High-throughput mass spectrometry maps the sepsis plasma proteome and differences in patient response

Yuxin Mi[1]†, Katie L. Burnham[1,2]‡, Philip D. Charles[3]‡, Raphael Heilig[3]‡§, Iolanda Vendrell[3,4]‡, Justin Whalley[1]¶, Hew D. Torrance[5], David B. Antcliffe[5,6], Shaun M. May[7], Matt J. Neville[8,9], Georgina Berridge[3], Paula Hutton[10], Cyndi G. Geoghegan[1], Jayachandran Radhakrishnan[1], Alexey I. Nesvizhskii[11], Fengchao Yu[11], GAinS Investigators#, Emma E. Davenport[1,2], Stuart McKechnie[10], Roger Davies[5], David J. P. O'Callaghan[5,6], Parind Patel[6], Ana G. del Arroyo[7], Fredrik Karpe[8,9], Anthony C. Gordon[5,6], Gareth L. Ackland[7], Charles J. Hinds[7], Roman Fischer[3,4]*,, Julian C. Knight[1,4,9]*,****

Sepsis, the dysregulated host response to infection causing life-threatening organ dysfunction, is a global health challenge requiring better understanding of pathophysiology and new therapeutic approaches. Here, we ap- plied high-throughput tandem mass spectrometry to delineate the plasma proteome for sepsis and comparator groups (noninfected critical illness, postoperative inflammation, and healthy volunteers) involving 2612 samples (from 1611 patients) and 4553 liquid chromatography–mass spectrometry analyses acquired through a single batch of continuous measurements, with a throughput of 100 samples per day. We show how this scale of data can delineate proteins, pathways, and coexpression modules in sepsis and be integrated with paired leukocyte transcriptomic data (837 samples from *n* = 649 patients). We mapped the plasma proteomic landscape of the host response in sepsis, including changes over time, and identified features relating to etiology, clinical pheno- types (including organ failures), and severity. This work reveals subphenotypes informative for sepsis response state, disease processes, and outcome; identifies potential biomarkers; and advances opportunities for a precision medicine approach to sepsis.

INTRODUCTION

Sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection (*1*). Currently, we lack effective immunomodulatory therapies to address the high mortality and global burden of this disease (*2*, *3*). Incomplete knowledge of pathophysiology and failure to define individual patient variation in the nature and timing of maladaptive host responses within the heterogeneous clinical syndrome of sepsis currently limit the design of clinical trials (*3*–*5*). Sepsis subphenotypes informative for immune response state, outcome, and therapeutic response are proposed on the basis of clinical, laboratory, and molecular stratifiers (*6*–*13*). However, establishing the nature of the sepsis host response and opportunities to optimally stratify patients to deliver precision medicine approaches has been constrained by incomplete knowledge of the sepsis plasma proteome. The plasma proteome reflects organ function through the secretome and tissue leakage products, offering the opportunity to identify key media- tors of the sepsis response; potential therapeutic targets; and bio- markers of individual variation in pathological state, disease severity, and outcome.

To date, there have been technological limitations to the high-throughput application of quantitative assays that are able to capture the high dynamic abundance range of proteins in the blood. Analysis of the sepsis plasma proteome has focused on mouse models, a small number of plasma cytokines and metabolites in patients, or mortality prediction and comparison with healthy individuals and sterile inflammation, involving relatively small numbers of cases (*14*–*24*).

Tandem mass spectrometry (MS) provides protein measurements in an untargeted and hypothesis-free manner suitable for discovery-led characterization of the sepsis blood proteome. Here, we show how, with higher-throughput automated and robust methods for sample preparation alongside MS-based data acquisition and data analysis, it is feasible to analyze >2500 nondepleted blood plasma samples in a single batch using a single liquid chromatography–MS (LC-MS) platform. We report the plasma proteome of >1000 adult patients with sepsis at multiple time points and integrate this with leukocyte transcriptomics to provide insights into the nature of the sepsis response and observed clinical heterogeneity.

RESULTS

High-throughput MS delineates the plasma proteome at scale

We aimed to characterize the acute sepsis plasma proteome. To do this, we analyzed patients with sepsis due to community acquired pneumonia (CAP) or fecal peritonitis (FP) admitted to the intensive care unit (ICU) and serially sampled during their admission (n = 1189 patients, 1879 sam- ples) as part of the UK Genomic Advances in Sepsis (UK GAinS) study (Fig. 1A and data file S1) (*10*, *25*). The median age was 65 years [interquar- tile range (IQR), 53 to 75], 54% were male, 64% mechanically ventilated, and 59% had shock; the median Sequential Organ Failure Assessment (SOFA) score was 6 (IQR, 3 to 8), and 28-day mortality was 17% (data file S1). These patients were divided into a discovery and a validation set, with additional cohorts as comparator groups, including a clinical trial of all cause sepsis requiring vasopressors (n = 45 patients, 154 samples), healthy volunteers (HVs) (n = 152 individuals and samples), and comparison with noninfectious causes of inflammation [elective surgery patients before and after operation (n = 149 patients, 351 samples)] and with nonin- fected ICU patients (n = 76 patients, 76 samples) (Fig. 1A and data file S1). Traditionally, MS-based proteomics is a low-throughput technique in the range of 10 to 20 samples per day when nanoflow ultraperformance liquid chromatography (UPLC) is used to maximize sensitivity (*26*). In this study, we developed a high-throughput quantitative proteomics work- flow, using a combination of Evosep One high-performance liquid chro- matography (HPLC) and Bruker timsTOF Pro (trapped ion mobility spectrometry time of flight) on a total of 2612 plasma samples from 1611 individuals (Fig. 1A) in one batch across 28 fully randomized acquisition plates to minimize assay variability between individuals and cohorts. In total, we acquired 4553 LC-MS analyses, including a prefractionated, super-depleted (*27*) master pool with continuous data acquisition at a throughput equivalent to 100 samples per day in data-dependent acquisi- tion mode (DDA-PASEF). The 4553 LC-MS injections comprised 2647 injections from samples (including 35 duplicates) and 1906 injections from library fractions and quality control. The complete dataset com- prised 250 million MS/MS spectra matching to a total of 2782 protein groups. In addition, we injected two nondepleted master pool samples every 24 cohort sample injections for subsequent identification trans- fer ("match between runs"), monitoring MS platform performance/sta- bility and correction for any variability as part of quality control (fig. S1). We customized data preprocessing to minimize potential techni- cal bias and maximize comparability between samples (workflow in fig. S1A). We analyzed raw protein intensities derived from Frag- Pipe, identifying 291 proteins reliably detected (in ≥50% samples) in at least one biological group, and removed 32 sample injections with few proteins detected, 35 duplicated MS injections or duplicated sample aliquots, five samples from excluded patients, and 22 pro- teins affected by cell residue contaminations in the plasma (fig. S1, A to E). We used variance stabilizing normalization to account for sys- tematic bias and applied k-nearest neighbors to impute missing val- ues based on the most similar proteins (for 170 proteins detected in ≥60% of the samples) or imputation by random draw from down-shifted normal distributions for the remainder. The processed data comprised 269 proteins in 2575 samples from 1598 individuals (Fig. 1A).

The proteome profile reveals an axis of severity across cohorts

We first sought to understand variation in plasma protein abundance and enrichment for biological processes across all cohorts. Reducing

the dimensionality of the data, we found that principal component (PC) 1 formed a sample gradient from HVs and elective surgery pre- operative cases, to postoperative cases and noninfectious critical ill- ness, to sepsis (Fig. 1B). Proteins with high positive loadings for PC1 included acute-phase [CRP (C-reactive protein), SAA1 (serum amy- loid A1), SAA2 (serum amyloid 2), SERPINA1 (serpin family A mem- ber 1, also referred to as alpha-1-antitrypsin), SERPINA3 (serpin family A member 3, also called alpha-1-antichymotrypsin), HP (haptoglobin), and C1RL (complement C1r subcomponent like)], S100 proinflammatory (S100A8, S100A9, and S100A12), innate im- mune or antibacterial [LCN2 (lipocalin 2), LBP (lipopolysaccharide binding protein), and USP15 (ubiquitin carboxyl-terminal hydrolase 15)], and extracellular matrix (ECM) proteins [TNC (tenascin C), MMP2 (matrix metalloproteinase-2), COL1A2 (collagen type I alpha 2 chain), and COL6A1 (collagen type 6 alpha 1 chain)], whereas lipid transport protein APOM (apolipoprotein M) had a high negative loading (Fig. 1C). We identified protein clusters on the basis of protein-protein interactions that were enriched for biological processes involving ECM, coagulation, lipid metabolism, acute-phase response, and neutrophil degranulation (fig. S1F).

We further analyzed functional groupings by deriving protein set en- richment scores using gene set enrichment analysis. We analyzed the resulting matrix by principal components analysis (PCA) and found that PC1 showed a gradient across sample cohorts with highest loadings for antimicrobial humoral response, cell chemotaxis, and positive regula- tion of response to external stimulus (toward sepsis/severe disease) and for lipoprotein metabolic process (toward nonsepsis controls). PC2 in- volved regulation of B cell activation and Fc receptor signaling pathway (Fig. 1D).

We also derived protein coexpression networks using all samples from the cohorts, grouping 184 proteins into 16 coexpression mod- ules that varied by comparator cohort and were enriched for acute- phase proteins (blue module, higher in sepsis), plasma lipoprotein assembly (yellow, higher in HVs), platelet degranulation (brown, high- er in sepsis), and immunoglobulins (black, higher in sepsis; purple, higher in noninfected ICU patients; and magenta, higher in sepsis and noninfected ICU patients) (Fig. 1, E and F, and fig. S1G).

Sepsis is associated with a distinct plasma proteome profile We proceeded to investigate differential protein abundance and pathway enrichment in sepsis versus other contexts (Fig. 2A). Here and subsequently, sepsis refers to the patients with CAP/FP sepsis admitted to ICU in the UK GAinS study unless stated otherwise. Across the six sepsis-comparator group contrasts in the discovery and validation cohorts (Fig. 2B), we found 11 proteins differentially abundant in all contrasts, all with highest abundance in sepsis. These involved the acute-phase response (CRP, LCN2, SERPINA1, and HP), ECM (MMP2, COL6A1, and TNC), protection from tissue damage (SERPINA1, HP, and TNC), neutrophil function (LCN2, MMP2, SERPINA1, S100A9, and S100A12), cytokine production (LCN2, MMP2, USP15, SERPINA1, HP, TNC, and S100A12), and galactose metabolism (B4GALT1, beta-1,4-galactosyltransferase 1) (Fig. 2, C and D, and fig. S2A). Elastic net prediction models based on protein differences in the discovery cohort could distinguish be- tween the validation populations [area under curve (AUC)] (95% confidence interval, CI) = 100.0% (99.9 to 100.0%) for Sepsis_ICU versus HV; 96.9% (94.6 to 99.2%) for Sepsis_ICU versus PostOp], illustrating that protein differences were stable across the different patient cohorts.

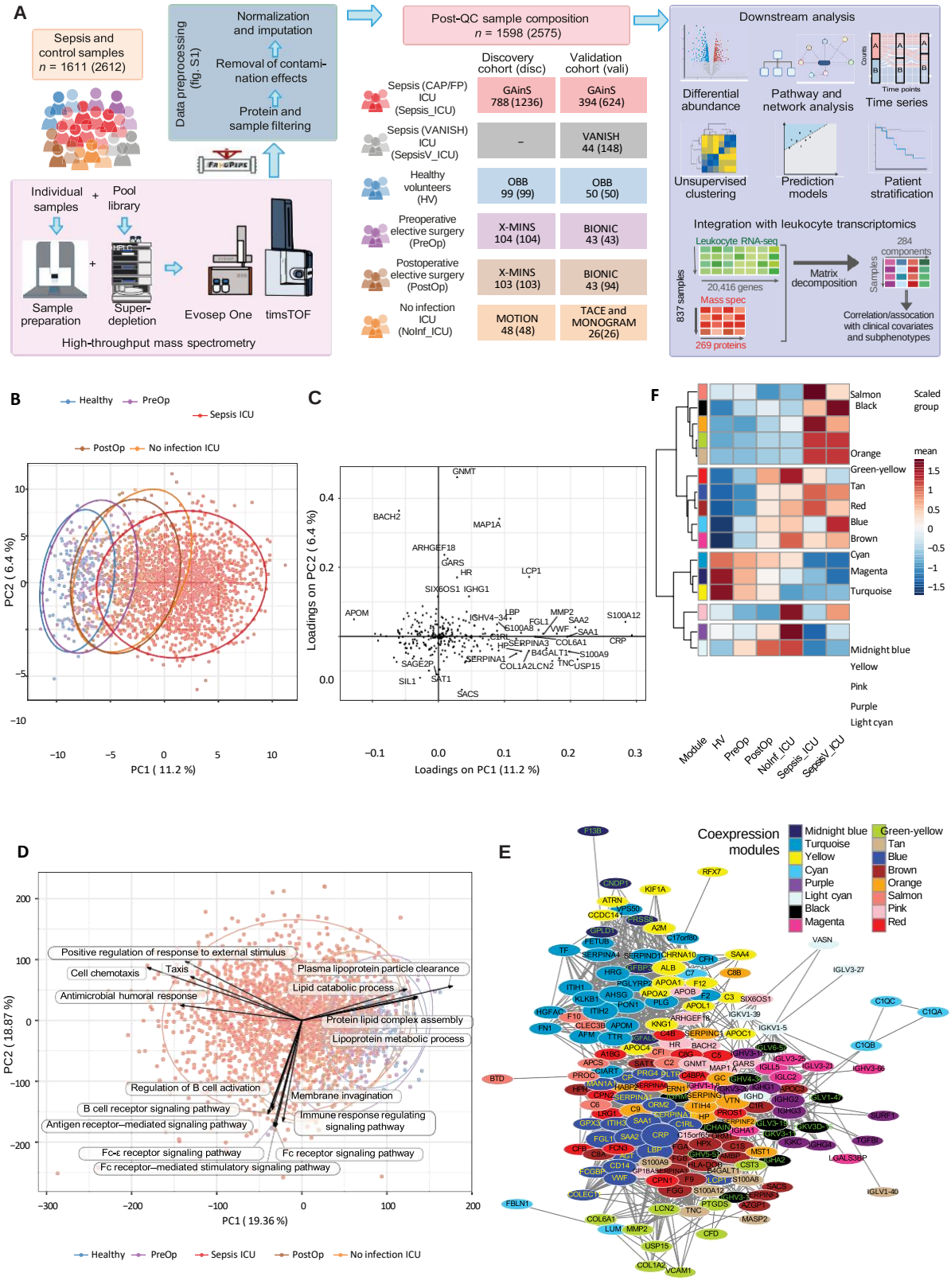


Fig. 1. Study overview and sample differentiation by plasma proteome. (**A**) Study design, workflow, and cohorts, with study name, numbers of individuals assayed (*n*), and number of samples (in brackets). Qc, quality control. (**B**) Pca based on protein abundance in all samples, showing Pc1 versus Pc2 with 95% data ellipses (assum- ing a multivariate *t* distribution). (**C**)

Protein loadings on Pc1 and Pc2. (**D**) Pca of enrichment score matrix on all samples from gene set enrichment analysis using protein abundance for single samples. arrows, gene ontology biological processes (top 8 loadings Pc1 and Pc2), length scaled to loading. (**E** and **F**) Protein coexpression network from weighted gene coexpression network analysis. (e) Module network, with edge weight denoting topological overlap between connected nodes; node size denoting within-module connectivity. (F) relationship of the coexpression modules with cohorts, showing the mean of module eigengenes. Patient numbers for (B), (d), and (F) are as shown in (a).

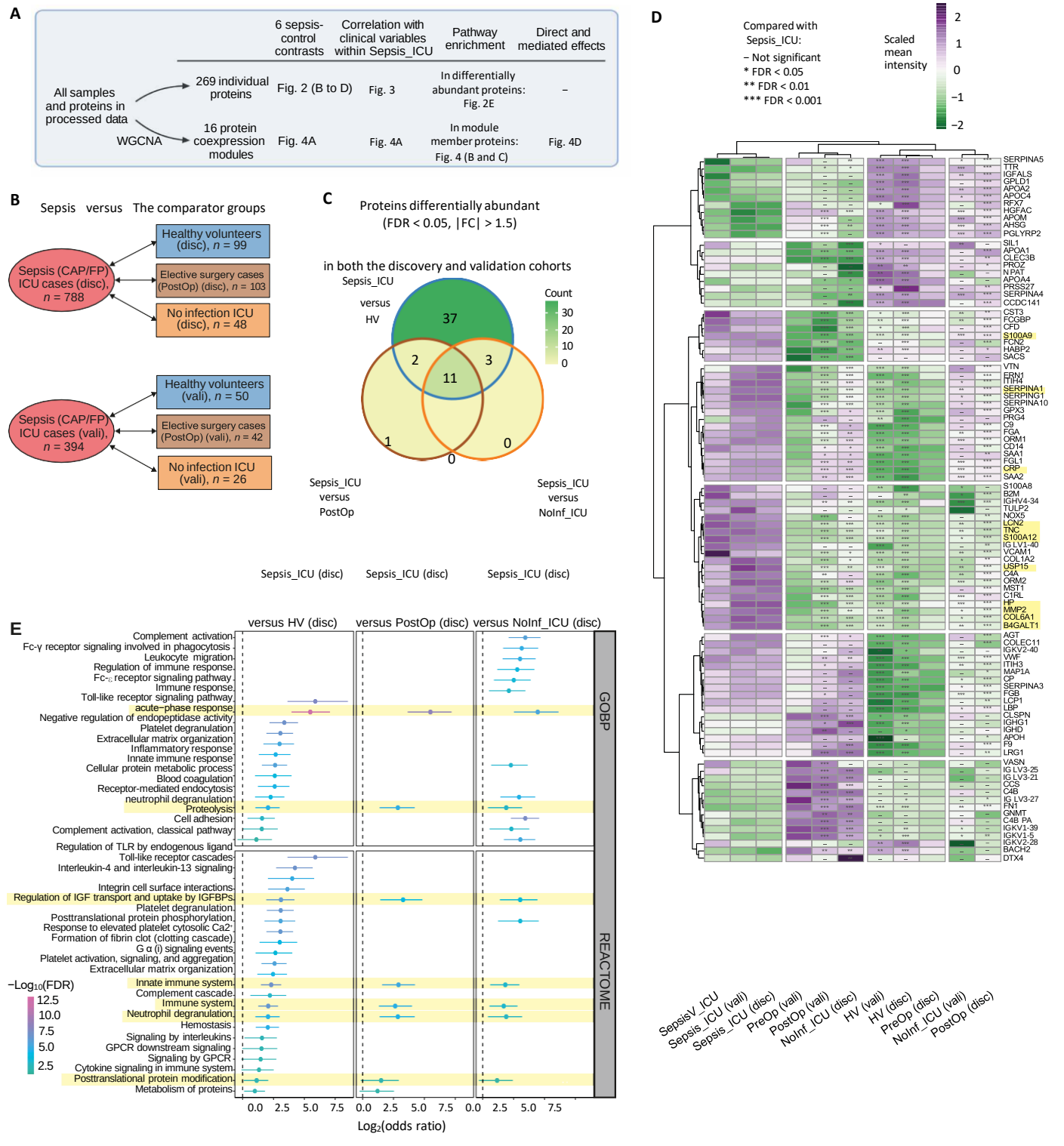


Fig. 2. Sepsis-specific proteomic response. (**A**) Schematic diagram of analyses shown in Figs. 2 to 4. (**B**) contrasts made between sepsis and comparator groups. *n*, number of individuals. (**C**) Venn diagram of differentially abundant (Fdr < 0.05 and |Fc| > 1.5) proteins overlapping between contrasts. (**D**) Summary heatmap of mean protein abundance in sepsis and comparator groups, scaled by row. a total of 94 proteins differentially abundant in any of the six contrasts in (a) are included, with Fdr thresholds shown; 11 proteins differentially abundant in all contrasts shaded yellow. only the first available samples per patient with sepsis were included. (**E**) Pathway enrichment of differentially abundant proteins. Terms significantly enriched in all discovery cohort contrasts (Fdr < 0.05) shaded in yellow. horizontal bars indicate 95% cis of $\log_2$(odds ratio). GoBP, gene ontology biological process.

More specifically, compared with HV, we found that 53 proteins were differentially abundant [false discovery rate (FDR) < 0.05, fold

change (|FC|) > 1.5] in both the discovery and validation cohorts (Fig. 2C and fig. S2B). Proteins more abundant in sepsis were implicated in the acute-phase response (CRP, SAA1, and SAA2), coagulation process [VWF (Von Willebrand factor), FGB (fibrinogen beta chain), and FGA], and immune or immune-regulatory functions [LBP, S100A9, FGL1 (fibrinogen-like protein 1), ORM1 (alpha-1- acid glycoprotein 1), and CD14 (cluster of differentiation 14, mono- cyte differentiation antigen], whereas abundances of apolipoproteins, α-2-HS-glycoprotein, hepatocyte growth factor activator, plasma serine protease inhibitor (SERPINA5), TTR (transthyretin inhibited

by inflammation), and transcription regulator protein BACH2 (regulates apoptosis and adaptive immunity) were reduced.

Comparing sepsis against sterile inflammation among postoperative samples from elective surgery or among noninfected ICU patients, we identified 14 and 14 proteins as differentially abundant in both the discovery and validation sepsis cohorts, respectively (Fig. 2C and fig. S2B). These included HP, TNC, B4GALT1, and S100A12 (versus postoperative surgery) with FCN2 (ficolin 2) not seen in sepsis versus HV contrast and S100A9, HP, SERPINA1, and B4GALT1 (versus noninfected ICU cases). Among the 11 proteins differentially abundant in all six sepsis-comparator group contrasts, CRP and S100A9 were also identified in the sterile surgery response comparing post- against preoperative samples, with the remaining nine proteins reflecting a more sepsis-specific response.

We then identified biological pathways consistently enriched in the sepsis contrasts. In the discovery cohort, sepsis differed from all comparator groups in acute-phase response, neutrophil degranula- tion, regulation of insulin-like growth factor (IGF) transport and uptake by IGF binding proteins, innate immune system, and post- translational protein modification (Fig. 2E). Immune and metabolic processes that differed in sepsis versus HV but were not different between sepsis and postoperative patients included Toll-like recep- tor signaling, clotting, and interleukin-4 (IL-4) and IL-13 signaling. These enriched terms were replicated in the validation cohort.

Specific plasma protein subsets associate with sepsis severity, clinical covariates, source, and progression

We then investigated whether specific plasma proteins were associated with particular clinical features of the sepsis response, combining samples from the sepsis discovery and validation cohorts. We first analyzed overall variance in the proteome within patients with sepsis. The largest component of variance, PC1, showed significant (FDR < 0.05) positive correlations with features relating to illness severity (fig. S2, C and D). In terms of individual plasma protein abundance, we identified a protein set [including PTGDS (prostaglandin D2 synthase), B2M (beta 2 microglobulin), CFD (complement factor D), LCN2, VWF, COL6A1, USP15, MMP2, COL1A2, CD14, PLTP (phospholipid transfer protein), and CRP] highly correlated with clinical variables reflecting more severe illness, including total SOFA, Acute Physiology And Chronic Health Evaluation (APACHE), occurrence of shock or renal failure, and prothrombin time; whereas a second set [including SERPIND1, C3 (complement component 3), APOA1 (apolipoprotein A-I), HRG (histidine-rich glycoprotein), KNG1 (kininogen-1), and VTN (vitronectin)] had strong negative associations (Fig. 3). We found five proteins (CRP, LCN2, USP15, COL1A2, and MMP2) that were significantly more abundant (FDR < 0.05 and FC > 1.5) in patients with FP compared with CAP (fig. S3A). Within CAP, LCN2, which limits bacterial growth by sequestering iron-containing siderophores (*28*), showed higher abundance in bacterial compared with viral infections (*28*).

Using the protein coexpression modules identified from all cohorts, we found that specific modules were significantly (FDR < 0.05) correlated with comparator group contrasts and with specific clinical variables (Figs. 2A and 4, A to C). For example, the blue module, enriched for acute-phase response proteins and positively correlated with sepsis in all comparator contrasts but that did not associate with mortality, showed modest association with organ dysfunction and the strongest association with high temperature. The tan module, containing S100 family proteins and enriched for neutrophil degranulation, showed a

positive correlation with sepsis, sepsis severity, and acute respiratory distress syndrome. The green-yellow module comprising many ECM proteins also positively correlated with sepsis and showed a stronger positive correlation with severity, together with renal impairment, lym- phopenia, low temperature, increased mortality, and greater age. The red module (complement activation) was associated with sepsis but with less severe disease. Lipoprotein metabolic processes, reflected by the yellow module, were negatively correlated with sepsis and with severity and mortality of the patients with sepsis.

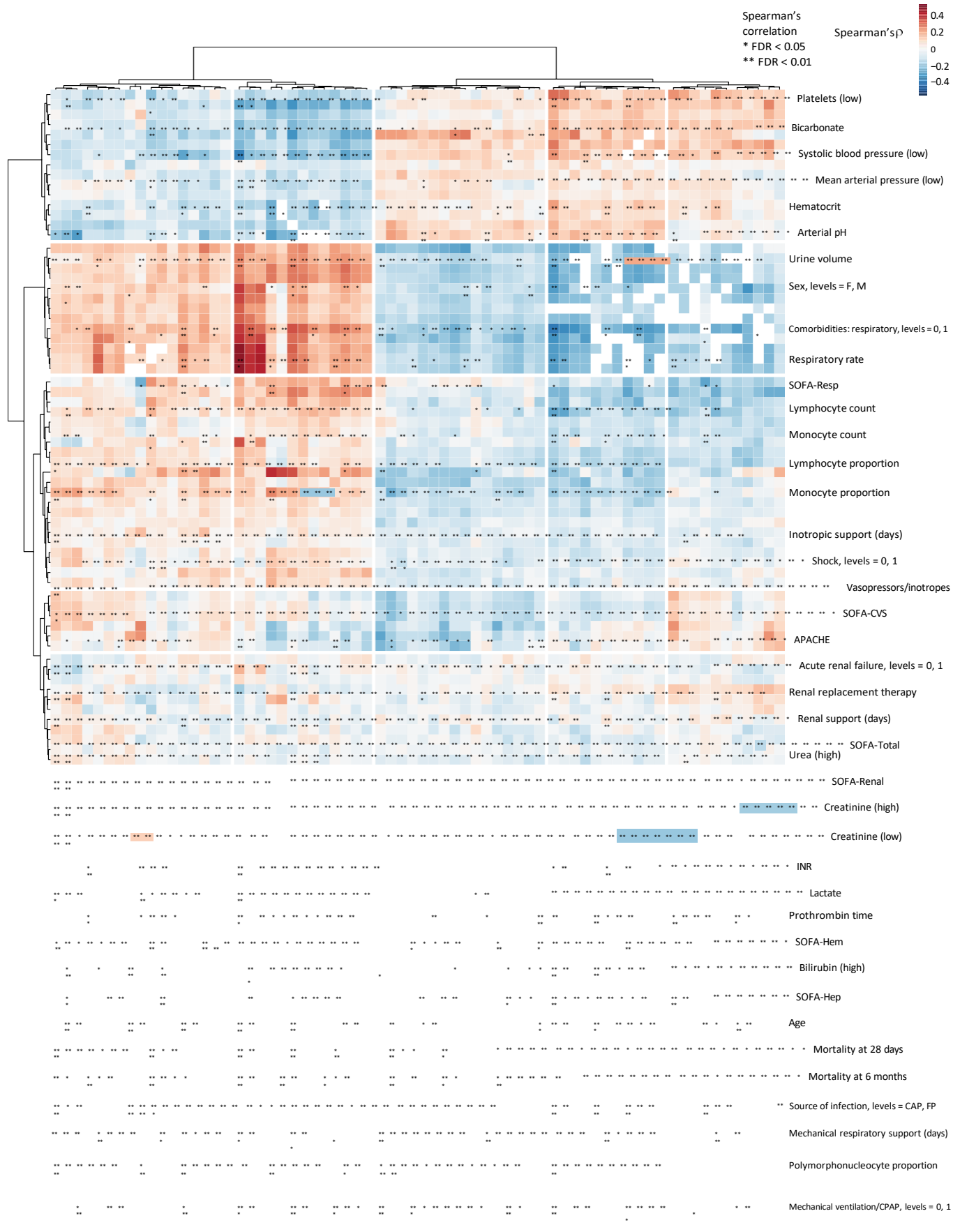
For coexpression modules that correlated with both 28-day mor- tality and SOFA scores, we applied mediation analysis to understand the potential causal relationships between these proteomic features and patient outcomes and to determine whether these influences are exerted through the dysfunction of specific organs. Using linear me- diator models and generalized linear outcome models, we tested the influence of each of the coexpression modules on mortality, medi- ated by either the total SOFA or individual organ scores.

This analysis showed that the effect of proteomic features on mortality is, in many cases, mediated by organ dysfunction, with six of the eight modules tested influencing the outcome both directly and indirectly through at least five of the organ system SOFA scores, involving hemostasis, negative regulation of endopeptidase activity, lipoprotein assembly, immunoglobulins, and neutrophil degranula- tion (Fig. 4D and fig. S3B). Other modules showed a more organ- specific effect. For example, the effect of more severely depleted complement (red module) on increased mortality was mediated by alternations in renal and neurological dysfunction (as measured by respective SOFA scores) but not the function of other organs.

Direct effects were stronger than the mediated effects except for models testing the total SOFA score or the ECM proteins (green- yellow module). This observation is consistent with individual organs mediating part of the effect and the rest being driven by direct mech- anisms and with the total SOFA score representing the summed ef- fect of multiple organ failure, which mediates a large proportion of the proteomic effect on mortality. ECM proteins showed a strong correlation with the cardiovascular and renal SOFA components (Fig. 4A), which corresponded to larger mediated effects on mortal- ity for these two organs than for direct effects. On the other hand, the analysis indicates that the positive effects of neutrophil degranu- lation proteins (tan module) and the negative effect of proteins regu- lating hemostasis (turquoise) on outcome are less mediated through particular organ dysfunction.

Among patients with sepsis who had serial samples, 12 of 16 coex- pression modules showed a change between ICU admission, day 3, and/ or day 5 using paired samples (fig. S3C and table S1). The blue (acute- phase), red (complement activation), and green-yellow (ECM) modules showed a consistent decrease over time from admission, and the change in green-yellow module from day 1 to day 5 positively correlated with the change in total SOFA score (FDR = 0.0005, $\rho = 0.33$ Spearman).

To further understand the extent to which organ failure and tissue damage contributed to particular circulating proteins in septic plasma, we used the protein tissue origins inferred by Malmström *et al.* (*29*) from constructing a murine protein tissue atlas. Within the 153 homologous proteins that overlapped between our processed data and their plasma dataset, most (n = 118) were predicted to originate from plasma, encom- passing many classical plasma proteins (fig. S3D). There were 23 non- plasma proteins with an even distribution across the blood cell types or other heavily vascularized organs (denoted as "Common") and a small number (n = 12) showing a more specific origin. Some of these proteins


Spearman's correlation
* FDR < 0.05
** FDR < 0.01
Spearman's ρ
0.4
0.2
0
−0.2
−0.4
Platelets (low)
Bicarbonate
Systolic blood pressure (low)
Mean arterial pressure (low)
Hematocrit
Arterial pH
Urine volume
Sex, levels = F, M
Comorbidities: respiratory, levels = 0, 1
Respiratory rate
SOFA-Resp
Lymphocyte count
Monocyte count
Lymphocyte proportion
Monocyte proportion
Inotropic support (days)
Shock, levels = 0, 1
Vasopressors/inotropes
SOFA-CVS
APACHE
Acute renal failure, levels = 0, 1
Renal replacement therapy
Renal support (days)
SOFA-Total
Urea (high)
SOFA-Renal
Creatinine (high)
Creatinine (low)
INR
Lactate
Prothrombin time
SOFA-Hem
Bilirubin (high)
SOFA-Hep
Age
Mortality at 28 days
Mortality at 6 months
Source of infection, levels = CAP, FP
Mechanical respiratory support (days)
Polymorphonucleocyte proportion
Mechanical ventilation/CPAP, levels = 0, 1

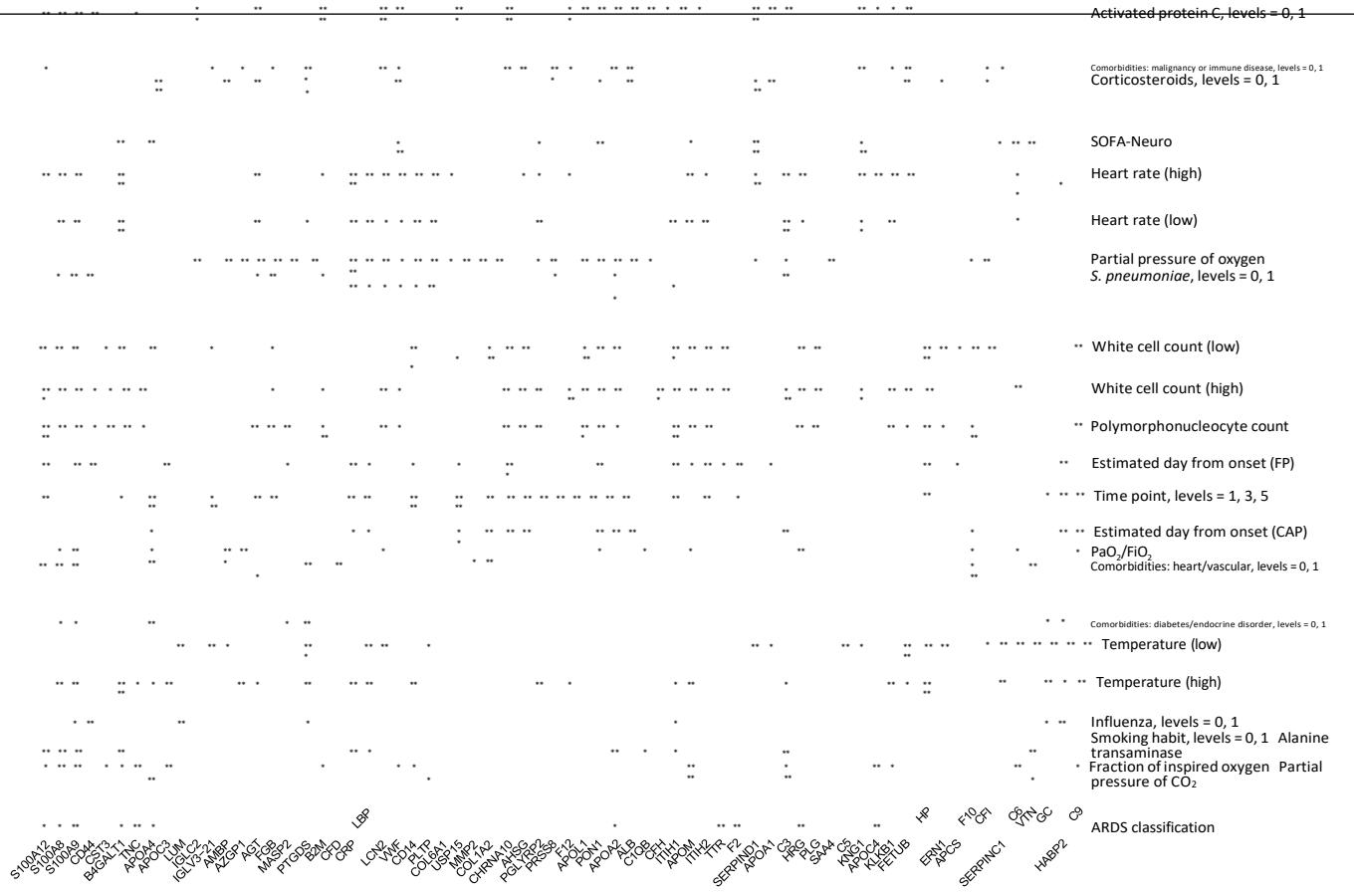


Fig. 3. Correlation of sepsis plasma proteome with clinical characteristics. heatmap of correlations between 66 clinical characteristics (data file S1) and protein abun- dance (*n* = 68 proteins shown), using the first sample per patient with sepsis (*n* = 1182). 0/1 indicates absence/presence of trait. Samples from the sepsis discovery and validation cohorts were combined for this analysis.

with a nonplasma tissue/cell specificity, including PTGDS, COL6A1, TTR, LUM (lumican), and S100A8, were among the proteins strongly correlated with sepsis severity and mortality, suggesting greater cell necrosis and tissue damage in the more severely ill patients.

Sepsis subphenotypes can be identified from the plasma proteome

We next investigated whether the plasma proteome was informative for defining sepsis subphenotypes. Consensus clustering on the pro- tein intensities for the sepsis discovery cohort using all time points identified three subgroups, which we denote as sepsis plasma proteome- based clusters (SPC1/2/3). These represented the optimal cluster stability and number shown by cumulative distribution of the con- sensus index and were independent of relatedness between serial samples (Fig. 5, A and B, and fig. S4, A and B).

Patients in SPC1 had more severe illness (reflected in SOFA scores and occurrence of shock and renal failure; Fig. 5C and ta- ble S2) and significantly higher mortality than those in SPC2 and SPC3 at both 28 days [SPC1 versus 2+3 hazard ratio (HR) (95% CI) = 2.5 (1.7 to 3.7), $P = 1.3 \times 10^{-6}$; fig. S4C and table S3] and 6 months [HR = 2.3 (1.7 to 3.2), $P = 5.4 \times 10^{-7}$; Fig. 5D] after sampling. Patients in SPC3 were younger than those in the other two clusters, and those in SPC2 had lower APACHE scores and intermediate lymphocyte and monocyte counts among the clusters (fig. S5A and summary statistics provided in table S2). SPC1 was enriched for patients with FP and earlier time point samples. Patients with CAP who clustered as SPC1 had worse respiratory function and required more respiratory support than those who clustered as SPC2 or SPC3, indicating that SPC1 identifies more severely ill patients, including after accounting for the original source of sepsis.

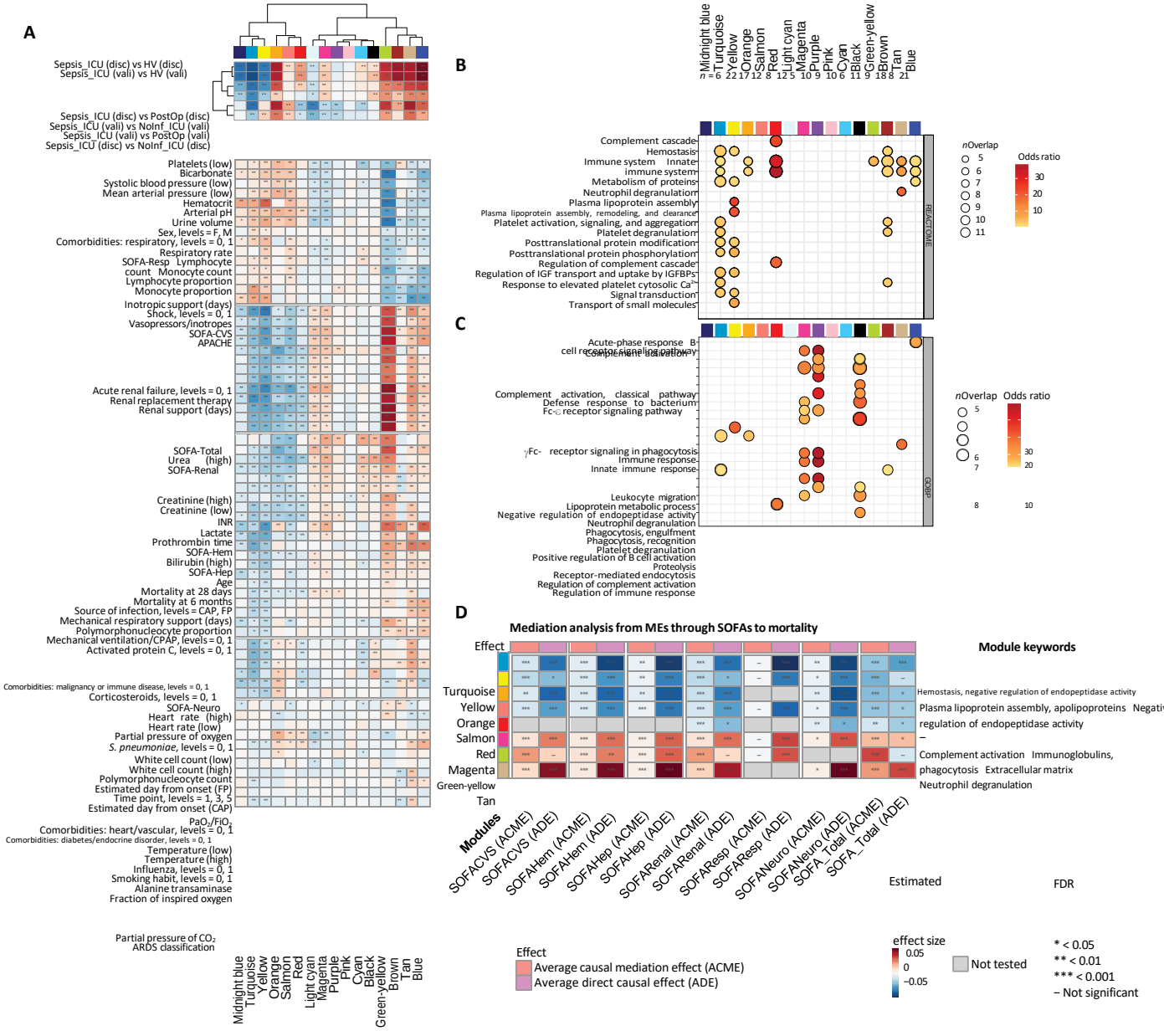


Fig. 4. Variation within sepsis plasma proteome. (**A**) heatmap of Spearman's correlation between coexpression module eigengenes, group contrasts, and clinical vari- ables using first sample per patient with sepsis (*n* = 1182). row order aligned to Fig. 3. (**B** and **C**) Balloon plots of pathway enrichment for module member proteins (module size shown as *n*) using reactome (B) and GoBP (c) annotations. order of coexpression modules aligned to (a). (**D**) Summary heatmap of mediation analysis test- ing SoFa scores as mediators and coexpression module eigengenes (Mes) as independent variables. SoFa-Me pairs without a significant (Fdr < 0.05) correlation in (B) were not tested for the causal mediation effect and grayed out.

Overall, more plasma proteins were differentially abundant when comparing SPC1 than SPC2 or SPC3 to HV (81, 74, and 55 proteins for SPC1, SPC2, and SPC3 patients with sepsis, respectively), in- cluding immunoglobulins and apolipoproteins specific to SPC1, en- richment for phagocytosis and positive regulation of B cell activation in SPC1, and lower abundances of immunoglobulins in SPC2 (fig. S5, B and C). Comparing SPC1 with SPC2 or with SPC3, we observed relatively greater activity of immune pathways, including interleu- kin signaling, Fc-γ or Fc-ε receptor signaling, leukocyte migration, complement activation, and ECM organization, and lower activity of lipoprotein metabolic processes (Fig. 5, E and F).

Proteomic patient subgroups are reproducible and involve specific pathways and biomarkers

To validate and further characterize these subgroups, we developed SPC prediction models based on the sepsis discovery cohort (fig. S6A). An elastic net model with 181 predictors selected from 269 input

candidates performed best in terms of test set accuracy [91.4%, test set *n* = 244, for each SPC area under the receiver operating charac- teristic (AUROC) curve ≥95%, sensitivity and specificity ≥85%; fig. S6, B and C]. We applied this model to the sepsis validation co- hort (*n* = 624 samples from 394 patients) to derive the SPC assign- ments, which replicated the associations with mortality (Fig. 6A, fig. S7A, and table S3) and measures of severity including lactate, cell counts, vasopressor and renal support, and SOFA scores (Fig. 6B and fig. S7B). Differential abundance and pathway enrichment anal- ysis between the validation cohort clusters and HV showed strong concordance with the discovery cohort (fig. S7, C to E). We further tested cluster prediction based on a small number of informative protein biomarkers. We derived a new minimal elastic net model with eight predictors [USP15, COL1A2, APOA2, MAP1A (microtubule-associated protein 1A), GNMT (glycine N- methyltransferase), TSPAN11 (tetraspanin-11), LCP1 (lymphocyte cytosolic protein 1, L-plastin), and ALB (albumin)], which successfully

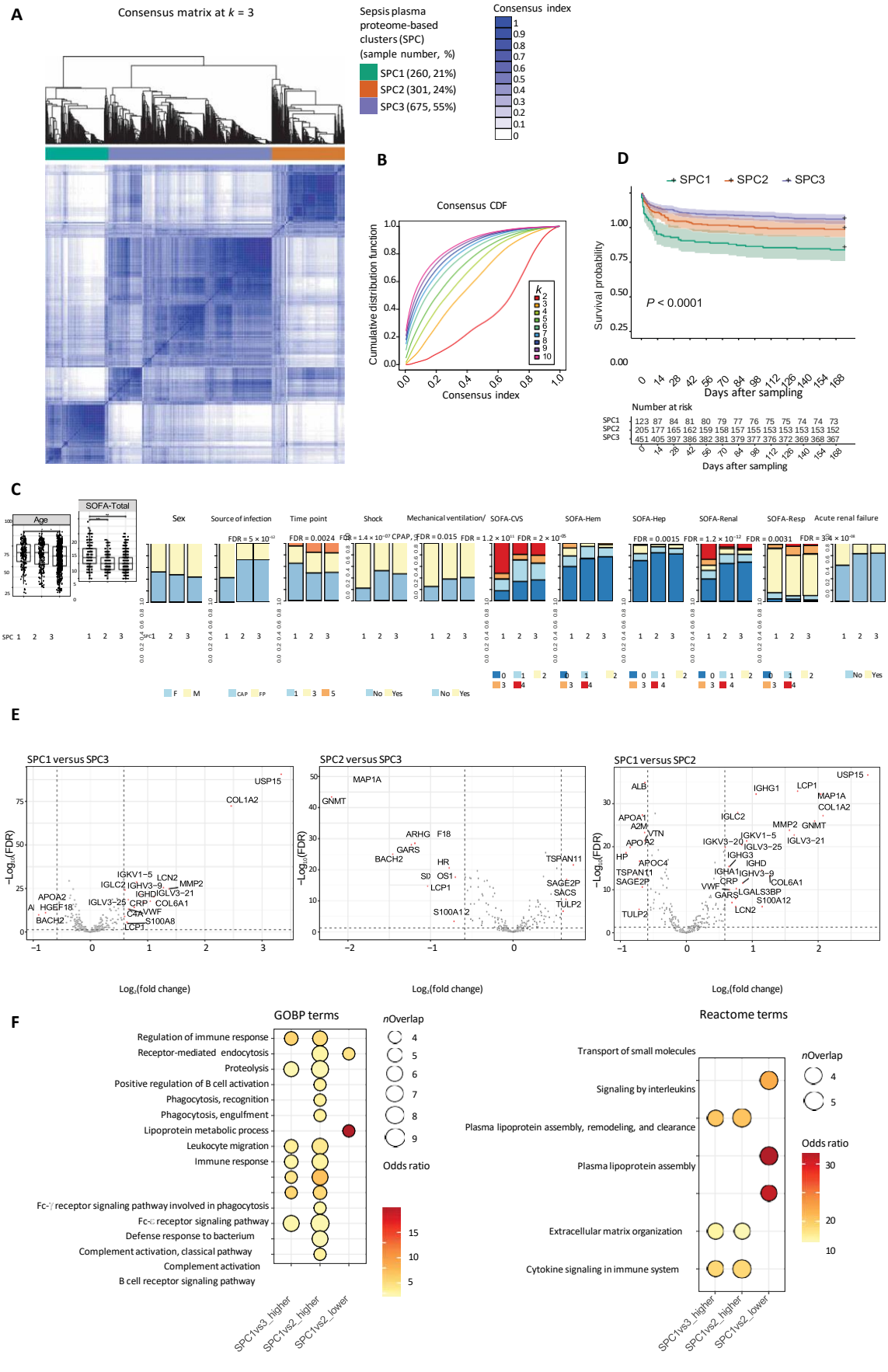

A
Consensus matrix at k = 3
Sepsis plasma proteome-based clusters (SPC) (sample number, %)
SPC1 (260, 21%)
SPC2 (301, 24%)
SPC3 (675, 55%)
Consensus index
B
Consensus CDF
Cumulative distribution function
Consensus index
D
SPC1 SPC2 SPC3
Survival probability
P < 0.0001
Days after sampling
Number at risk
C
Age
SOFA-Total
Sex
Source of infection
Time point
Shock
Mechanical ventilation/ CPAP
SOFA-CVS
SOFA-Hem
SOFA-Hep
SOFA-Renal
SOFA-Resp
Acute renal failure
E
SPC1 versus SPC3
SPC2 versus SPC3
SPC1 versus SPC2
Log2(fold change)
F
GOBP terms
Regulation of immune response
Receptor-mediated endocytosis
Proteolysis
Positive regulation of B cell activation
Phagocytosis, recognition
Phagocytosis, engulfment
Lipoprotein metabolic process
Leukocyte migration
Immune response
Fc-γ receptor signaling pathway involved in phagocytosis
Fc-ε receptor signaling pathway
Defense response to bacterium
Complement activation, classical pathway
Complement activation
B cell receptor signaling pathway
nOverlap
Odds ratio
SPC1vs3_higher
SPC1vs2_higher
SPC1vs2_lower
Reactome terms
Transport of small molecules
Signaling by interleukins
Plasma lipoprotein assembly, remodeling, and clearance
Plasma lipoprotein assembly
Extracellular matrix organization
Cytokine signaling in immune system

Fig. 5. SPCs in discovery cohort. (**A** and **B**) consensus clustering (1236 samples from *n* = 788 patients). (a) cluster dendrogram and heatmap of consensus index (fre- quency of the sample pair being in the same cluster) for three clusters with % samples for each SPc noted. (B) cumulative distribution function (cdF) curves of the con- sensus index for increasing cluster number (*k*). (**C**) Bar plots of representative clinical variables between SPcs (*x* axis). Bar height represents proportion of each value of the variable. age: SPc1 median 67 years (iQr, 56 to 77), SPc2 67 (53 to 77), SPc3 63 (51 to 74) (Kruskal-Wallis Fdr = 0.014). Total SoFa score: median SPc1 7 (iQr, 5 to 10),
SPc2 = SPc3 5 (3 to 7), dunn's Fdr SPc1 versus 2 = 2.3×10^{-10}, SPc1 versus 3 = 1.2×10^{-11}. Shock: SPc1 = 77%, SPc2 = 49%, SPc3 = 54% (χ^2 Fdr = 1.4×10^{-7}). renal failure: SPc1 = 38%, SPc2 = 17%, SPc3 = 16% (χ^2 Fdr = 3.4×10^{-8}). Proportion FP (of FP+caP): SPc1 = 59%, SPc2 = 28%, SPc3 = 27% (χ^2 Fdr = 5.0×10^{-12}). (**D**) Kaplan-
Meier survival curves by SPc at 6 months after sampling. For patients with multiple samples (day 1/3/5), cluster assignment from the latest available sample was used. Global *P* values from log-rank tests; shading, 95% cis. (**E**) differential protein abundance between SPcs. (**F**) GoBP and reactome terms enriched (Fdr < 0.05) in differen- tially abundant proteins. only contrasts with enriched terms detected are shown.

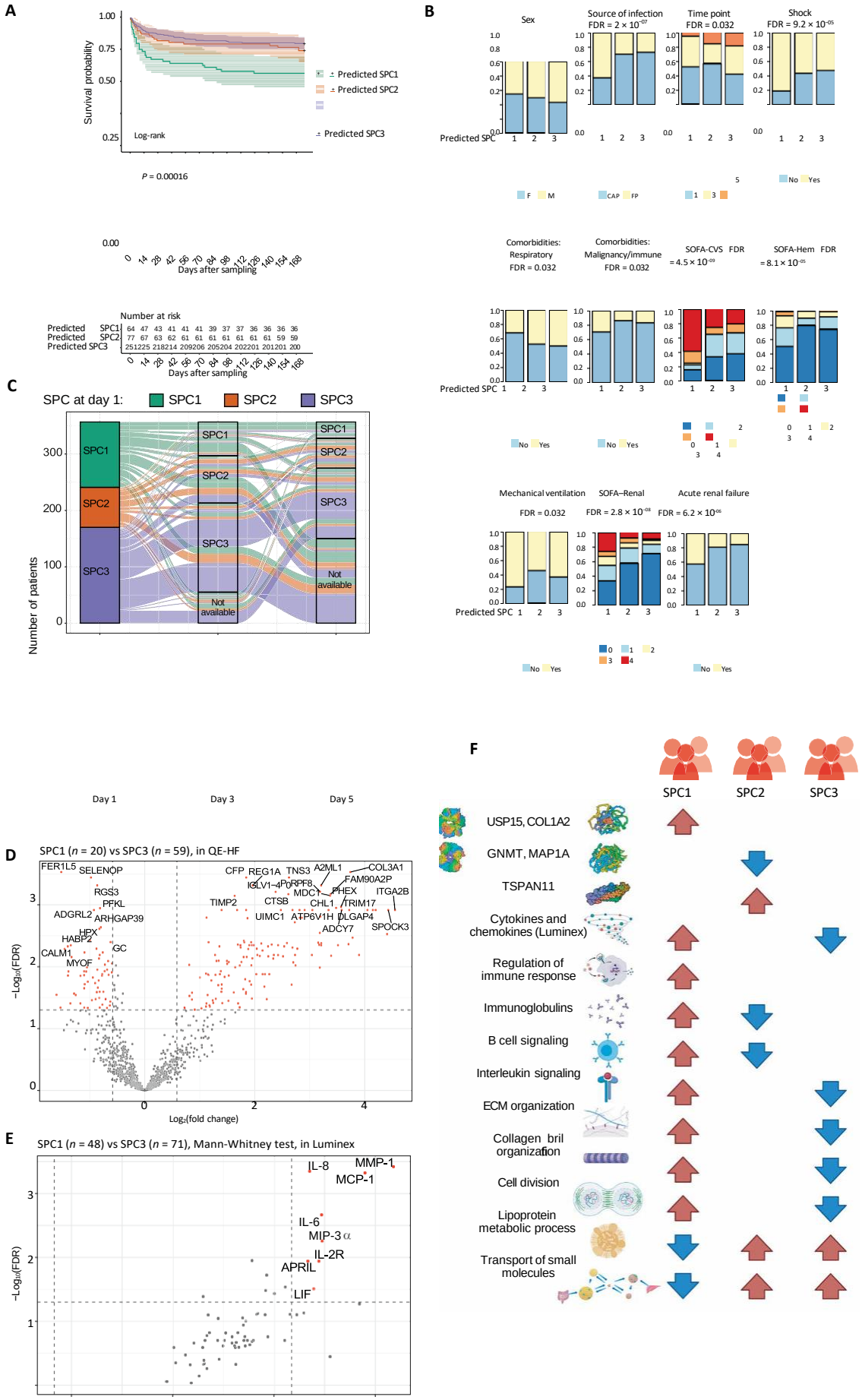

A
Survival probability
1.00
0.75
0.50
0.25
Log-rank
P = 0.00016
Predicted SPC1
Predicted SPC2
Predicted SPC3
0.00
0 14 28 42 56 70 84 98 112 126 140 154 168
Days after sampling
Number at risk
Predicted SPC1 64 47 43 41 41 41 39 37 37 36 36 36 36
Predicted SPC2 77 67 63 62 61 61 61 61 61 61 61 59 59
Predicted SPC3 251225 218214 209206 205204 202201 201201 200
0 14 28 42 56 70 84 98 112 126 140 154 168
Days after sampling
B
Sex
Source of infection
FDR = 2 × 10-07
Time point
FDR = 0.032
Shock
FDR = 9.2 × 10-05
Predicted SPC
F M
CAP FP
1 3 5
No Yes
Comorbidities: Respiratory
FDR = 0.032
Comorbidities: Malignancy/immune
FDR = 0.032
SOFA-CVS FDR
= 4.5 × 10-09
SOFA-Hem FDR
= 8.1 × 10-05
Mechanical ventilation
FDR = 0.032
SOFA–Renal
FDR = 2.8 × 10-08
Acute renal failure
FDR = 6.2 × 10-06
C
SPC at day 1:
SPC1
SPC2
SPC3
Number of patients
300
200
100
0
Day 1
Day 3
Day 5
Not available
D
SPC1 (n = 20) vs SPC3 (n = 59), in QE-HF
FER1L5
SELENOP
RGS3
PFKL
ADGRL2
ARHGAP39
HPX
HABP2
CALM1
MYOF
GC
CFP
REG1A
TNS3
A2ML1
COL3A1
IGLV1-40
PPP1R8
FAM90A2P
MDC1
PHEX
ITGA2B
TIMP2
CTSB
CHL1
TRIM17
UIMC1
ATP6V1H
DLGAP4
ADCY7
SPOCK3
−Log10(FDR)
Log2(fold change)
E
SPC1 (n = 48) vs SPC3 (n = 71), Mann-Whitney test, in Luminex
IL-8
MMP-1
MCP-1
IL-6
MIP-3α
IL-2R
APRIL
LIF
−Log10(FDR)
F
SPC1
SPC2
SPC3
USP15, COL1A2
GNMT, MAP1A
TSPAN11
Cytokines and chemokines (Luminex)
Regulation of immune response
Immunoglobulins
B cell signaling
Interleukin signaling
ECM organization
Collagen bril organization
Cell division
Lipoprotein metabolic process
Transport of small molecules

Fig. 6. Validation and molecular characteristics of sepsis plasma proteome-based clusters (SPCs). (**A** and **B**) Sepsis validation cohort. (a) Kaplan-Meier survival curves (*n* = 392) at 6 months after sampling. For each patient, cluster assignment from latest available sample were used. (B) Bar plots comparing categorical clinical variables between SPcs. (**C**) SPc movement days 1 to 5 of icU admission (*n* = 346 patients with sepsis with a day 1 and at least one subsequent sample available, discovery and validation cohorts combined). Flow widths are proportional to number of patients with the corresponding SPc transition and are color coded according to starting SPc. (**D** and **E**) differential abundance analysis between SPcs using a Qe-hF mass spectrometer (d) or a luminex immunoassay (e) in subsets of sepsis samples. (**F**) Sum- mary of molecular characteristics for each SPc (see table S6). red and blue arrows indicate higher or lower abundance of the corresponding proteins, respectively.

classified 79.5% of the discovery cohort test set, with a 72.7% sensitivity and 94.2% specificity for SPC1. In the validation cohort, this prediction showed good consistency with the 181-predictor model, with an 82.5% overlapping assignment and AUROCs > 87% (fig. S6C). We also investigated patient transitions between clusters. Cluster membership was associated with the length of time after ICU ad- mission (χ^2 P = 0.0012). Analyzing 526 patients with multiple time points available, 57.4% of patients changed group over time, most frequently from SPC1 to SPC3, consistent with a general trajectory of recovery (Fig. 6C and table S4).

To more fully characterize the plasma proteome in these clusters, we first profiled a subset of the sepsis samples composed of 148 sam- ples from 100 patients on a Q-Exactive high field (QE-HF) mass spectrometer after depleting 12 highly abundant proteins. This per- mitted measurement of many more proteins (1123 detected in ≥70% samples) but has potential limitations associated with the depletion process. We identified 144 proteins with higher and 63 proteins with lower abundance in SPC1 versus SPC3 and no signal in the two con- trasts with SPC2 (Fig. 6D) while recognizing the reduced power to detect differences with smaller sample numbers. Pathway enrich- ment analysis again highlighted immune response pathways, ECM organization, and lipoprotein metabolism differentiating SPC1 and SPC3, along with IL-4 and IL-13 signaling, collagen organization, and the cell cycle. Second, we analyzed 65 cytokines and other signal- ing molecules in 204 samples from 146 patients with SPC assign- ments, assayed using the Luminex immunoassay. We found greater activity in chemotaxis and IL-6 regulated pathways in SPC1 versus SPC3 with significantly increased (FDR < 0.05 and FC > 1.5) chemo- kines MCP-1 (monocyte chemoattractant protein-1), IL-8 (interleu- kin 8), and MIP-3α (macrophage inflammatory protein 3 alpha); cytokines involved in B cell proliferation [APRIL (tumor necrosis factor ligand superfamily member 13) and IL-6 (interleukin 6)] and immune in- hibitory functions [IL-2R (interleukin 2 receptor) and LIF (leukemia inhibitory factor)]; and the interstitial collagenase MMP-1 (matrix metalloproteinase-1) (Fig. 6E and fig. S7F).

Overall, SPC1 was characterized by higher abundance in plasma of immune response proteins, including specific cytokines and immunoglobulins, and more collagen and ECM components, implying a greater degree of tissue damage in these patients (Fig. 6F and table S5). Lipoprotein metabolism and transport were comparatively down-regulated in SPC1. SPC2, by contrast, had lower immunoglobulin abundances and B cell signaling pathway proteins, whereas in SPC3, interleukin signaling and cytokine concentrations were relatively reduced.

Integration of the plasma proteome and leukocyte transcriptome reveals components contributing to the sepsis response

We next sought to maximize the informativeness of the sepsis plasma proteomics (MS) by integrating with paired white blood cell (WBC) transcriptomics [RNA sequencing (RNA-seq)] for 837 samples (649 patients) using matrix decomposition (*30*). We identified 284 latent components, each comprising vectors of scores (loadings) that indicate the contribution of individual proteins or genes linked by that component. We proceeded to identify which components showed association with disease severity, source of sepsis, clinical covariates, and disease subphenotypes, focusing on the 76 components with significant contributions from proteins (based on posterior inclusion probability > 0.5) (Fig. 7, fig. S8, and data file S2).

The component showing the strongest disease severity association (component 141) involving contributions from the plasma proteome linked less severe disease (lower total SOFA score) with proteins im- plicated in lipid biology [APOA1 and PON1 (paraoxonase 1)] and HRG; differential expression of genes enriched for human leukocyte antigen class II; and overall pathway enrichment for negative regula- tion of endopeptidase activity, platelet degranulation, and regulation of complement activation (Fig. 7A). Components 187 and 164, in- volving WBC transcriptomic differences in metabolic and immune processes, also strongly associated with disease severity (fig. S8, A and B, and data file S2).

Components 266 and 134 correlated with the source of sepsis (CAP versus FP) (Fig. 7B, fig. S8C, and data file S2). The most statistically correlated component involving plasma proteins (component 266) linked contributions from genes and cognate proteins for mul- tiple immunoglobulin variable and constant chains and showed en- richment for receptor-mediated endocytosis (FDR = 0.016) in FP (Fig. 7B). We also identified correlations with time from ICU admis- sion (Fig. 7C, fig. S8, D and E, components 241, 106, and 174) in- cluding a component (241) that involved proteins PRSS8 (serine protease-8), CD5L (CD5-antigen like; involved in lipid synthesis and macrophage apoptosis), and FN1 (fibronectin; involved in cell adhesion and motility) and regulation of chemotaxis and signaling pathways indicated by differential gene expression (data file S2).

The components showing the strongest association with SPCs (242, 133, and 204) all involved protein abundances only and implicated proteins enriched for ECM and metabolism (USP15, COL1A2, MMP2, and VWF), complement, and immunoglobulin variable chains (Fig. 7D and data file S2). We previously reported WBC transcriptome–derived sepsis response signatures (SRS) associated with differential outcome and response to therapy (*10*, *31*), includ- ing patients with the SRS signature 1 (SRS1) who show granulopoi- etic dysfunction, relative immune compromise, and high mortality (*12*, *13*). Most components associated with SRS did not include con- tributions from proteins (fig. S8, F and G, components 92 and 232), but one component (160) demonstrated that SRS1 was associated with contributions from proteins including COLEC11 (collectin subfamily member 11; role in innate immunity and apoptosis), CRP, DEFA1 (human alpha defensin 1), LBP, ECPAS (Ecm29 proteasome adapter and scaffold protein), and CPN2 (carboxypeptidase N sub- unit 2), along with genes enriched for secreted soluble factors, G protein–coupled receptor ligand binding, neutrophil degranulation, and immunoregulatory interactions (Fig. 7E).

Transcriptomic and proteomic profiling reveal complementary but distinct sepsis subphenotypes and response states

We further explored the relationship between plasma proteome- and leukocyte transcriptome–derived sepsis subphenotypes by analyz- ing 1016 patients (1361 samples) with both SPC and SRS assignments. Considering the first available time points, we found that 70% of SPC1 patients were also assigned to SRS1 in the discovery cohort, compared with 37 and 34% in SPC2 and SPC3, respectively (71, 48, and 31% in validation cohort) (χ^2 $P < 0.0001$; Fig. 8, A and B). There was great- er likelihood of transition from SPC1 or SRS1 to another state than in the opposite directions (fig. S9, A and B).

We identified differentially abundant proteins between SRS groups (Fig. 8C), some of which overlapped with the proteins discriminating SPC1 from SPC2+3 (Fig. 8D), including higher abundance of CRP, LCN2, USP15, COL1A2, SAA2, MMP2, S100A8, TNC, and S100A12 in both SRS1 and SPC1. On the other hand, a set of immunoglobulins, HP and APOA2, differed only between SPCs and SAA2 only in the SRS contrast. Gene expression differences between SRS groups and SPC groups were strongly correlated (fig. S9C). Pathways enriched in the differential proteins and genes showed shared and specific features (table S6). These included cytokine signaling and innate immunity inferred from higher-abundance proteins in both SRS1 and SPC1, together with neutrophil degranulation and oxidation-reduction (up-regulated) and adaptive immune response and T cell costimulation (down-regulated) in both SRS1 and SPC1 from gene expression analysis. Differences included major histocompatibility complex class II genes down-regulated uniquely in SRS1 and interferon signaling and cell division terms only enriched in the SPC analysis (table S6).

Lastly, given that SRS1 and SPC1 both associated with poor outcome (Fig. 8E), we tested whether the two classifications can be combined to further inform risk stratification. We found that the patients assigned to both SRS1 and SPC1 (~11% patients) had the highest mor- tality rate of 33.3% at 28 days (31.7% in validation cohort), HR = 3.9 (95% CI 2.3 to 6.7), $P < 0.0001$ (discovery) [HR = 3.0 (1.5 to 6.0) $P = 0.002$ (validation)] versus SPC3 non-SRS1 patients (~43%) who had the lowest mortality of 10.4% (12.8% in validation cohort) (Fig. 8F and fig. S9, D and E).

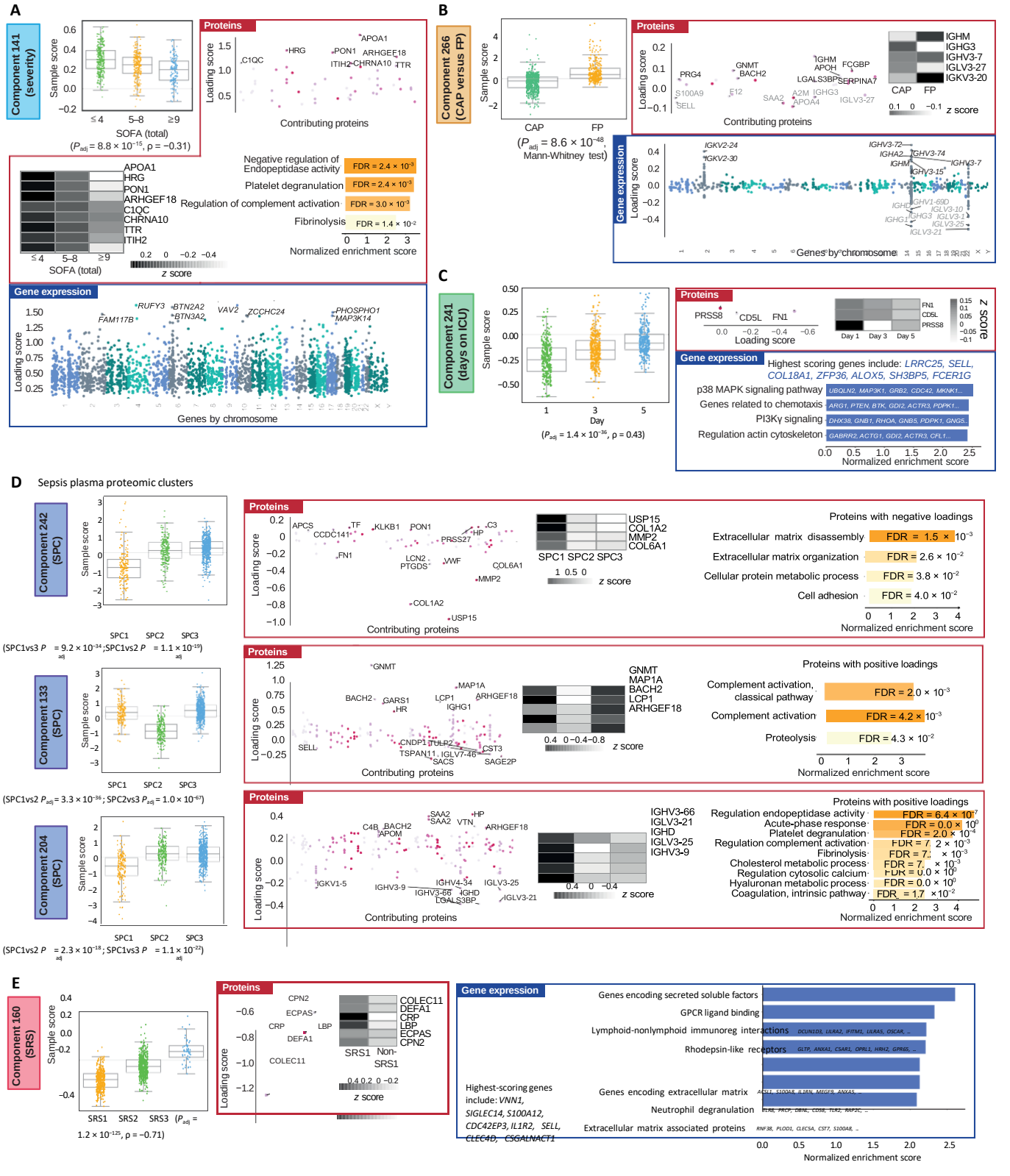
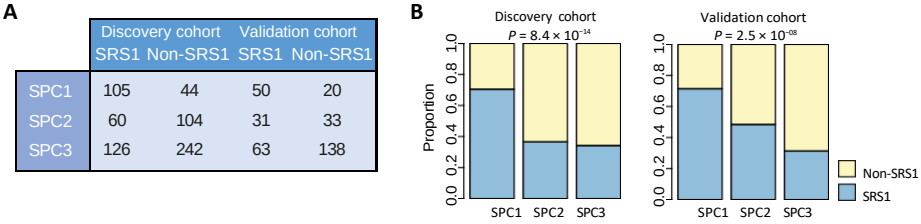


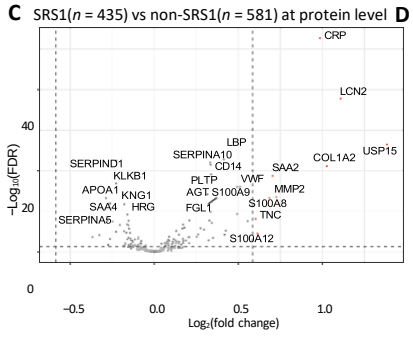
Fig. 7. Integration of plasma proteomics and leukocyte transcriptomics. Matrix decomposition of proteomic and transcriptomic data (837 samples from *n* = 649 pa- tients with sepsis). components that included protein contributions and were most statistically correlated and biologically informative for a given clinical feature or sub- phenotype comparison are shown for disease severity (**A**), source of infection (**B**), days in icU (**C**), SPc (**D**), and SrS (**E**) (see data file S2B). For each component, the figure shows the sample loading scores by group (box plots), protein or gene loading scores for those significantly contributing to the component (posterior inclusion probabil- ity > 0.5), *z* scores of representative proteins, and pathway enrichment

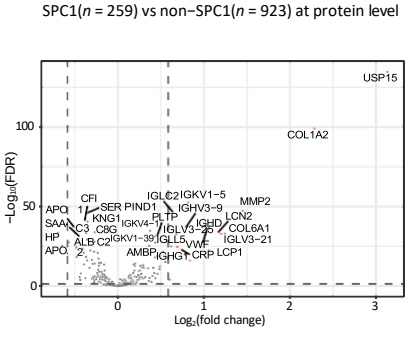
of significantly contributing proteins or genes.

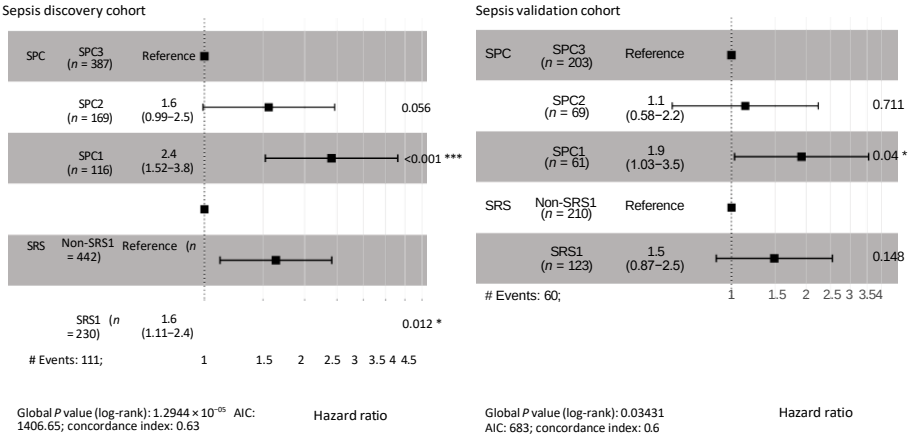
A

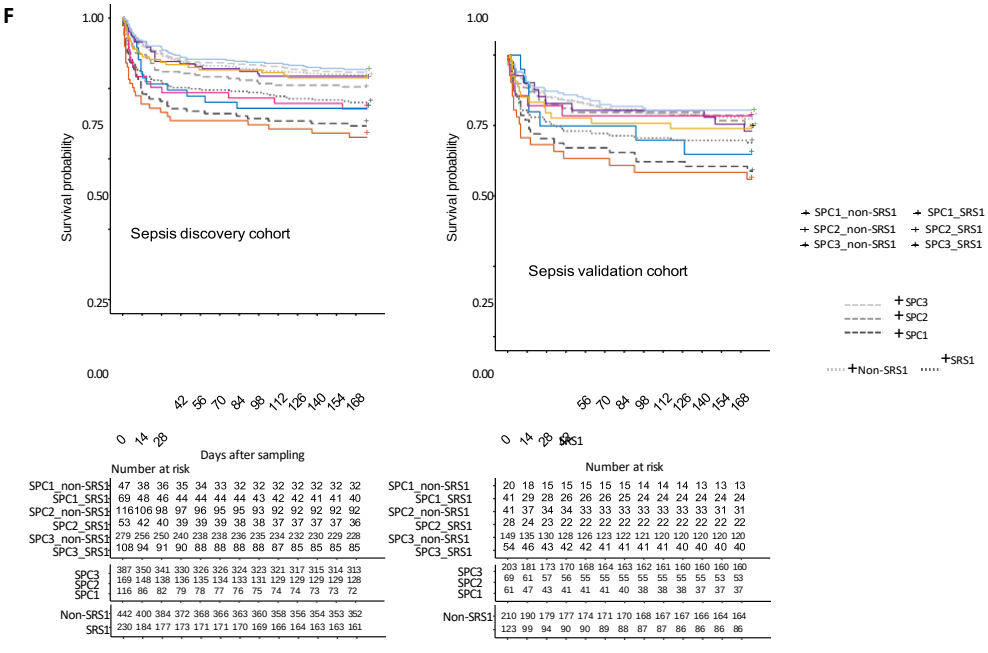
	Discovery cohort		Validation cohort	
	SRS1	Non-SRS1	SRS1	Non-SRS1
SPC1	105	44	50	20
SPC2	60	104	31	33
SPC3	126	242	63	138











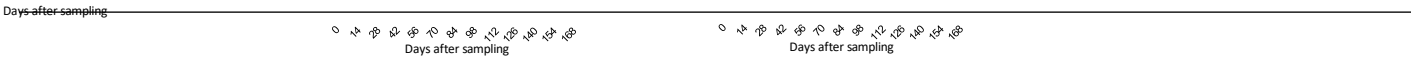


Fig. 8. Interaction of proteomic (SPC) and transcriptomic (SRS) patient subgroups. (**A** and **B**) overlap in SrS and SPc assignments in patients with sepsis by (a) patient numbers and (B) proportions. First available samples per patient used. (**C** and **D**) Protein differential abundance analysis for SrS (c) and SPc (d). (**E**) Multivariate cox pro- portional hazard regression on 28-day mortality considering both SPc and SrS assignments. aic, akaike's information criterion. (**F**) Kaplan-Meier curves comparing sur- vival at 6 months after sampling by SPc, SrS, or both assignments combined. (e) and (F) cluster assignments of the last sample (within icU day 1/3/5) per patient with both SrS and SPc assignments were used.

DISCUSSION

In this study, we have generated a comprehensive map of the human sepsis plasma proteome in terms of sample numbers, quantification methodologies, and comparator groups. We developed and applied a mass spectrometry–based approach at scale to understand how sep- sis differs from health, sterile inflammatory states, and noninfected critical illness as well as to reveal and define heterogeneity in the sep- sis response. The sepsis plasma proteome reflects mechanisms un- derlying the dysregulated host response to infection as well as the wider consequences of organ dysfunction (reduced metabolism and excretion for example) and tissue injury. Proteomics therefore provides an opportunity to identify aspects of pathogenesis together with measures of organ dysfunction and disease severity.

We identified specific proteins, coexpression modules, and networks that are differentially abundant in sepsis. These involved innate immunity, acute-phase response, neutrophil function, cytokine production, lipometabolism, tissue damage protection, and ECM or- ganization. By including multiple nonsepsis controls, we showed that these signals were consistently associated with sepsis. Our large sep- sis cohort with deep clinical phenotyping allowed us to determine that more severe illness is associated with specific proteins as well as modules enriched for S100 family proteins and extracellular matrix proteins (positive correlation), complement, and lipoprotein meta- bolic proteins (negative correlation). Some key proteomic markers of specific organ function showed correlations with the corresponding clinically accepted criteria as well as with overall mortality. For ex- ample, apolipoproteins, mainly synthesized in the liver, were corre- lated with bilirubin, alanine transaminase, and prothrombin time, and protein markers of declining renal function [PTGDS, B2M, CST3 (cystatin C), and LCN2] (*32*) had associations with all nine clinical features tested that reflect renal failure as well as positive correlations with total SOFA and mortality. These results indicate an association between the liver and kidney dysfunction and the plasma response proteome. One limitation is that the interpretation is limited to the two most common etiologies of sepsis (CAP and FP), but at the

Compared with a supervised strategy of directly modeling on clin- ical severity and mortality, which is often confounded by multiple contributors to the final cause of death in sepsis, our approach of stratifying patients by unsupervised clustering of protein abundanc- es or gene expression can effectively divide patients into more homogeneous subgroups with shared molecular mechanisms or physiology, which consistently associate with different risks of mor- tality. This approach may also more closely align with biological processes that could be targeted by immunomodulatory therapies. Further work is needed to understand the relationship of SPC with other transcriptomic and clinically defined subphenotypes reported in the literature as well as to establish the mechanisms driving the SPCs, whether they are reflective of treatable traits, and their thera- peutic utility. Animal and human studies, for example, have already highlighted lipoproteins as potential therapeutic targets in sepsis (*36*) and COVID-19 (*37*). Future investigations of the translational value of SPC in comparison to, or in combination with, SRS (and other biomarker-defined or clinically defined subphenotypes) requires prospective validation in urgent care settings and prospective incorporation into biomarker-led clinical trials where patients are allocated to treatment on the basis of testing for relevant sub- phenotypes.

The increasing availability of high-dimensional proteomic data for clinical and molecular disease phenotyping requires innovative approaches to analyze and integrate such datasets (*38*, *39*). Here, we leveraged analytical strategies developed for transcriptomics to investigate protein coexpression networks and modules and signatures of response. We further showed how matrix decomposition allows integration of paired plasma proteomic and leukocyte transcriptomic data, linking information from both datatypes. Our results are informative for illness severity and disease group and provide evidence that they are functionally related in a particular pathogenic process.

Recent advances in affinity-based proteome profiling platforms, including SOMAscan and Olink, have made them important tools in omics studies of various critical illnesses. These often comple- ment the LC-MS approach by achieving deep detection of thousands of proteins (*40–42*) and other more specific study aims, including multiomic profiling (*43*, *44*), measuring the postmortem tissue proteome (*45*), and determining changes in protein abundances over time (*46*)

. The number of proteins measured by SOMAscan are usually much larger than an LC-MS–based clinical proteomics study. However, the latter analysis of plasma samples allows unprejudiced discovery anal- ysis of potential biomarkers without the bias of affinity reagents compromising clinical, patient-specific measurements.

For example, many proteins included in our cleaned-up data are not targeted in the cur- rent SOMAscan panel (*47*), including top proteins distinguishing the SPC2 subset of patients (MAP1A, ARHGEF18, HR, TSPAN11, and SAGE2P) or that differentiate between sepsis and HV (ORM1 and TTR). In addition, the substantially lower costs and sample us- age make LC-MS–based analysis more accessible in larger cohort studies. The inclusion of >1000 critically ill sepsis patients (>1800 samples) in one batch in this work is critical for elucidating the het- erogeneity within our clinically defined patient group. The LC-MS approach also retains the potential to analyze proteoforms and post- translational modifications and is more robust to possible partial degradations of the proteins.

In this study, we show that medium to high-throughput proteomics across multiple large cohorts in a single batch is feasible on a single LC-MS platform. A simple semiautomatic sample preparation strategy in combination with the MS-based analysis of >2500 clinical, nondepleted plasma samples and a further ~2000 quality control and library samples that can be acquired through continuous measurements at a rate of 100 samples per day now reaches throughputs used by other proteomics technologies, such as Olink or SOMAscan. Inclusion of regular injections of a matrix pool to correct for variability is important; here, this comprised a pool of all 2612 samples in the study injected every 24 samples analyzed. Although measured proteome depth is limited because of the extreme dynamic range of plasma protein abundance, we achieved good cov- erage of the acute-phase proteome and markers of disease routinely quantified in single measurement assays such as enzyme-linked im- munosorbent assay (ELISA). In subsets of the samples, we also used the QE-HF system with depleted samples and Luminex assays (equiv- alent to a multiplex ELISA) to cover the deeper proteome, which verified the distinction between the patient subgroups and provided further information elucidating the biology. The throughput, cost effectiveness, and robustness of modern LC-MS platforms mean that this technology is now competitive with standard clinical practice measurements, such as ELISA, for absolute protein quantitation us- ing heavy isotope–labeled peptide standards, marking a transition from a pure discovery tool toward a more clinical point-of-care ap- plication in the coming years.

Study limitations include the detection limit in large-scale clini- cal MS studies. Further work is needed to fully establish the com- plete sepsis proteome across a wide dynamic range of protein abundance and size and to differentiate protein variation (proteoforms) (*48*). Recent advances in data-independent acquisition for MS would be compatible with high-throughput proteomics platforms and offer future opportunities to increase depth and data completeness. More widely, future work using MS will require streamlining sample and data processing workflows toward clinical certification and point- of-care use. Further work is also needed to quantify disease-relevant tissue-specific proteomes and establish pathological mechanisms, recognizing that obtaining tissue samples from the critically ill is challenging. Biomarker discovery for patient stratification will re- quire prospective validation and demonstration of whether they can inform specific therapeutic interventions, whereas identified pro- teins and pathways in critically ill patients represent potential future therapeutic targets. Additional work is also required to determine whether the maximally informative protein biomarkers can be ap- plied in combination with other "omic" platforms and currently available clinical or laboratory variables to better stratify patients. In conclusion, our study shows the feasibility and informativeness of high-throughput proteomics using MS as part of a multimodal tool kit for understanding the nature of our individual response to severe infection and moving toward a more precision medicine approach that may also be applicable to other disease states. **MATERIALS AND METHODS**

Study design

This was an observational study designed to understand the sepsis proteomic response and individual heterogeneity by assaying the plasma proteome of multiple sepsis and nonsepsis comparator cohorts and integrating with paired leukocyte transcriptomic data. Study cohorts included UK GAinS (*10*, *25*) [patients admitted to ICU with sepsis (*49*) due to CAP or FP (*50*–*52*)], VANISH (*53*) (clinical trial cohort of all-cause sepsis requiring vasopressors), HVs from the Oxford BioBank (*54*), elective surgery patients [X-MINS (*55*) and BIONIC (*56*)], and noninfected ICU patients [MOTION (*57*), MONOGRAM, and TACE (*58*)] with cohorts described in the Supplementary Materials (numbers of individuals and samples given in data file S1). Our processed timsTOF sepsis proteomics data were split into two nonoverlapping cohorts for some analyses. A discovery cohort (n = 1041 patients, 1590 samples) was used for interrogation of sepsis-specific proteome profiles and proteome-based sepsis subphenotypes, which were then investigated in the validation cohort (n = 557 patients, 985 samples). Sample sizes were determined by available patient and sample numbers recruited into each of the study cohorts, with statistical power calculation based on a smaller-scale sepsis proteomics dataset indicating that a minimum of 171 biological replicates was needed in each group to detect dif- ference at 80% power and FDR ≤ 0.05 for 70% of the analytes in the sepsis-control contrast and 269 biological replicates for the within- sepsis contrast. From the whole timsTOF dataset, 32 measurements that failed quality control, 35 duplicate measurements, and five sam- ples from excluded patients (because of withdrawal from study or no clinical information) were excluded from the processed data. No sample was further excluded in the downstream analysis. The pro- cessed timsTOF MS dataset did not contain repeated measurements of the same samples. For MS data acquisition, samples from all co- horts were fully randomized across the acquisition plates. Data gen- eration was performed on these plates without distinguishing the study cohort or knowing the SPC clusters at the time of generation.

Patients from the GAinS or OBB studies were separated into the discovery or validation cohorts by random draws.

Statistical analysis

Differential abundance analysis for proteins and differential expression analysis for genes were performed by fitting the intensities in linear models using the limma R package (*59*), using only the first available sample of each patient and including age and sex as covariates. The Benjamini-Hochberg procedure was used to adjust for multiple testing. Significance of differential abundance was defined as FDR < 0.05 and FC > 1.5 unless otherwise specified. Comparisons between the pre- and postoperative samples are paired and with no additional covariates. All tests are two-sided.

Cytokine concentrations measured by Luminex assay were compared by Wilcoxon rank sum tests (Mann-Whitney tests) using the first available sample of each patient. Categorical clinical variables were compared using χ^2 tests without Yate's correction. Numerical clinical variables were compared between two groups using Wilcoxon rank sum tests and compared between three groups using Kruskal-Wallis test and Dunn's post hoc tests. Only the first available samples after ICU admission of each patient were included in comparing clinical variables.

Survival differences were assessed and Kaplan-Meier curves plotted using the R packages survival and survminer (*60*, *61*). The input data were a data frame specifying the time to event from the sampling day, the event (death or end of 28-day or 6-month observation), and the patient subgroups. For patients with multiple time points sampled, cluster assignment from the last available sample within 5 days of ICU admission was used. *P* values were given by log-rank tests. HRs and the CIs were calculated using univariate (SPC) or multivariate (SPC and SRS) Cox proportional hazard models.

Methods for LC-MS/MS using the high-throughput Evosep One–Bruker timsTOF Pro platform, leukocyte whole-genome expression profiling, timsTOF protein identification and quantification, timsTOF data preprocessing, QE-HF MS, and other statistical analyses, including unsupervised clustering and matrix decomposition, are described in Supplementary Materials and Methods.

REFERENCES AND NOTES

1. M. Singer, c. S. deutschman, c. Seymour, M. Shankar-hari, d. annane, M. Bauer, r. Bellomo, G. r. Bernard, J. d. chiche, c. M. coopersmith, r. S. hotchkiss, M. M. levy, J. c. Marshall, G. S. Martin, S. M. opal, G. d. rubenfeld, T. der Poll, J. l. Vincent, d. c. angus, The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* **315**, 801–810 (2016).
2. K. e. rudd, S. c. Johnson, K. M. agesa, K. a. Shackelford, d. Tsoi, d. r. Kievlan, d. V. colombara, K. S. ikuta, n. Kissoon, S. Finfer, c. Fleischmann-Struzek, F. r. Machado, K. K. reinhart, K. rowan, c. W. Seymour, r. S. Watson, T. e. West, F. Marinho, S. i. hay, r. lozano, a. d. lopez, d. c. angus, c. J. l. Murray, M. naghavi, Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of disease Study. *Lancet* **395**, 200–211 (2020).
3. T. van der Poll, M. Shankar-hari, W. J. Wiersinga, The immunology of sepsis. *Immunity* **54**, 2450–2464 (2021).
4. J. c. Marshall, Why have clinical trials in sepsis failed? *Trends Mol. Med.* **20**, 195–203 (2014).
5. n. l. Stanski, h. r. Wong, Prognostic and predictive enrichment in sepsis. *Nat. Rev. Nephrol.* **16**, 20–31 (2020).
6. c. W. Seymour, J. n. Kennedy, S. Wang, c. c. h. chang, c. F. elliott, Z. Xu, S. Berry, G. clermont, G. cooper, h. Gomez, d. T. huang, J. a. Kellum, Q. Mi, S. M. opal, V. Talisa, T. Van der Poll, S. Visweswaran, Y. Vodovotz, J. c. Weiss, d. M. Yealy, S. Yende, d. c. angus, derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* **321**, 2003–2017 (2019).
7. a. Baghela, o. M. Pena, a. h. lee, B. Baquir, r. Falsafi, a. an, S. W. Farmer, a. hurlburt, a. Mondragon-cardona, J. d. rivera, a. Baker, U. Trahtemberg, M. Shojaei, c. e. Jimenez-canizales, c. c. dos Santos, B. Tang, h. r. Bouma, G. V. cohen Freue, r. e. W. hancock, Predicting sepsis severity at first clinical presentation: The role of endotypes and mechanistic signatures. *EBioMedicine* **75**, 103776 (2022).
8. T. e. Sweeney, T. d. azad, M. donato, W. a. haynes, T. M. Perumal, r. henao, J. F. Bermejo-Martin, r. almansa, e. Tamayo, J. a. howrylak, a. choi, G. P. Parnell, B. Tang, M. nichols, c. W. Woods, G. S. Ginsburg, S. F. Kingsmore, l. omberg, l. M. Mangravite, h. r. Wong, e. l. Tsalik, r. J. langley, P. Khatri, Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters. *Crit. Care Med.* **46**, 915–925 (2018).
9. h. r. Wong, J. T. caldwell, n. Z. cvijanovich, S. l. Weiss, J. c. Fitzgerald, M. T. Bigham, P. n. Jain, a. Schwarz, r. lutfi, J. nowak, G. l. allen, n. J. Thomas, J. r. Grunwell, T. Baines, M. Quasney, B. haileselassie, c. J. lindsell, Prospective clinical testing and experimental validation of the Pediatric Sepsis Biomarker risk Model. *Sci. Transl. Med.* **11**, eaax9000 (2019).
10. e. e. davenport, K. l. Burnham, J. radhakrishnan, P. humburg, P. hutton, T. c. Mills, a. rautanen, a. c. Gordon, c. Garrard, a. V. S. hill, c. J. hinds, J. c. Knight, Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. *Lancet Respir. Med.* **4**, 259–271 (2016).
11. B. P. Scicluna, l. a. van Vught, a. h. Zwinderman, M. a. Wiewel, e. e. davenport, K. l. Burnham, P. nürnberg, M. J. Schultz, J. horn, o. l. cremer, M. J. Bonten, c. J. hinds, h. r. Wong, J. c. Knight, T. van der Poll, F. M. de Beer, l. d. J. Bos, J. F. Frencken, M. e. Koster-Brouwer, K. van de Groep, d. M. Verboom, G. J. Glas, r. T. M. van hooijdonk, a. J. hoogendijk, M. a. huson, P. M. K. Klouwenberg, d. S. Y. ong, l. r. a. Schouten, M. Straat, e. Witteveen, l. Wieske, classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *Lancet Respir. Med.* **5**, 816–826 (2017).
12. e. cano-Gamez, K. l. Burnham, c. Goh, a. allcock, Z. h. Malick, l. overend, a. Kwok, d. a. Smith, h. Peters-Sengers, d. antcliffe, G. S. investigators, S. McKechnie, B. P. Scicluna, T. van der Poll, a. c. Gordon, c. J. hinds, e. e. davenport, J. c. Knight, an immune dysfunction score for stratification of patients with acute infection based on whole-blood gene expression. *Sci. Transl. Med.* **14**, eabq4433 (2022).
13. a. J. Kwok, a. allcock, r. c. Ferreira, e. cano-Gamez, M. Smee, K. l. Burnham, Y. X. Zurke, a. novak, M. darwent, T. Baron, c. Brown, S. Beer, a. espinosa, T. Panduro, d. Georgiou, J. Martinez, h. Thraves, e. Perez, r. Fernandez, a. Sobrino, V. Sanchez, r. Magallano, K. dineen, J. Wilson, S. McKechnie, a. J. Mentzer, c. Monaco, i. a. Udalova, c. J. hinds, J. a. Todd, e. e. davenport, J. c. Knight, neutrophils and emergency granulopoiesis drive immune suppression and an extreme response endotype during sepsis. *Nat. Immunol.* **24**, 767–779 (2023).
14. r. J. langley, e. l. Tsalik, J. c. Van Velkinburgh, S. W. Glickman, B. J. rice, c. Wang, B. chen, l. carin, a. Suarez, r. P. Mohney, d. h. Freeman, M. Wang, J. You, J. Wulff, J. Will Thompson, M. arthur Moseley, S. reisinger, B. T. edmonds, B. Grinnell, d. r. nelson, d. l. dinwiddie, n. a. Miller, c. J. Saunders, S. S. Soden, a. J. rogers, l. Gazourian, l. e. Fredenburgh, a. F. Massaro, r. M. Baron, a. M. K. choi, G. ralph corey, G. S. Ginsburg, c. B. cairns, r. M. otero, V. G. Fowler, e. P. rivers, c. W. Woods, S. F. Kingsmore, an integrated clinico-metabolomic model improves prediction of death in sepsis. *Sci. Transl. Med.* **5**, 195ra95 (2013).
15. a. de coux, Y. Tian, K. Y. deleon-Pennell, n. T. nguyen, l. e. de castro Brás, e. r. Flynn, P. l. cannon, M. e. Griswold, Y. F. Jin, M. a. Puskarich, a. e. Jones, M. l. lindsey, Plasma glycoproteomics reveals sepsis outcomes linked to distinct proteins in common pathways. *Crit. Care Med.* **43**, 2049–2058 (2015).
16. G. G. F. leite, B. l. Ferreira, a. K. Tashima, e. S. nishiduka, e. cunha-neto, M. Karina, c. Brunialti, B. P. Scicluna, r. Salomão, combined transcriptome and proteome leukocyte's profiling reveals up-regulated module of genes/proteins related to low density neutrophils and impaired transcription and translation processes in clinical sepsis. *Front. Immunol.* **12**, 744799 (2021).
17. M. Kiehntopf, d. Schmerler, F. M. Brunkhorst, r. Winkler, K. ludewig, d. osterloh, F. Bloos, K. reinhart, T. deufel, Mass spectometry-based protein patterns in the diagnosis of sepsis/systemic inflammatory response syndrome. *Shock* **36**, 560–569 (2011).
18. n. K. Sharma, a. K. Tashima, M. K. c. Brunialti, e. r. Ferreira, r. J. S. Torquato, r. a. Mortara, F. r. Machado, M. assuncao, o. rigato, r. Salomao, Proteomic study revealed cellular

assembly and lipid metabolism dysregulation in sepsis secondary to community- acquired pneumonia. *Sci. Rep.* **7**, 15606 (2017).

19. T. Thavarajah, c. c. dos Santos, a. S. Slutsky, J. c. Marshall, P. Bowden, a. romaschin, J. G. Marshall, The plasma peptides of sepsis. *Clin. Proteomics* **17**, 26 (2020).
20. G. Pimienta, d. M. heithoff, a. rosa-campos, M. Tran, J. d. esko, M. J. Mahan, J. d. Marth, J. W. Smith, Plasma proteome signature of sepsis: a functionally connected protein network. *Proteomics* **19**, e1800389 (2019).
21. a. G. Toledo, G. Golden, a. r. campos, h. cuello, J. Sorrentino, n. lewis, n. Varki, V. nizet, J. W. Smith, J. d. esko, Proteomic atlas of organ vasculopathies triggered by *Staphylococcus aureus* sepsis. *Nat. Commun.* **10**, 4656 (2019).
22. c. d. Fjell, S. Thair, J. l. hsu, K. r. Walley, J. a. russell, J. Boyd, cytokines and signaling molecules predict clinical outcomes in sepsis. *PLOS ONE* **8**, e79207 (2013).
23. n. K. Sharma, B. l. Ferreira, a. K. Tashima, M. K. c. Brunialti, r. J. S. Torquato, a. Bafi, M. assuncao, l. c. P. azevedo, r. Salomao, lipid metabolism impairment in patients with sepsis secondary to hospital acquired pneumonia, a proteomic analysis. *Clin. Proteomics* **16**, 29 (2019).
24. h. Miao, S. chen, r. ding, evaluation of the molecular mechanisms of sepsis using proteomics. *Front. Immunol.* **12**, 733537 (2021).
25. a. rautanen, T. c. Mills, a. c. Gordon, P. hutton, M. Steffens, r. nuamah, J. d. chiche, T. Parks, S. J. chapman, e. e. davenport, K. S. elliott, J. Bion, P. lichtner, T. Meitinger, T. F. Wienker, M. J. caulfield, c. Mein, F. Bloos, i. Bobek, P. cotogni, V. Sramek, S. Sarapuu, M. Kobilay, V. M. ranieri, J. rello, G. Sirgo, Y. G. Weiss, S. russwurm, e. M. Schneider, K. reinhart, P. a. h. holloway, J. c. Knight, c. S. Garrard, J. a. russell, K. r. Walley, F. Stüber, a. V. S. hill, c. J. hinds, Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study. *Lancet Respir. Med.* **3**, 53–60 (2015).
26. r. aebersold, M. Mann, Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
27. l. a. Beer, B. Ky, K. T. Barnhart, d. W. Speicher, in-depth, reproducible analysis of human plasma using igY 14 and SuperMix immunodepletion. *Methods Mol. Biol.* **1619**, 81–101 (2017).
28. r. r. Shields-cutler, J. r. crowley, c. d. Miller, a. e. Stapleton, W. cui, J. P. henderson, human metabolome-derived cofactors are required for the antibacterial activity of siderocalin in urine. *J. Biol. Chem.* **291**, 25901–25910 (2016).
29. e. Malmström, o. Kilsgård, S. hauri, e. Smeds, h. herwald, l. Malmström, J. Malmström, large-scale inference of protein tissue origin in gram-positive sepsis plasma using quantitative targeted proteomics. *Nat. Commun.* **7**, 10261 (2016).
30. V. hore, a. Viñuela, a. Buil, J. Knight, M. i. Mccarthy, K. Small, J. Marchini, Tensor decomposition for multiple-tissue gene expression experiments. *Nat. Genet.* **48**, 1094–1100 (2016).
31. d. B. antcliffe, K. l. Burnham, F. al-Beidh, S. Santhakumaran, S. J. Brett, c. J. hinds, d. ashby, J. c. Knight, a. c. Gordon, Transcriptomic signatures in sepsis and a differential response to steroids from the VaniSh randomized trial. *Am. J. Respir. Crit. Care Med.* **199**, 980–986 (2019).
32. e. l. Tsalik, l. K. Willig, B. J. rice, J. c. Van Velkinburgh, r. P. Mohney, J. e. Mcdunn, S. F. Kingsmore, r. J. langley, renal systems biology of patients with systemic inflammatory response syndrome. *Kidney Int.* **88**, 804–814 (2015).
33. d. M. Maslove, B. Tang, M. Shankar-hari, P. r. lawler, d. c. angus, J. K. Baillie, r. M. Baron, M. Bauer, T. G. Buchman, c. S. calfee, c. c. dos Santos, e. J. Giamarellos-Bourboulis, a. c. Gordon, J. a. Kellum, J. c. Knight, a. leligdowicz, d. F. Mcauley, a. S. Mclean, d. K. Menon, n. J. Meyer, l. l. Moldawer, K. reddy, J. P. reilly, J. a. russell, J. e. Sevransky, c. W. Seymour, n. i. Shapiro, M. Singer, c. Summers, T. e. Sweeney, B. T. Thompson, T. van der Poll, B. Venkatesh, K. r. Walley, T. S. Walsh, l. B. Ware, h. r. Wong, Z. e. Zador, J. c. Marshall, redefining critical illness. *Nat. Med.* **28**, 1141–1148 (2022).
34. Z. Zador, a. landry, M. d. cusimano, n. Geifman, Multimorbidity states associated with higher mortality rates in organ dysfunction and sepsis: a data-driven analysis in critical care. *Crit. Care* **23**, 247 (2019).
35. K. l. Burnham, e. e. davenport, J. radhakrishnan, P. humburg, a. c. Gordon, P. hutton, e. Svoren-Jabalera, c. Garrard, a. V. S. hill, c. J. hinds, J. c. Knight, Shared and distinct aspects of the sepsis transcriptomic response to fecal peritonitis and pneumonia. *Am. J. Respir. Crit. Care Med.* **196**, 328–339 (2017).
36. S. Tanaka, d. couret, a. Tran-dinh, J. duranteau, P. Montravers, a. Schwendeman, o. Meilhac, high-density lipoproteins during sepsis: From bench to bedside. *Crit. Care* **24**, 134 (2020).
37. V. chidambaram, h. Shanmugavel Geetha, a. Kumar, M. G. Majella, r. K. Sivakumar, d. Voruganti, J. l. Mehta, P. c. Karakousis, association of lipid levels with coVid-19 infection, disease severity and mortality: a systematic review and meta-analysis. *Front. Cardiovasc. Med.* **9**, 862999 (2022).
38. a. T. rajczewski, P. d. Jagtap, T. J. Griffin, an overview of technologies for MS-based proteomics-centric multi-omics. *Expert Rev. Proteomics* **19**, 165–181 (2022).
39. S. Graw, K. chappell, c. l. Washam, a. Gies, J. Bird, M. S. robeson, S. d. Byrum, Multi-omics data integration considerations and study design for biological systems and disease. *Mol. Omi.* **17**, 170–185 (2021).
40. l. r. Van nynatten, M. Slessarev, c. M. Martin, a. leligdowicz, M. r. Miller, M. a. Patel, M. daley, e. K. Patterson, G. cepinskas, d. d. Fraser, novel plasma protein biomarkers from critically ill sepsis patients. *Clin. Proteomics* **19**, 50 (2022).
41. i. Paranjpe, P. Jayaraman, c.-Y. Su, S. Zhou, S. chen, r. Thompson, d. M. del Valle, e. Kenigsberg, S. Zhao, S. Jaladanki, K. chaudhary, S. ascolillo, a. Vaid, e. Gonzalez-Kozlova, J. Kauffman, a. Kumar, M. Paranjpe, r. o. hagan, S. Kamat, F. F. Gulamali, h. Xie, J. harris, M. Patel, K. argueta, c. Batchelor, K. nie, S. dellepiane, l. Scott, M. a. levin, J. c. he, M. Suarez-Farinas, S. G. coca, l. chan, e. U. azeloglu, e. Schadt, n. Beckmann, S. Gnjatic, M. Merad, S. Kim-Schulze, B. richards, B. S. Glicksberg, a. W. charney, G. n. nadkarni, Proteomic characterization of acute kidney injury in patients hospitalized with SarS-coV2 infection. *Commun. Med.* **3**, 81 (2023).
42. e. duijvelaar, J. Gisby, J. Peters, h. J. Bogaard, J. aman, longitudinal plasma proteomics reveals alveolar-capillary barrier disruption in critically ill coVid-19 patients. *Nat. Commun.* **15**, 744 (2023).
43. r. Batra, r. Uni, o. M. akchurin, S. alvarez-Mulett, l. G. Gómez-escobar, e. Patino, K. l. hoffman, W. Simmons, W. Whalen, K. chetnik, M. Buyukozkan, e. Benedetti, K. Suhre, e. Schenck, SSchmidt, M. e. choi, J. Krumsiek, Urine-based multi-omic comparative analysis of coVid-19 and bacterial sepsis-induced ardS. *Mol. Med.* **29**, 13 (2023).
44. J. Wu, Y. Vodovotz, S. abdelhamid, F. X. Guyette, M. B. Yaffe, d. S. Gruen, a. cyr, d. o. okonkwo, U. K. Kar, n. Krishnamoorthi, r. G. Voinchet, i. M. Billiar, M. h. Yazer, r. a. namas, B. J. daley, r. S. Miller, B. G. harbrecht, J. a. claridge, h. a. Phelan, B. S. Zuckerbraun, P. i. Johansson, J. Stensballe, J. h. Morrissey, r. P. Tracy, S. r. Wisniewski, M. d. neal, J. l. Sperry, T. r. Billiar, Multi-omic analysis in injured humans: Patterns align with outcomes and treatment responses. *Cell Rep. Med.* **2**, 100478 (2021).
45. c. d. russell, a. Valanciute, n. n. Gachanja, J. Stephen, r. Penrice-randal, S. d. armstrong, S. clohisey, B. Wang, W. al Qsous, W. a. Wallace, G. c. oniscu, J. Stevens, d. J. harrison, K. dhaliwal, J. a. hiscox, J. K. Baillie, a. r. akram, d. a. dorward, c. d. lucas, Tissue proteomic analysis identifies mechanisms and stages of immunopathology in fatal coVid-19. *Am. J. Respir. Cell Mol. Biol.* **66**, 196–205 (2022).
46. J. Bonaroti, i. Billiar, h. Moheimani, J. Wu, r. namas, S. li, U. K. Kar, Y. Vodovotz, M. d. neal, J. l. Sperry, T. r. Billiar, Plasma proteomics reveals early, broad release of chemokine, cytokine, TnF, and interferon mediators following trauma with delayed increases in a subset of chemokines and cytokines in patients that remain critically ill. *Front. Immunol.* **13**, 1038086 (2022).
47. SomaScan Menu, Somalogic, https://menu.somalogic.com/.
48. r. d. Melani, V. r. Gerbasi, l. c. anderson, J. W. Sikora, T. K. Toby, J. e. hutton, d. S. Butcher, F. negrão, h. S. Seckler, K. Srzentić, l. Fornelli, J. M. camarillo, r. d. leduc, a. J. cesnik, e. lundberg, J. B. Greer, r. T. Fellers, M. T. robey, c. J. dehart, e. Forte, c. l. hendrickson, S. e. abbatiello, P. M. Thomas, a. i. Kokaji, J. levitsky, n. l. Kelleher, The Blood Proteoform atlas: a reference map of proteoforms in human hematopoietic cells. *Science* **375**, 411–418 (2022).
49. M. M. levy, M. P. Fink, J. c. Marshall, e. abraham, d. angus, d. cook, J. cohen, S. M. opal, J. l. Vincent, G. ramsay, 2001 SccM/eSicM/accP/aTS/SiS international Sepsis definitions conference. *Crit. Care Med.* **31**, 1250–1256 (2003).
50. d. c. angus, T. J. Marrie, d. Scott obrosky, G. clermont, T. T. dremsizov, c. coley, M. J. Fine, d. e. Singer, W. n. Kapoor, Severe community-acquired pneumonia: Use of intensive care services and evaluation of american and British Thoracic Society diagnostic criteria. *Am. J. Respir. Crit. Care Med.* **166**, 717–723 (2002).
51. a. P. Walden, G. M. clarke, S. McKechnie, P. hutton, a. c. Gordon, J. rello, J. d. chiche, F. Stueber, c. S. Garrard, c. J. hinds; eSicM/eccrn GenoSept investigators, Patients with community acquired pneumonia admitted to european intensive care units: an epidemiological survey of the GenoSept cohort. *Crit. Care* **18**, r58 (2014).
52. a. Tridente, G. M. clarke, a. Walden, S. McKechnie, P. hutton, G. h. Mills, a. c. Gordon, P. a. h. holloway, J. d. chiche, J. Bion, F. Stuber, c. Garrard, c. J. hinds, Patients with faecal peritonitis admitted to european intensive care units: an epidemiological survey of the GenoSept cohort. *Intensive Care Med.* **40**, 202–210 (2014).
53. a. c. Gordon, a. J. Mason, n. Thirunavukkarasu, G. d. Perkins, M. cecconi, M. cepkova, d. G. Pogson, h. d. aya, a. anjum, G. J. Frazier, S. Santhakumaran, d. ashby, S. J. Brett; VaniSh investigators, effect of early vasopressin vs norepinephrine on kidney failure in patients with septic shock: The VaniSh randomized clinical Trial. *JAMA* **316**, 509–518 (2016).
54. F. Karpe, S. K. Vasan, S. M. humphreys, J. Miller, J. cheeseman, a. l. dennis, M. J. neville, cohort profile: The oxford Biobank. *Int. J. Epidemiol.* **47**, 21–21g (2018).
55. S. M. May, a. reyes, G. Martir, J. reynolds, l. G. Paredes, S. Karmali, r. c. M. Stephens, d. Brealey, G. l. ackland, acquired loss of cardiac vagal activity is associated with myocardial injury in patients undergoing noncardiac surgery: Prospective observational mechanistic cohort study. *Br. J. Anaesth.* **123**, 758–767 (2019).

56. h. d. Torrance, P. Zhang, e. r. longbottom, Y. Mi, J. P. Whalley, a. allcock, a. J. Kwok, e. cano-Gamez, c. G. Geoghegan, K. l. Burnham, d. B. antcliffe, e. e. davenport, r. M. Pearse, M. J. o'dwyer, c. J. hinds, J. c. Knight, a. c. Gordon, a transcriptomic approach to understand patient susceptibility to pneumonia after abdominal surgery. *Ann. Surg.* **279**, 510–520 (2024).
57. P. B. Patel, S. J. Brett, d. o'callaghan, a. anjum, M. cross, J. Warwick, a. c. Gordon, Methylnaltrexone for the treatment of opioid-induced constipation and gastrointestinal stasis in intensive care patients. results from the MoTion trial. *Intensive Care Med.* **46**, 747–755 (2020).
58. d. J. P. o'callaghan, K. P. o'dea, a. J. Scott, M. Takata, a. c. Gordon, Monocyte tumor necrosis factor-α-converting enzyme catalytic activity and substrate shedding in sepsis and noninfectious systemic inflammation. *Crit. Care Med.* **43**, 1375–1385 (2015).
59. M. e. ritchie, B. Phipson, d. Wu, Y. hu, c. W. law, W. Shi, G. K. Smyth, limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
60. T. M. Therneau, P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model* (Springer, 2000).
61. a. Kassambara, M. Kosinski, P. Biecek, survminer: drawing Survival curves using "ggplot2" (comprehensive r archive network, 2021); https://cran.r-project.org/ package=survminer.
62. S. M. May, T. e. F. abbott, a. G. del arroyo, a. reyes, G. Martir, r. c. M. Stephens, d. Brealey, B. h. cuthbertson, d. n. Wijeysundera, r. M. Pearse, G. l. ackland, Microrna signatures of perioperative myocardial injury after elective noncardiac surgery: a prospective observational mechanistic cohort study. *Br. J. Anaesth.* **125**, 661–671 (2020).
63. d. antcliffe, B. Jiménez, K. Veselkov, e. holmes, a. c. Gordon, Metabolic profiling in patients with pneumonia on intensive care. *EBioMedicine* **18**, 244–253 (2017).
64. d. B. antcliffe, a. M. Wolfer, K. P. o'dea, M. Takata, e. holmes, a. c. Gordon, Profiling inflammatory markers in patients with pneumonia on intensive care. *Sci. Rep.* **8**, 14736 (2018).
65. h. Keshishian, M. W. Burgess, h. Specht, l. Wallace, K. r. clauser, M. a. Gillette, S. a. carr, Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nat. Protoc.* **12**, 1683–1701 (2017).
66. covid-19 Multi-omics Blood aTlas (coMBaT) consortium, a blood atlas of coVid-19 defines hallmarks of disease severity and specificity. *Cell* **185**, 916–938.e58 (2022).
67. n. Bache, o. hoerning, l. Falkenby, o. Vorm, evosep one: a gradient off-set focusing UhPlc instrument for robust and high throughput proteomics; https://evosep.com/ wp-content/uploads/2017/09/tech-note-evosep-a4-15.09.pdf.
68. n. Bache, P. e. Geyer, d. B. Bekker-Jensen, o. hoerning, l. Falkenby, P. V. Treit, S. doll, i. Paron, J. B. Müller, F. Meier, J. V. olsen, o. Vorm, M. Mann, a novel lc system embeds analytes in pre-formed gradients for rapid, ultra-robust proteomics. *Mol. Cell. Proteomics* **17**, 2284–2296 (2018).
69. n. Prianichnikov, h. Koch, S. Koch, M. lubeck, r. heilig, S. Brehmer, r. Fischer, J. cox, MaxQuant software for ion mobility enhanced shotgun proteomics. *Mol. Cell. Proteomics* **19**, 1058–1069 (2020).
70. d. B. antcliffe, Y. Mi, S. Santhakumaran, K. l. Burnham, T. Prevost, J. Ward, T. Marshall, c. Bradley, F. al-Beidh, P. hutton, S. McKechnie, e. e. davenport, c. J. hinds, c. M. o'Kane, d. Mcauley, M. Shankar-hari, a. c. Gordon, J. c. Knight, inflammatory sub-phenotypes in sepsis: relationship to outcomes, treatment effect and transcriptomic sub-phenotypes. medrxiv 2022.07.12.22277463 [Preprint] (2022). https://doi.org/10.1101/2022.07.12.22277463.
71. F. Yu, S. e. haynes, G. c. Teo, d. M. avtonomov, d. a. Polasky, a. i. nesvizhskii, Fast quantitative analysis of timsToF PaSeF data with MSFragger and ionQuant. *Mol. Cell. Proteomics* **19**, 1575–1585 (2020).
72. a. T. Kong, F. V. leprevost, d. M. avtonomov, d. Mellacheruvu, a. i. nesvizhskii, MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nat. Methods* **14**, 513–520 (2017).
73. F. da Veiga leprevost, S. e. haynes, d. M. avtonomov, h. Y. chang, a. K. Shanmugam, d. Mellacheruvu, a. T. Kong, a. i. nesvizhskii, Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* **17**, 869–870 (2020).
74. a. Keller, a. i. nesvizhskii, e. Kolker, r. aebersold, empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392 (2002).
75. a. i. nesvizhskii, a. Keller, e. Kolker, r. aebersold, a statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).
76. F. Yu, S. e. haynes, a. i. nesvizhskii, ionQuant enables accurate and sensitive label-free quantification with Fdr-controlled match-between-runs. *Mol. Cell. Proteomics* **20**, 100077 (2021).
77. UniProt consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, d480–d489 (2021).
78. P. e. Geyer, e. Voytik, P. V. Treit, S. doll, a. Kleinhempel, l. niu, J. B. Müller, M. Buchholtz, J. M. Bader, d. Teupser, l. M. holdt, M. Mann, Plasma proteome profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* **11**, e10427 (2019).
79. W. huber, a. von heydebreck, h. Sültmann, a. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18** (Suppl. 1), S96–S104 (2002).
80. a. chawade, e. alexandersson, F. levander, normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *J. Proteome Res.* **13**, 3114–3120 (2014).
81. T. Välikangas, T. Suomi, l. l. elo, a systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.* **19**, bbw095 (2016).
82. n. a. Karp, W. huber, P. G. Sadowski, P. d. charles, S. V. hester, K. S. lilley, addressing accuracy and precision issues in iTraQ quantitation. *Mol. Cell. Proteomics* **9**, 1885–1897 (2010).
83. T. hastie, r. Tibshirani, B. narasimhan, G. chu, impute: imputation for Microarray data (Bioconductor, 2024); https://bioconductor.org/packages/devel/bioc/manuals/impute/man/impute.pdf.
84. J. T. leek, W. e. Johnson, h. S. Parker, e. J. Fertig, a. e. Jaffe, Y. Zhang, J. d. Storey, l. c. Torres, sva: Surrogate Variable analysis (Bioconductor, 2021); https://bioconductor.riken.jp/packages/3.12/bioc/manuals/sva/man/sva.pdf.
85. h. Fang, B. Knezevic, K. l. Burnham, J. c. Knight, XGr software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med.* **8**, 129 (2016).
86. P. Shannon, a. Markiel, o. ozier, n. S. Baliga, J. T. Wang, d. ramage, n. amin, B. Schwikowski, T. ideker, cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
87. J. h. Morris, l. apeltsin, a. M. newman, J. Baumbach, T. Wittkop, G. Su, G. d. Bader, T. e. Ferrin, clusterMaker: a multi-algorithm clustering plugin for cytoscape. *BMC Bioinformatics* **12**, 436 (2011).
88. P. langfelder, S. horvath, WGcna: an r package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
89. M. reich, T. liefeld, J. Gould, J. lerner, P. Tamayo, J. P. Mesirov, GenePattern 2.0. *Nat. Genet.* **38**, 500–501 (2006).
90. r core Team, *R: A Language and Environment for Statistical Computing* (r Foundation for Statistical computing, 2015).
91. J. c. Brunson, ggalluvial: layered grammar for alluvial plots. *J. Open Source Softw.* **5**, 2017 (2020).
92. J. Pearl, interpretation and identification of causal mediation. *Psychol. Methods* **19**, 459–481 (2014).
93. d. Tingley, T. Yamamoto, K. hirose, l. Keele, K. imai, Mediation: r package for causal mediation analysis. *J. Stat. Softw.* **59**, 1–38 (2014).
94. M. d. Wilkerson, d. n. hayes, consensusclusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
95. e. a. Thévenot, a. roux, Y. Xu, e. ezan, c. Junot, analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and oPlS statistical analyses. *J. Proteome Res.* **14**, 3322–3335 (2015).
96. J. Friedman, T. hastie, r. Tibshirani, regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
97. M. Kuhn, caret: classification and regression Training (comprehensive r archive network, 2021); https://cran.r-project.org/package=caret.
98. a. liaw, M. Wiener, classification and regression by randomForest. *R News* **2**, 18–22 (2002).

Acknowledgments: We thank participating patients and volunteers from the oxford BioBank (www.oxfordbiobank.org.uk). We also thank laboratory members (of J.c.K. and r.F. groups and the MS laboratory at the Target discovery institute oxford, led by B. M. Kessler) for useful discussions.

Funding: This research and participating researchers were supported by the following funders: national institute for health and care research (nihr) through the comprehensive clinical research network for patient recruitment, Wellcome Trust [Wellcome Trust investigator award (204969/Z/16/Z) (J.c.K.), core funding to the Wellcome centre for human Genetics (090532/Z/09/Z and 203141/Z/16/Z), and Wellcome Sanger institute (206194 and 220540/Z/20/a)], Medical research council (Mr/V002503/1) (J.c.K., e.e.d., and a.c.G.), caMS iFMS 2018-i2M-2-002 (J.c.K.), china Scholarship council–University of oxford Scholarship (Y.M.), BJa/rca career development award (G.l.a.), British oxygen company (G.l.a.), British heart Foundation (rG/14/4/30736 and rG/19/5/34463) (G.l.a.), nihr advanced Fellowship (nihr 300097) (G.l.a.), nihr oxford Biomedical research centre (J.c.K., oxford BioBank, and oxford Bioresource), nihr imperial Biomedical research centre (d.a., r.d., d.J.P.o., P.P., a.c.G., and Bionic), nihr research Professor award (rP-2015-06-018) (a.c.G.), nihr research for Patient Benefit awards (MonGraMS, PB-PG-0613-31073 and VaniSh, PB-PG-0610-22350), nihr academic clinical Fellowship (acF-2019-21-011) (h.d.T.), intensive care Foundation Young investigators awards (Tace and MonoGraMS), national institute of academic anaesthesia (Bionic, BJa/rcoa Project grant WKr0-2020-0019), and european Society of anaesthesiologists (Bionic). The views expressed are those of the author(s) and not necessarily those of the nhS, the nihr, or the department of health. **Author contributions:** J.c.K. and r.F. conceptualized the study. Y.M. and P.d.c. curated the data. Formal analysis was performed by Y.M., K.l.B., P.d.c., J.W., and r.F. Y.M., K.l.B., r.h., i.V., and c.G.G. performed the investigations.

Methodology was developedby Y.M. and P.d.c. J.c.K. and r.F. provided project administration. J.W., h.d.T., d.B.a., S.M.M., M.J.n., P.h., c.G., J.r., e.e.d., S.Mc.K., r.d., d.J.P.o., P.P., a.G.a., F.K., a.c.G., G.l.a., c.J.h., and J.c.

provided resources (recruited patients, provided samples, materials, computing resources, or analysis tools). a.n. and F.Y. developed the software. J.c.K. and r.F. supervised the study. c.J.h., J.c.K., a.c.G., G.l.a., F.K., and P.P. supervised the contributing cohorts. Y.M., K.l.B., J.c.K., i.V., and r.F. visualized the data and results. The original draft was written by Y.M., K.l.B., J.c.K., r.F., and i.V. all authors reviewed and edited the draft. **Competing interests:** a.c.G. has received consulting fees from astraZeneca, unrelated to this project. a.i.n. and F.Y. receive royalties from the University of Michigan for the sale of MSFragger and ionQuant software licenses to commercial entities. all license transactions are managed by the University of Michigan innovation Partnerships office, and all proceeds are subject to a university technology transfer policy. The other authors declare that they have no competing interests. **Data and materials availability:** all data associated with this study are present in the paper or the Supplementary Materials. raw and processed timsToF MS data are available on the Proteomics identification database (Pride) (accession id PXd039875). rna-seq and microarray gene expression data for GainS study samples are available in the european Genome-Phenome archive (eGad00001008730) and arrayexpress (e-MTaB-4421, e-MTaB-4451, e-MTaB-5273, and e-MTaB-5274). raw data from figures are in data file S3.

GAinS Investigators

in addition to the GainS investigators who are authors (Y.M., K.I.B., P.h., c.G.G., e.e.d., S.M., a.c.G., c.J.h., and J.c.K.), the following GainS investigators, listed alphabetically by institution, were involved in patient recruitment, sample collection, or sample processing:

nigel Webster[12], helen Galley[12], Jane Taylor[12], Sally hall[12], Jenni addison[12], Sian roughton[12], heather Tennant[12], achyut Guleri[13], natalia Waddington[13], dilshan arawwawala[14], John durcan[14], alasdair Short[14], Karen Swan[14], Sarah Williams[14], Susan Smolen[14], christine Mitchell-inwang[14], emily errington[15], Maie Templeton[15], Pyda Venatesh[16], Geraldine Ward[16], Marie Mccauley[16], Simon Baudouin[17,33], charley higham[17], Jasmeet Soar[19], Sally Grier[19], elaine hall[19], Stephen Brett[19], david Kitson[19], robert Wilson[19], laura Mountford[19], Juan Moreno[19], Peter hall[20], Jackie hewlett[20], christopher Garrard[21], Julian Millo[21], duncanYoung[21], Penny Parsons[21], alex Smiths[21], roserFaras-arraya[21], Jasmeet Soar[22], Parizade raymode[22], Jonathan Thompson[23], Sarah Bowrey[23], Sandra Kazembe[23], natalie rich[23], Prem andreou[23], dawn hales[23], emma roberts[23], Simon Fletcher[24], Melissa rosbergen[24], Georgina Glister[24], Jeronimo Moreno cuesta[25], Julian Bion[26], Joanne Millar[26], elsa Jane Perry[26], heather Willis[26], natalie Mitchell[26], Sebastian ruel[26], ronald carrera[26], Jude Wilde[26], annette nilson[26], Sarah lees[26], atul Kapila[27], nicola Jacques[27], Jane atkinson[27], abby Brown[27], heather Prowse[27], anton Krige[28], Martin Bland[28], lynne Bullock[28], donna harrison[28], Gary Mills[29,30], John humphreys[29, 30], Kelsey armitage[29, 30], Shond laha[31], Jacqueline Baldwin[31], angela Walsh[31], nicola doherty[31], Stephen drage[32], laura ortiz-ruiz de Gordoa[32], Sarah lowes[32], charley higham[33], helen Walsh[33], Veritycalder[33], catherineSwan[33], heatherPayne[33], davidhiggins[34], Sarahandrews[34], SarahMappleback[34], chris Garrard[35,36], d. Watson[35,36], eleanor Mclees[35,36], alice Purdy[35,36], Martin Stotz[37], adaeze ochelli-okpue[37], Stephen Bonner[38], iain Whitehead[38], Keith hugil[38], Victoria Goodridge[38], louisa cawthor[38], Martin Kuper[39], Sheik Pahary[39], Geoffrey Bellingan[40], richard Marshall[40], hugh Montgomery[40], Jung hyun ryu[40], Georgia Bercades[40], Susan Boluda[40], andrew Bentley[41], Katie Mccalman[41], Fiona Jefferies[41], andrew Kwok[1], narelle Maugeri[1], Jayachandran radhakrishnan[1], and alice allcock[1].

affiliations 1 to 11 can be found on the first page of the paper.

[12]aberdeen royal infirmary, aberdeen aB25 2Zn, UK.
[13]Blackpool Victoria hospital, Blackpool FY3 8nr, UK.
[14]Broomfield hospital, chelmsford cM1 7eT, UK.
[15]charing cross hospital, london W6 8rF, UK.
[16]coventry and Warwickshire University hospital, coventry cV2 2dX, UK.
[17]Freeman hospital, newcastle upon Tyne ne7 7dn, UK.
[18]Frenchay hospital, Bristol, UK and Southmead hospital, Bristol BS16 1Je, UK.
[19]hammersmith hospital, london W12 0hS, UK. [20]huddersfield royal infirmary, huddersfield hd3 3ea, UK. [21]John radcliffe hospital, headington, oxford oX3 9dU, UK. [22]Kettering General hospital, Kettering nn16 8UZ, UK.
[23]leicester royal infirmary, leicester le1 5WW, UK.
[24]norfolk and norwich University hospital, norwich nr4 7UY, UK.
[25]north Middlesex hospital, london n19 1QX, UK. [26]Queen elizabeth hospital, Birmingham B15 2GW, UK. [27]royal Berkshire hospital, reading rG1 5an, UK.
[28]royal Blackburn hospital, Blackburn BB2 3hh, UK. [29]royal hallamshire hospital, Sheffield S10 2JF, UK. [30]northern General hospital, Sheffield S5 7aU, UK. [31]royal Preston hospital, Preston Pr2 9hT, UK.
[32]royal Sussex county hospital, Brighton Bn2 5Be, UK. [33]royal Victoria infirmary, newcastle upon Tyne ne1 4lP, UK. [34]Southend hospital, Westcliff-on-Sea SS0 0rY, UK.
[35]St Bartholomew's hospital, london ec1a 7Be, UK.
[36]royal london hospital, london e1 1Fr, UK.
[37]St Mary's hospital, london W2 1nY, UK.
[38]James cook University hospital, Middlesbrough TS4 3BW, UK.
[39]Whittington hospital, london n18 5nF, UK.
[40]University college london hospital, Uclh, london nW1 2BU, UK.
[41]Wythenshawe hospital, Manchester M23 9lT, UK.