et al.* High-throughput mass spectrometry maps the sepsis plasma proteome and differences in response. *medRxiv*, 2022.08.07.22278495 (2022).
19. Scott, A. M. *et al.* Generalized precursor prediction boosts identification rates and accuracy in mass spectrometry based proteomics. *Communications Biology* **6**, 628 (2023).
20. Hartman, E., Scott, A. M., Malmström, L. & Malmström, J. Interpreting biologically informed neural networks for enhanced biomarker discovery and pathway analysis (2023).
21. Scott, A. M., Hartman, E., Malmström, J. & Malmström, L. Explainable machine learning for the identification of proteome states via the data processing kitchen sink. *bioRxiv*. eprint: <https://www.biorxiv.org/content/early/2023/08/31/2023.08.30.555506.full.pdf>. <https://www.biorxiv.org/content/early/2023/08/31/2023.08.30.555506> (2023).
22. Álvarez, M. B. *et al.* Next generation pan-cancer blood proteome profiling using proximity extension assay. *Nature Communications* **14**, 4308 (2023).
23. Toussaint, P. A. *et al.* Explainable artificial intelligence for omics data: a systematic mapping study. *Briefings in Bioinformatics* **25**, bbad453. ISSN: 1467-5463 (2023).
24. Torun, F. M. *et al.* Transparent Exploration of Machine Learning for Biomarker Discovery from Proteomics and Omics Data. *Journal of Proteome Research*. ISSN: 1535-3893 (2022).
25. Carrasco-Zanini, J. *et al.* Proteomic signatures for identification of impaired glucose tolerance. *Nature Medicine* **28**, 2293–2300. ISSN: 1078-8956 (2022).
26. Niu, L. *et al.* Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nature Medicine* **28**, 1277–1287. <https://www.nature.com/articles/s41591-022-01850-y> (June 2022).
27. Karlsson, C. A. *et al.* Streptococcus pyogenes Infection and the Human Proteome with a Special Focus on the Immunoglobulin G-cleaving Enzyme IdeS*. *Molecular & Cellular Proteomics* **17**, 1097–1111. ISSN: 1535-9476 (2018).
28. Altman, N. & Krzywinski, M. The curse(s) of dimensionality. *Nature Methods* **15**, 399–400. ISSN: 1548-7091 (2018).
29. Rosenqvist, M., Fagerstrand, E., Lanbeck, P., Melander, O. & Åkesson, P. Sepsis Alert – a triage model that reduces time to antibiotics and length of hospital stay. *Infectious Diseases* **49**, 507–513. ISSN: 2374-4235 (2017).
30. Mellhammar, L. *et al.* New, Useful Criteria for Assessing the Evidence of Infection in Sepsis Research. *Critical Care Explorations* **4**, e0697 (2022).
31. Soto, K. *et al.* Cystatin C as a Marker of Acute Kidney Injury in the Emergency Department. *Clinical Journal of the American Society of Nephrology* **5**, 1745–1754. ISSN: 1555-9041 (2010).
32. Blairon, L., Wittebole, X. & Laterre, P.-F. Lipopolysaccharide-Binding Protein Serum Levels in Patients with Severe Sepsis Due to Gram-Positive and Fungal Infections. *The Journal of Infectious Diseases* **187**, 287–291. ISSN: 0022-1899 (2003).

33. Szklarczyk, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* **51**, D638–D646. ISSN: 0305-1048 (2022).
34. Osipov, A. *et al.* The Molecular Twin artificial-intelligence platform integrates multi-omic data to predict outcomes for pancreatic adenocarcinoma patients. *Nature Cancer*, 1–16 (2024).
35. Bossuyt, P. M. *et al.* STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ : British Medical Journal* **351**, h5527. ISSN: 0959-8138 (2015).
36. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **17** (2020).
37. Willforss, J., Chawade, A. & Levander, F. NormalyzerDE: Online Tool for Improved Normalization of Omics Expression Data and High-Sensitivity Differential Expression Analysis. *Journal of Proteome Research* (2019).
38. Pham, T. V., Henneman, A. A. & Jimenez, C. R. iq: an R package to estimate relative protein abundances from ion quantification in DIA-MS-based proteomics. *Bioinformatics* **36**, 2611–2613. <https://academic.oup.com/bioinformatics/article/36/8/2611/5697917> (Apr. 2020).
39. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, 9440–9445. ISSN: 0027-8424 (2003).
40. Lundberg, S. M. & Lee, S.-I. in *Advances in Neural Information Processing Systems 30* (eds Guyon, I. *et al.*) 4765–4774 (Curran Associates, Inc., 2017). <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
41. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **13-17-Aug**, 785–794. eprint: 1603.02754. <https://arxiv.org/abs/1603.02754v3> (Mar. 2016).

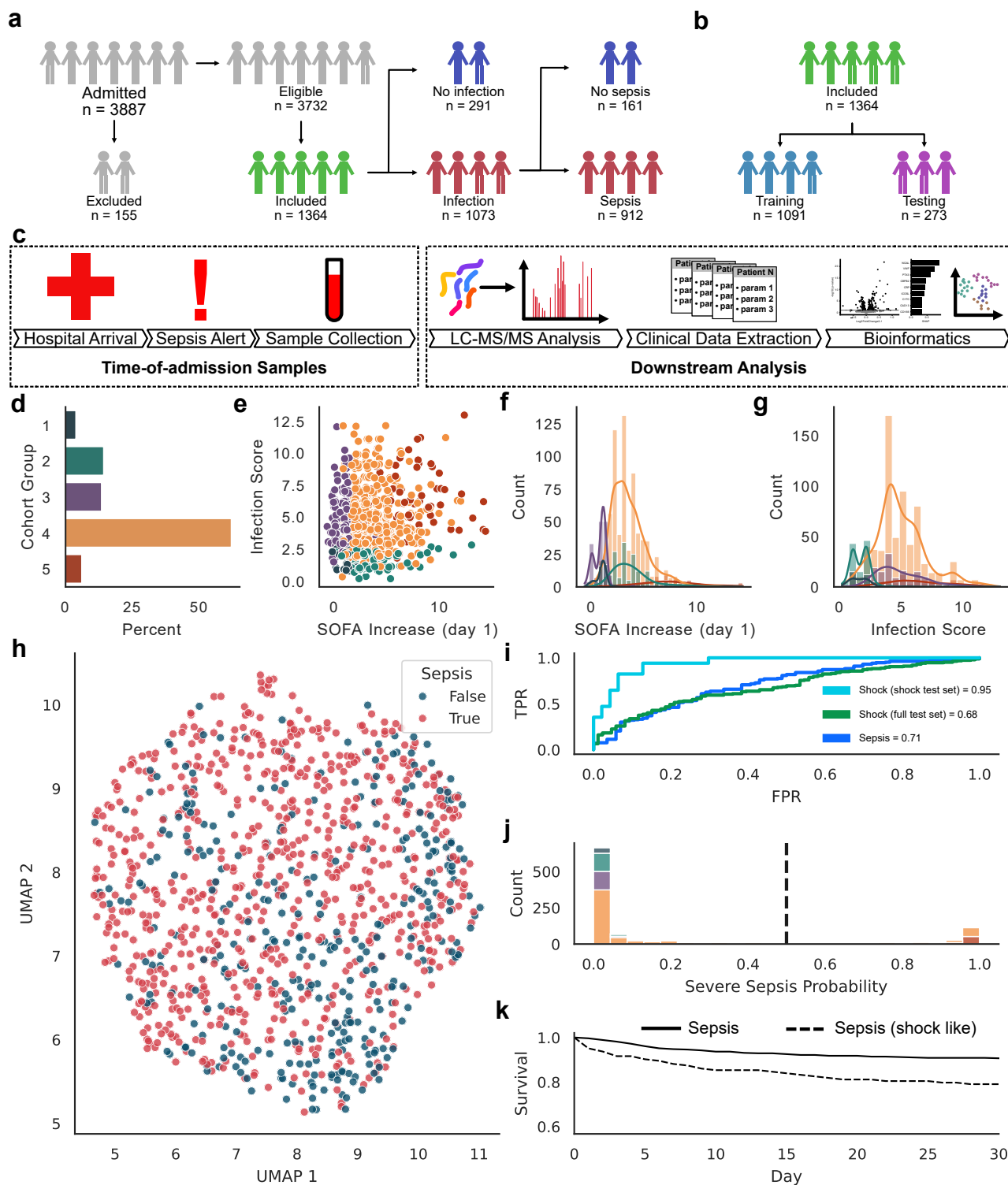


Figure 1: Cohort overview and the definition of sepsis proteome signatures: **a** Strobe chart visualizing the inclusion criteria for the patients in the study. **b** Split of the 1364 included patients into a training set (n=1091) and a test set (n=273). **c** An overview of the selection of patients and the acquisition of data used in the study. Time-of-admission samples were collected at the emergency department from patients suspected of sepsis and analyzed using LC-MS/MS and clinical parameters were extracted for each patient in the cohort. **d** A barplot showing the percentage distributions of 5 groups within the cohort (1-3 = non sepsis, 4-5 = sepsis) based on clinical parameters. **e** A scatter plot showing the distribution of SOFA score increase against the Linder-Mellhammar Criteria of Infection (LMCI) score colored by cohort group. **f-g** Distributions of SOFA increase for day 1 and LMCI score colored by the cohort group. **h** A UMAP of the proteome from the 1364 patient samples. Colors indicate sepsis (True = cohort group 4-5) or non-sepsis (False = cohort group 1-3). **i** ROC curves and AUC values for sepsis (full test data) and septic shock (full and reduced test data) classifiers. **j** Histogram of re-scored probabilities for all patients in the training set calculated using the septic shock classifier, colored by cohort group. The dotted line represents the 0.5 probability cutoff used to separate the sepsis-3 (4) patients. **k** Survival curves for sepsis-3 (group 4) patients predicted as high probability severe (shock like) or low probability severe using the septic shock classifier.

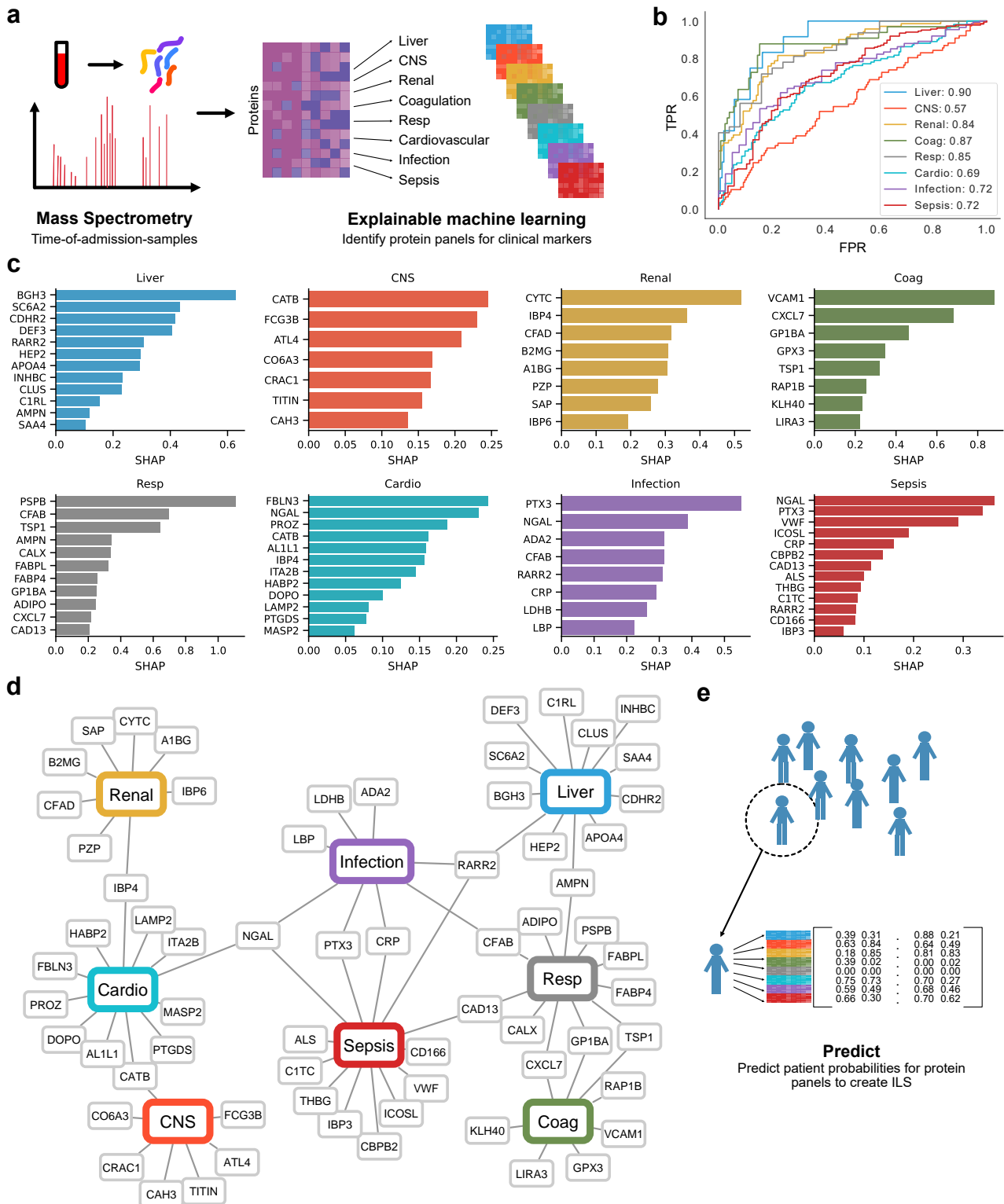


Figure 2: Explainable machine learning uncovers specific molecular panels in sepsis: **a** All time-of-admission plasma samples from the training data were analyzed using LC-MS/MS to obtain proteome maps for each patient sample. From the global list of identified proteins, 8 different clinical markers used in the diagnosis of sepsis (i.e. organ dysfunction, infection, sepsis) were analyzed and proteins were selected using explainable machine learning that were the most descriptive for each proteome state. **b** ROC curves displaying the performance of the selected protein panels on the held-out test data set. **c** Bar plots showing the mean importance, calculated using SHAP, of the selected proteins after 100 bootstrap iterations. **d** All proteins shown in (c) displayed as a network graph. Edges are unweighted and based on the inclusion of a protein in a panel. **e** An overview of the workflow used to predict organ dysfunction, infection, and sepsis for all patients in the training cohort to create the interpretable latent space. Probabilities for each of the 8 panels are predicted for each patient in the training data to assemble an 8-dimension latent space where each dimension represents a probability for one of the 8 panels.

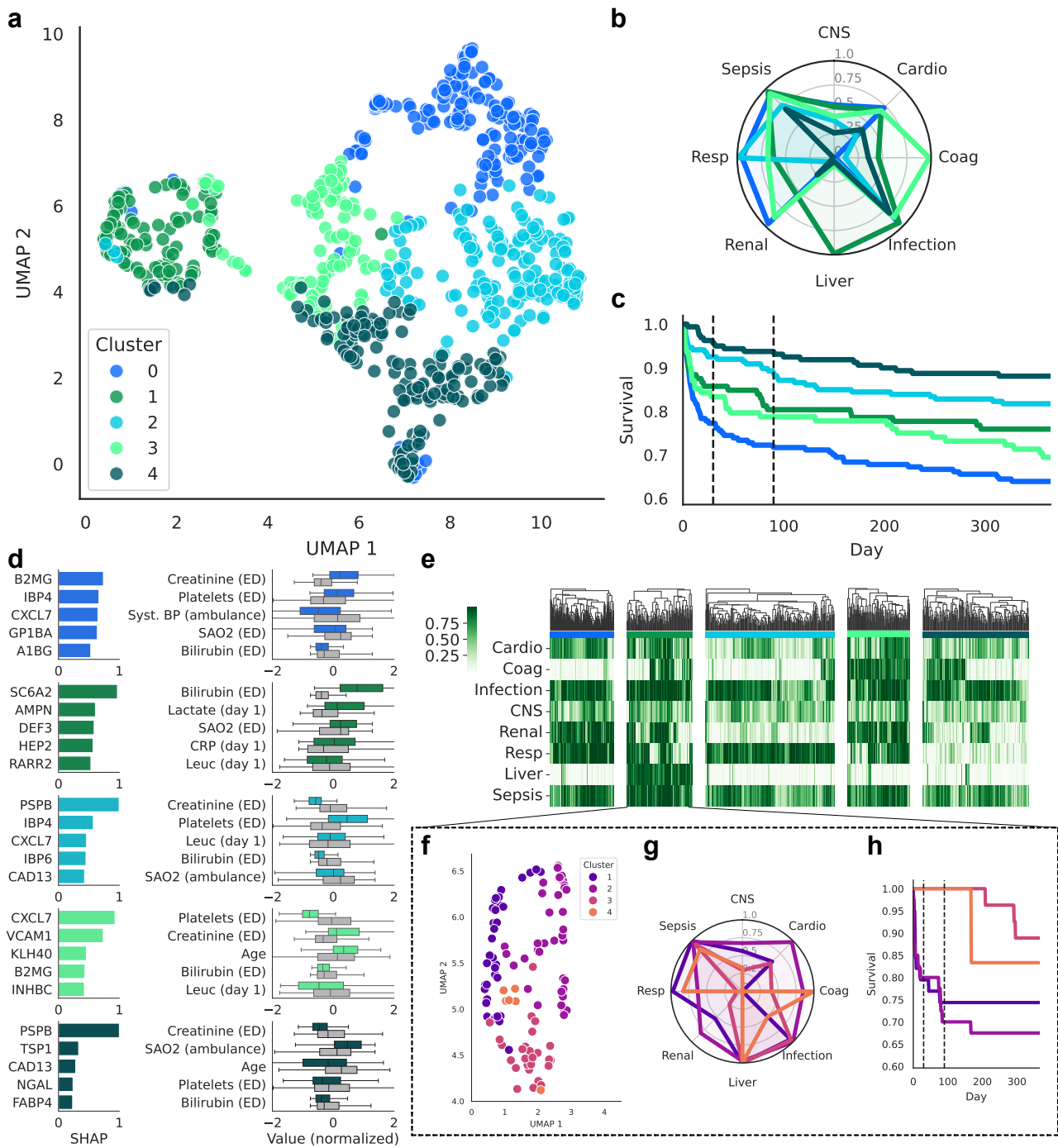


Figure 3: Identifying heterogeneous subgroups within sepsis: **a** A UMAP of the 8-dimensional latent space, subdivided into 5 clusters using the k-means algorithm on the 8-dimensional latent space and colored by the assigned cluster. The same colors are used throughout the figure to represent the 5 clusters (0-4). **b** The percentage organ dysfunction, infection, and sepsis was determined for each clusters and plotted in a radar plot and colored according to clusters. **c** The survival profiles of the 5 clusters across 365 days. The dotted lines highlight survival at 30 and 90 days respectively. **d** The top-5 most important proteins and clinical parameters selected using explainable machine learning in a one-vs-all analysis for each cluster. The gray box plots represent the values for all other clusters combined. The boxes represent the quartiles of the data, while the whiskers represent the remaining distribution. Outliers are not shown. **e** Cluster maps for each of the 5 clusters and 8 clinical parameters displaying the raw predicted probabilities of the latent space shown using a green color scale. **f** A zoomed in UMAP of cluster 1 from (a) colored by 4 subclusters identified using hierarchical clustering. These colors are used for the same subclusters in **g-h** **g** A radar plot of the percentage organ dysfunction, infection, and sepsis for the 4 unique subclusters found within cluster 1. **h** Survival profiles for the 4 subclusters within cluster 1. Similar graphs for the other 4 main clusters shown in (a) are available in the supplementary material.

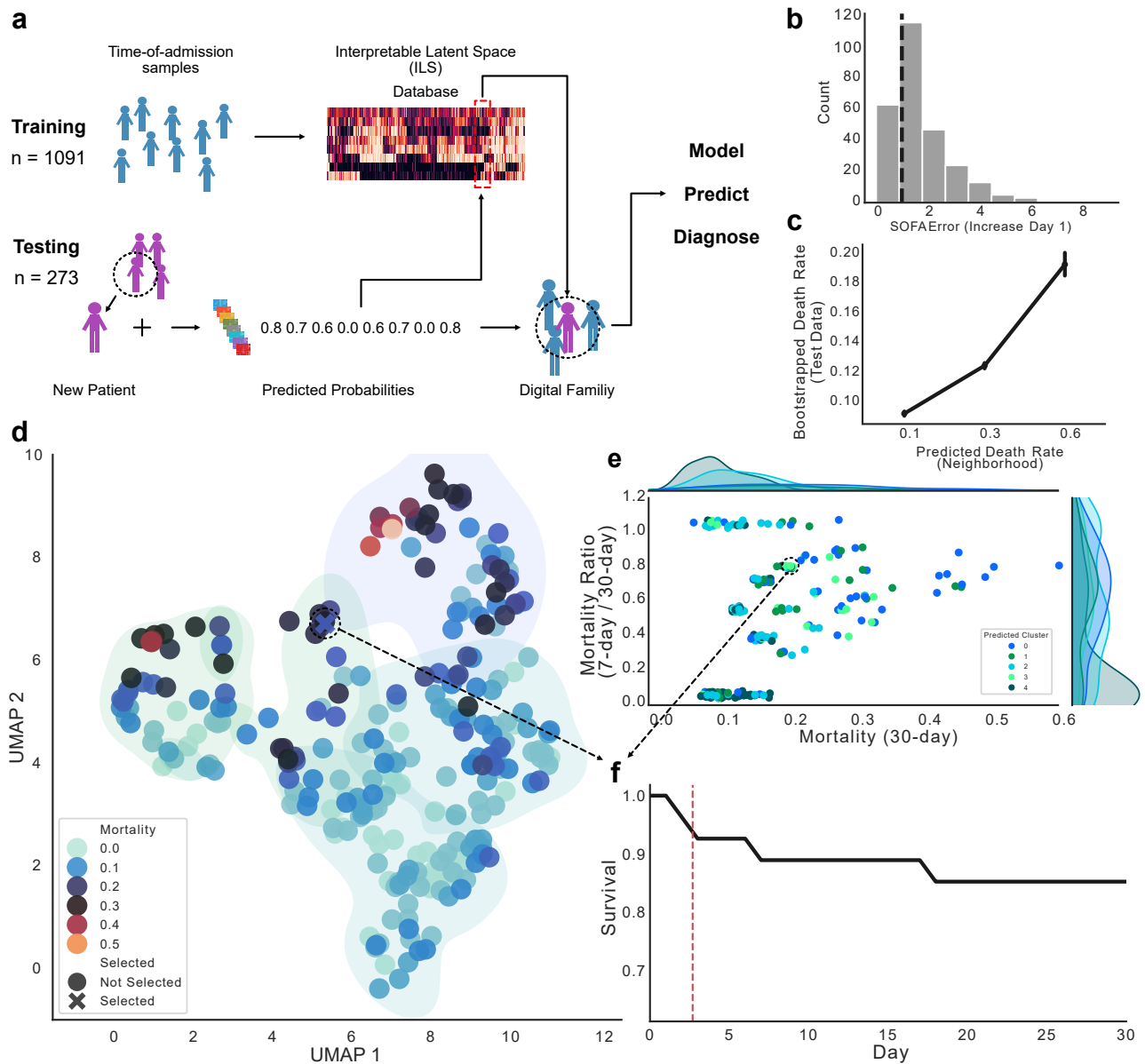
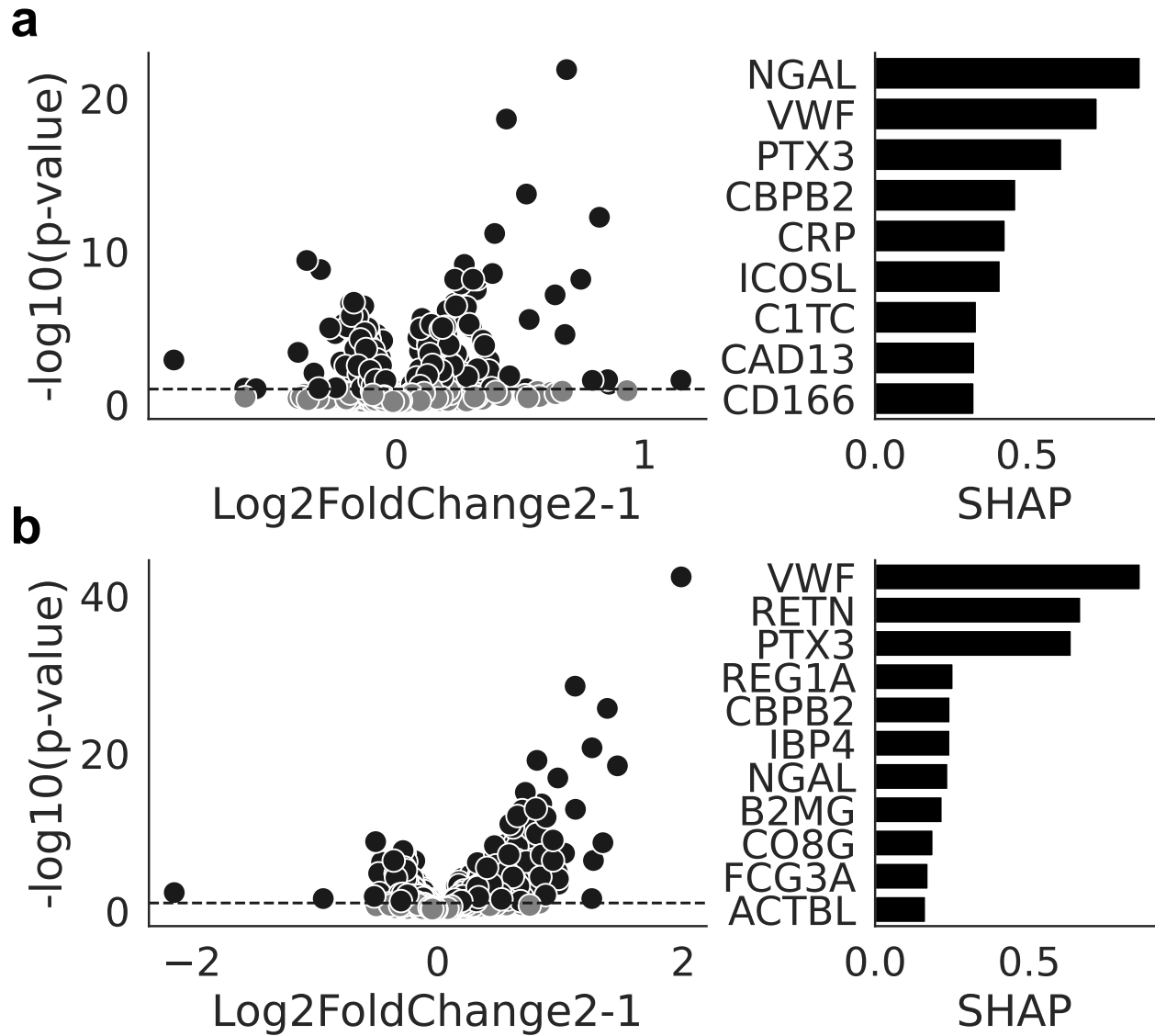


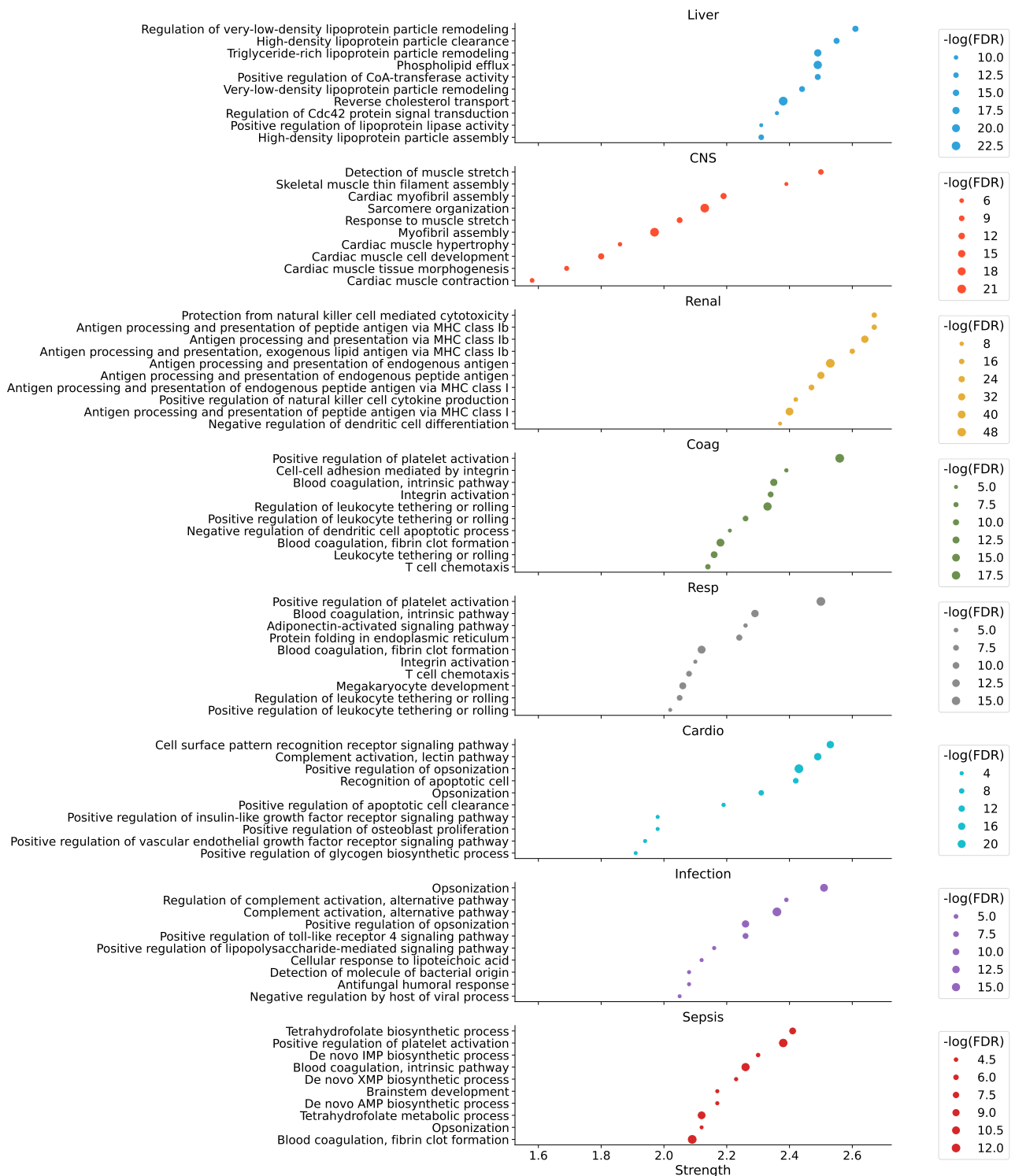
Figure 4: Digital family analysis facilitates accurately modelling of unknown disease trajectories.

a Using the proteome from time-of-admission samples of the training data and the detailed clinical data, the interpretable latent space (ILS) is predicted (Figure 2). This ILS is used as a database to determine groups of patients that are most similar to a new patient. Probabilities for new patients from the held-out test set were predicted for each of the 8 panels from Figure 2 and the nearest neighbors consisting of 2.5% of the database were selected as the digital family for each new patient. **b** A histogram of the error in predicted SOFA score, determined by taking the absolute value of the median SOFA score increase of the digital family subtracted from the actual SOFA score increase of the patient from the held-out test set. The dotted line represents the median SOFA error at 1.0. **c** A line plot showing the relationship between the predicted 30-mortality of the digital family, calculated as the percentage of how many in the digital family have survival less than 30-days, against the bootstrapped (n=1000) mortality rates in the test data. **d** Digital families were created for each patient in the held-out test set, and a UMAP was plotted of that test set using the same projection from Figure 3a. The colors indicate the predicted 30-day mortality of the digital family for each patient. The shaded colored outlines represent the 5 clusters from Figure 3. **e** Scatter plot of the 7-day survival ratio against the predicted 30-day mortality for each digital family with jitter added so all points are visible. Histograms for each axis are shown above and to the right of the plot. The coloring is based on the predicted cluster from Figure 3 of each patient in the held-out test set. **f** A selected survival profile for a patient with relatively high predicted 30-day mortality, but a higher 5-day survival gradient. This selected patient died within 5 days after admission to the hospital. Dotted circles show where this particular patient is placed in **d** and **e**.

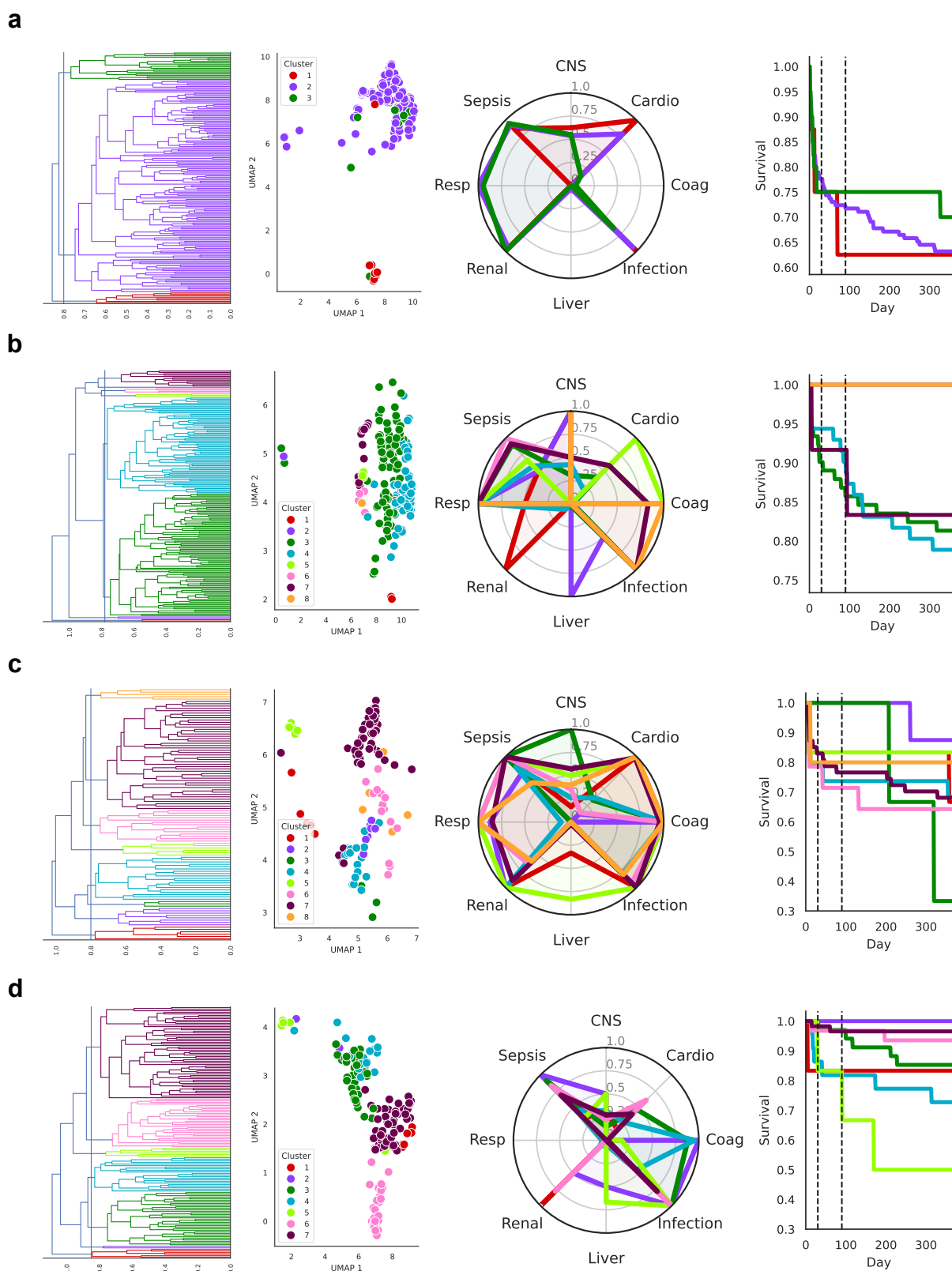
6 Supplementary Figures



Supplementary Figure S1: Proteomic comparison of the cohort **a** Volcano plot and selected proteins for the comparison of sepsis (4, 5) to non-sepsis (1, 2, 3). **b** Volcano plot and selected proteins for the comparisons of septic shock (5) to non-sepsis (1, 2, 3).



Supplementary Figure S2: Biological processes associated with the selected sepsis protein panels: This figure shows scatterplots for the top-10 GO Biological Processes found using STRING-DB to predicted high-confidence (≥ 0.9) interactions (up to 20 maximum interactions) for each protein panel. The size of the dots is based on the $-\log$ of the FDR.



Supplementary Figure S3: Subclusters of the identified subphenotypes a Subclusters of group 0, **b** subclusters of group 2, **c** subclusters of group 3, **d** subclusters of group 4

7 Supplementary Tables

Demographics	
Age, median (IQR)	76.0 (66.5-83.8)
Gender, female, n (%)	524 (38.4)
Comorbidities	
Diabetes Mellitus, n (%)	358 (26.2)
Cardiovascular disease, n (%)	565 (41.4)
Respiratory disease, n (%)	262 (19.2)
Liver disease, n (%)	20 (1.5)
Renal disease, n (%)	140 (10.3)
Malignancy, n (%)	234 (17.2)
Immunodeficiency, n (%)	150 (11.0)
Charlson Comorbidity Index (IQR)	5 (3-6)
No comorbidities, n (%)	384 (28.2)
Organ dysfunction	
SOFA increase day 1, median (IQR)	3 (2-4)
CNS SOFA increase ≥ 1 day 1	399 (29.3)
Cardiovascular SOFA increase ≥ 1 day 1	583 (42.7)
Respiratory SOFA increase , 73 day 1	1217 (89.3)
Renal SOFA increase ≥ 1 day 1	450 (32.9)
Liver SOFA increase ≥ 1 day 1	192 (14.1)
Coagulation SOFA increase ≥ 1 day 1	179 (13.1)
Diagnosis	
No, possible or indeterminable infection, n (%)	290 (21.3)
Infection probable or verified, n (%)	1073 (78.7)
Infection foci	n = 1073
Pneumonia, n (%)	160 (14.9)
Urinary tract infection, n (%)	282 (26.3)
Abdominal infection, n (%)	49 (4.6)
Skin or soft tissue infection, n (%)	99 (9.2)
Bone or joint infection, n (%)	12 (11.2)
Upper respiratory tract, n (%)	16 (1.5)
COVID-19, n (%)	40 (3.7)
Blood stream infection, n (%)	73 (6.8)
Other, n (%)	296 (21.7)
Bacteriemia	n = 586
Escherichia Coli, n (%)	220 (37.5)
Other Enterobacteriales, n (%)	119 (20.3)
Beta-hemolytic streptococci, n (%)	49 (8.4)
Staphylococcus Aureus, n (%)	61 (10.4)
Streptococcus pneumoniae, n (%)	27 (4.6)
Enterococci, n (%)	37 (6.3)
Sepsis	n = 932
Septic shock, n (%)	84 (9.0)
Critical care, n (%)	124 (13.3)
In-hospital mortality, n (%)	108 (11.6)
28-day mortality, n (%)	119 (12.7)
90-day mortality, n (%)	158 (16.9)

Supplementary Table T1: Demographic and clinical characteristics, n=1364

MS Type	Cycle ID	1/K0 Begin [Vs/cm2]	1/K0 End [Vs/cm2]	Start Mass [m/z]	End Mass [m/z]
PASEF	1	0.83	1.6	681.04	708.33
PASEF	1	0.6	0.83	258.48	409.55
PASEF	2	0.88	1.6	708.33	733.83
PASEF	2	0.6	0.88	409.55	444.26
PASEF	3	0.89	1.6	733.83	760.32
PASEF	3	0.6	0.89	444.26	474.76
PASEF	4	0.9	1.6	760.32	790.62
PASEF	4	0.6	0.9	474.76	502.25
PASEF	5	0.92	1.6	790.62	823.07
PASEF	5	0.6	0.92	502.25	525.91
PASEF	6	0.93	1.6	823.07	858.4
PASEF	6	0.6	0.93	525.91	548.77
PASEF	7	0.94	1.6	858.4	899.09
PASEF	7	0.6	0.94	548.77	571.27
PASEF	8	0.96	1.6	899.09	948.46
PASEF	8	0.6	0.96	571.27	592.32
PASEF	9	0.98	1.6	948.46	1010.99
PASEF	9	0.6	0.98	592.32	612.8
PASEF	10	0.99	1.6	1010.99	1079.07
PASEF	10	0.6	0.99	612.8	634.8
PASEF	11	1.02	1.6	1079.07	1187.54
PASEF	11	0.6	1.02	634.8	656.97
PASEF	12	1.08	1.6	1187.54	1443.66
PASEF	12	0.6	1.08	656.97	681.04

Supplementary Table T2: LC-MS/MS Acquisition Windows.

Organ system	SOFA score			
	1	2	3	4
Respiration PaO_2/FIO_2 mmHg corresponding SpO_2 without O_2 support	<400 <95%	<300 <91%	<200 with respiratory support	<100
Coagulation Platelets $\times 10^9/L$	<150	<100	<50	<20
Liver Bilirubin $\mu mol/L$	>19	>32	>101	>204
Cardiovascular Hypotension	MAP < 70mmHg	Dopamine ≤ 5 $\mu g/kg/min$ or dobutamine	Dopamine >5 or epinephrine or norepinephrine $\leq 0,1 \mu g/kg/min$ for at least one hour	Dopamine >15 or epinephrine or norepinephrine $>0.1 \mu g/kg/min$
Central Nervous System (CNS) Glasgow Coma Scale	13-14	10-12	6-9	<6
Renal Creatinine $\mu mol/L$ or urine output mL/day	>110	>170	>300 <500	>440 <200

Supplementary Table T3: SOFA definitions.