

Supplemental Materials

A machine learning and centrifugal microfluidics platform for bench-side prediction of sepsis.

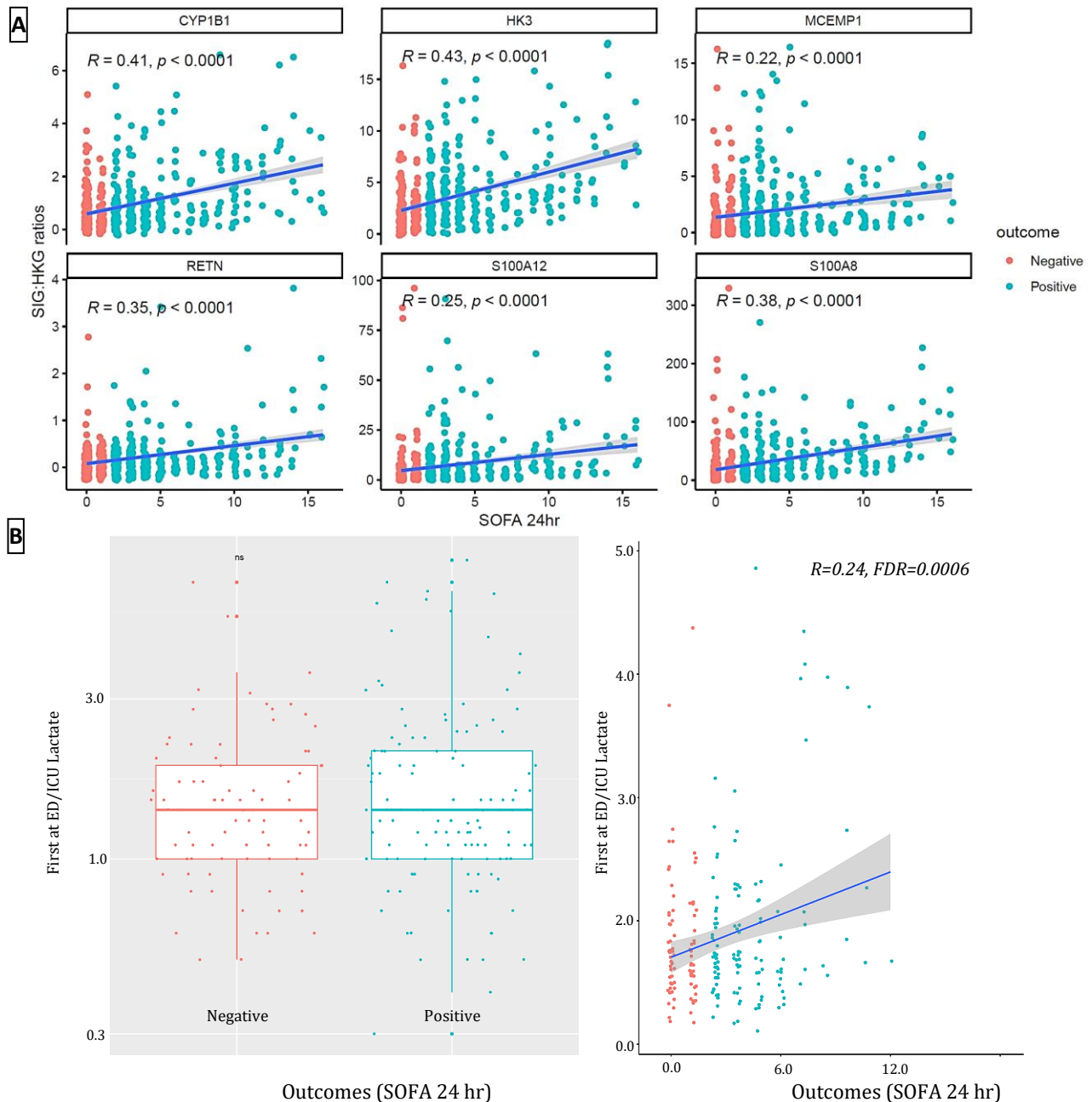
Lidija Malic, first author

Claudia C. dos Santos^{4,2,#} Corresponding author

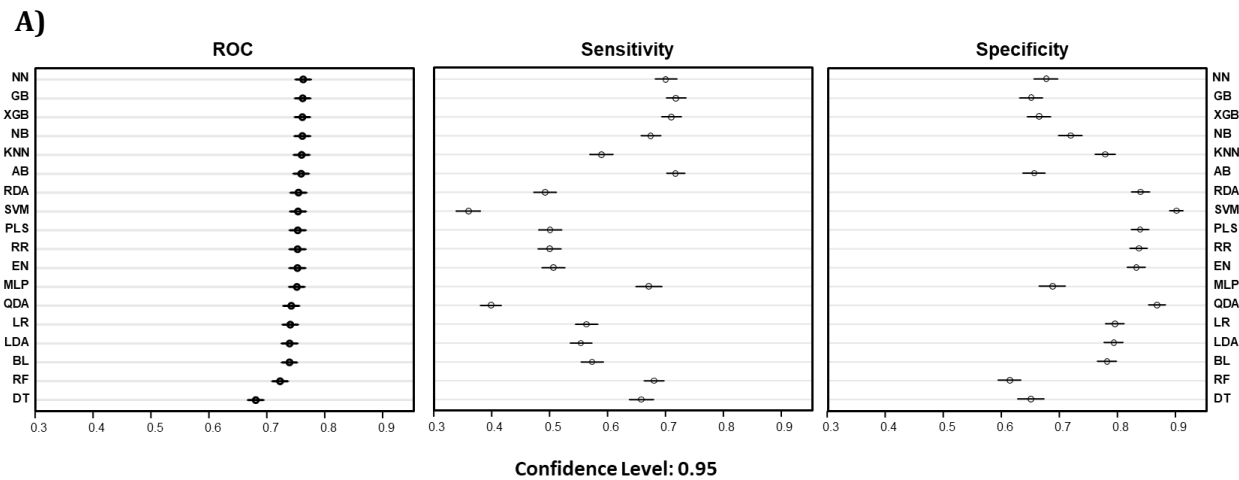
Supplemental Figure 1: Analysis of Sepset signature genes (data from GSE185263). A)

Relationship between 24-hour SOFA score and the relative gene expression of each of the Sepset-signature genes demonstrating a relative increase in expression as a function of 24-hour SOFA score. Red = Sepsis negative SOFA <2; Blue = sepsis positive SOFA ≥ 2 . The lines are drawn by linear regression with the statistical data using Pearson's Correlation Coefficient cited in each subfigure.

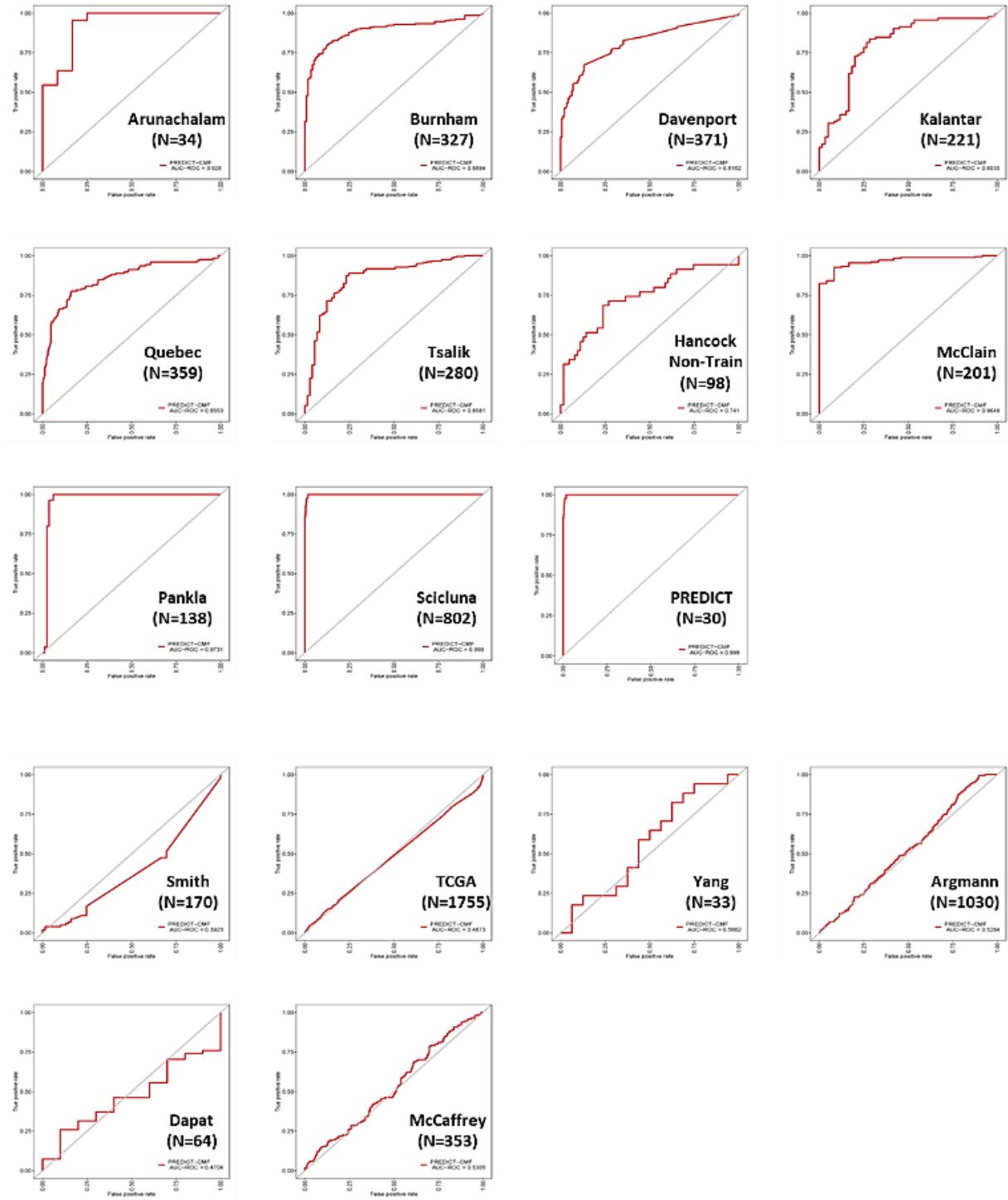
B) Relationship between baseline SOFA and serum lactate with clinical deterioration. Boxplots show no statistical (Wilcoxon rank-sum test) differences between lactate level between sepsis samples (positive = SOFA ≥ 2) vs. non-sepsis samples (negative = SOFA < 2). The scatterplot shows poor correlation (Pearson's Correlation Coefficient) between lactate level vs. SOFA score ($R = 0.24$, suggesting lactate, as a marker of clinical severity, accounted for $\sim 5\%$ variation in SOFA scores).



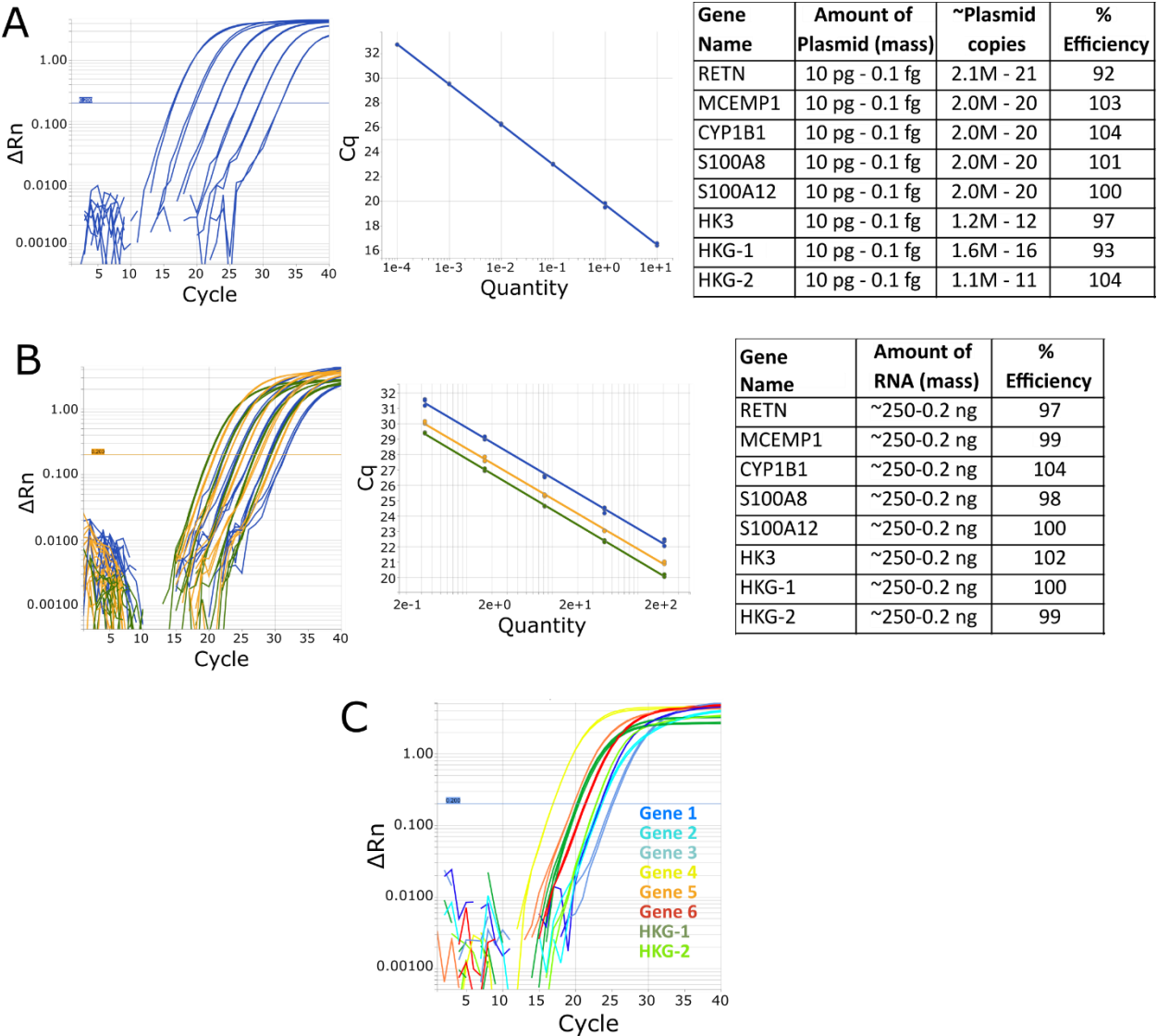
Supplemental Figure 2: Selection of the Sepset (and 2 HKG) gene expression signature (data from GSE185263). Schematic of the patient cohort analyzed and the methodology of machine learning/algorithm development to predict clinical outcomes based on expression levels. We used an ensemble of single feature selection algorithms to identify the optimal biomarker gene set in terms of both predictive performance and robustness. **A)** AUC-ROC of Training data set. The ensemble approach combined results from thirteen different selectors: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Red Deer Algorithm (RDA), Gradient Boosting Machine (GBM), Neural Network (NN), Partial Least Squares regression (PLS), Lasso Regression (LASSO), Elastic net (ELASTIC), Radio Frequency (RF), Background subtraction (BG), Naïve Bayesian (NB), Linear Discriminant Analysis (LDA), and Flexible Gradient Boosting (GB and XGB). **B)** The AUROC for the ability of the Sepset classifier to correctly classify label-free samples from data-sets downloaded from GEO (not used in the derivation) with known group assignment (septic vs non-septic as defined by suspicion of sepsis + SOFA>2).



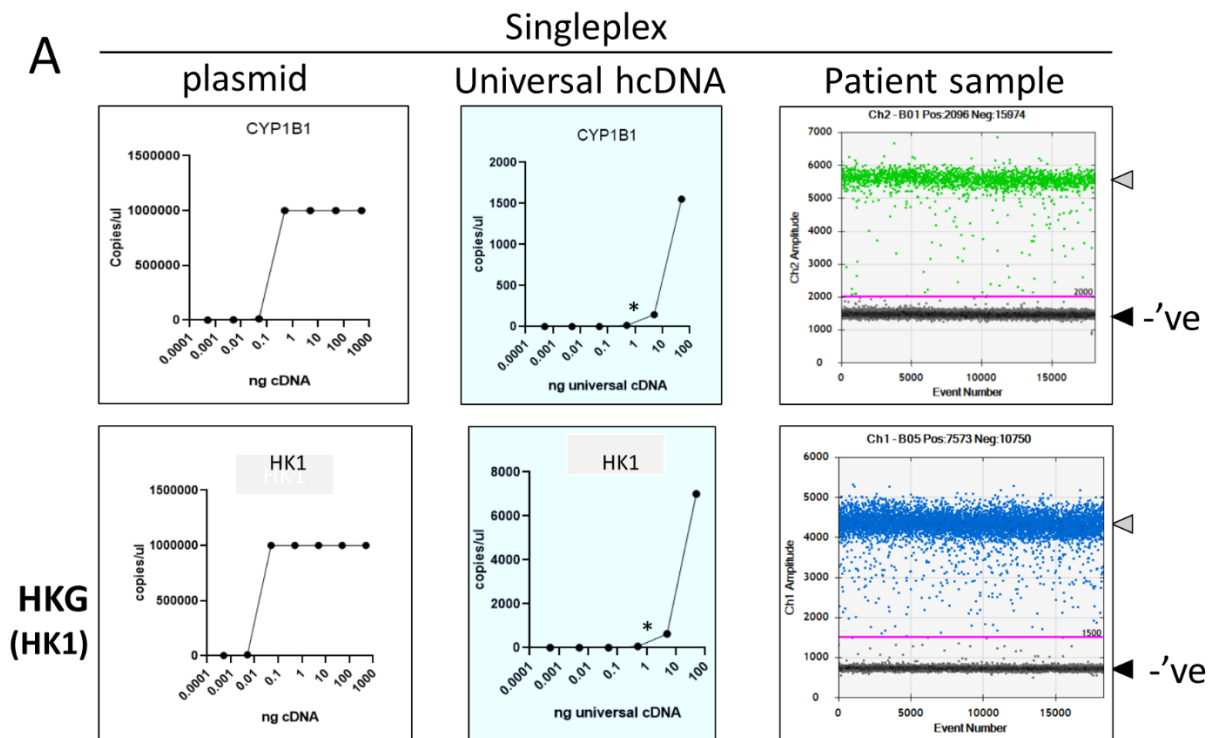
B)



Supplemental Figure 3: Development and performance summary of RT-qPCR primers and probes used in the Sepset signature. PCR primers were developed for each sepsis gene signature gene as well as two HKGs and their reaction efficiency was evaluated in a single-plex reaction using gene specific plasmids as the reaction template (A). Shown is the representative amplification curves and corresponding standard curve for the MCEMP1 target along with a table summarizing the reaction efficiencies across all genes. In all cases, the efficiency was close to 100% and consistent RT-qPCR results could be obtained with as low as 10-20 copies of template in the reaction. The PCR primers were then combined in triplex reactions (B) and their reaction efficiencies were determined using RNA isolated from healthy donor blood (n=4 replicates). Shown is a representative amplification curve and corresponding standard curve for one of the multiplex reaction wells. Reaction efficiencies for all the RT-qPCR reactions were close to 100% and consistent results could be obtained with as little as ~0.2 ng of RNA used as input for the RT-qPCR reaction. A representative amplification curve for the RT-qPCR assay showing the expression pattern of all six signature genes as well as the two housekeeping genes is also shown (C).

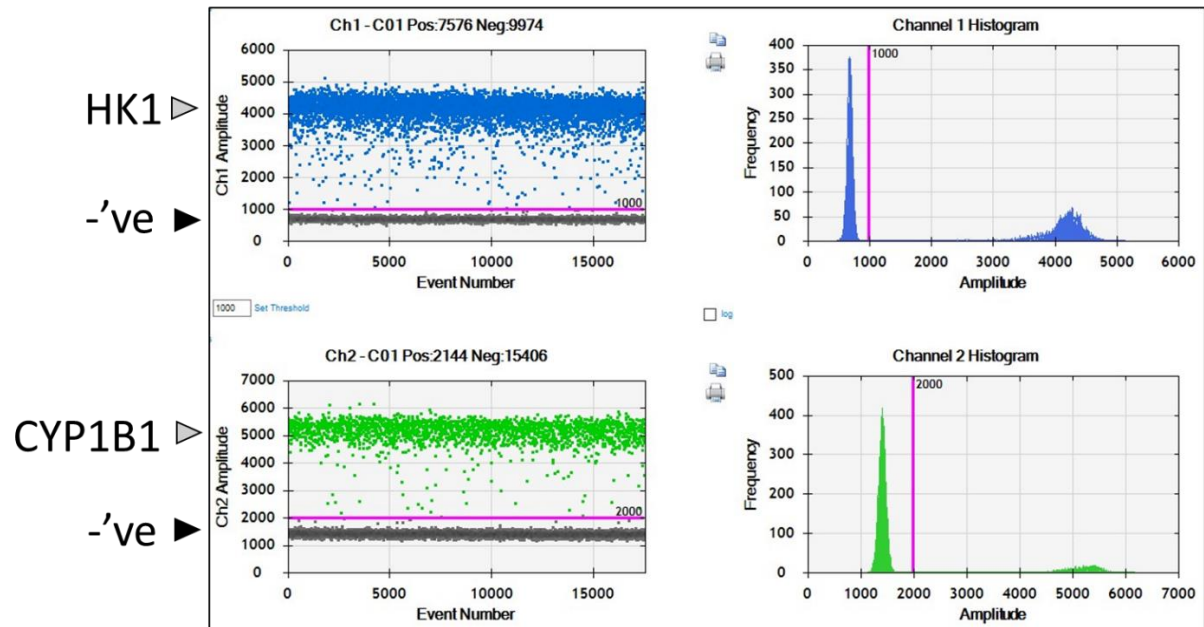


Supplemental Figure 4: Benchmarking of Sepset signature probes. A. (A. Left panels) Gene specific plasmids were obtained and purified and used as template for the probes using the BioRad AutoDG and QX200 ddPCR system (“gold standard”). Shown are CYP1B1 as a representative for the genes of interest (top) and HK1 as representative housekeeping gene (HK1, bottom). Serial dilutions of the plasmid were used to determine LOI and dynamic range of the probes. (A. Middle panels) Serial dilutions of universal human cDNA were used as template for “gold standard” ddPCR to determine the concentration of input (500 pg) for optimal profiling of patient samples. (A. Right panels) Representative single-plex (CYP1B1 and HK1) sample tracings from a patient sample showing probe signal (positive partition) and negative partition at optimal (500 pg) input concentration. B. Probes were utilized in duplex using (500 pg) patient samples and counts were assessed from each patient, in triplicate. (left panel) Sample tracings showing partitioning of positive and negative droplets. Thresholding was performed on the 2D tracing (C. top right panel) and clear demarcation of double negative droplets (bottom left quadrant), HK1 positive droplets (top left quadrant), CYP1B1 positive droplets (bottom right quadrant) and double positive droplets (top right quadrant) are apparent. Probe alone control rendered only double negative droplets (C. bottom right panel).



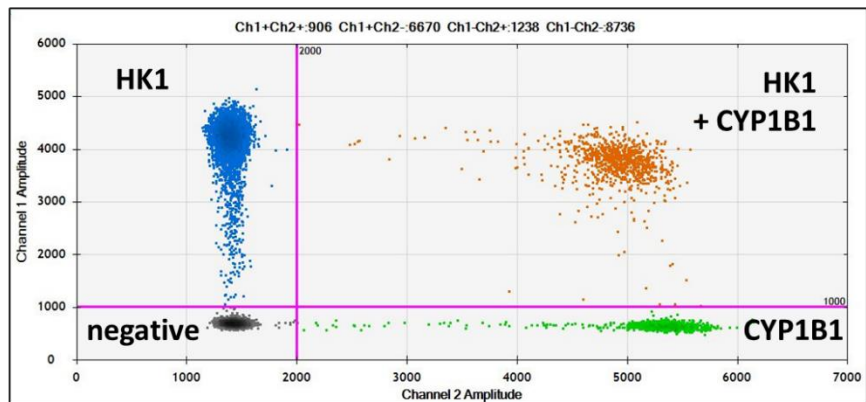
B

Multiplex on patient sample

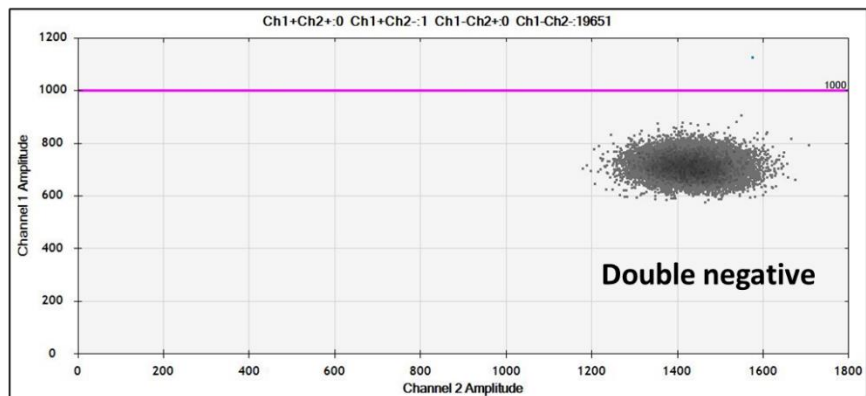


C

HK1 +
CYP1B1



Probe alone



Supplemental Figure 5: Microfluidic implementation. The platform executes a predefined sequence of spinning and pneumatic actuation steps to automate the entire analytical protocol on the cartridge. These steps are illustrated by the sequence of schematic images showing the cartridge's status at various points during the automated RNA extraction protocol and subsequent ddPCR (see Supplemental Tables S8 and S9 for details). Arrows indicate the direction of fluid flow at each step. Activated ports are highlighted with green and red circles, representing the application of positive and negative pressures, respectively.

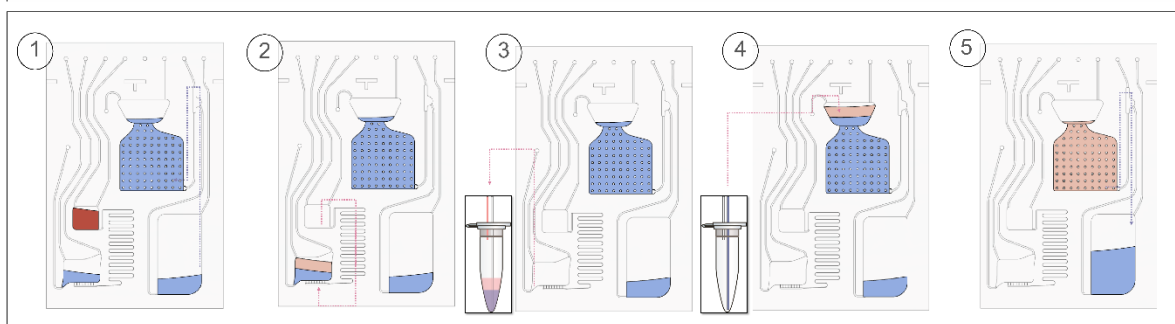
(A) RNA Extraction: The automated RNA extraction protocol begins with the introduction of the sample into the RNA extraction chamber and the installation of the cartridge on the platform (step 0). The software then executes the pre-programmed sequence, initiating rotation of the platform. The first step involves the transfer of Proteinase K solution to the extraction chamber (step 1a), followed by bubble mixing (steps 1b and 1c). Lysis/binding buffer containing magnetic nanoparticles (MNPs) is then transferred to the sample (step 2a), mixed (step 2b), and incubated for 10 minutes at 55 °C (step 2c). The rotation speed is increased to capture MNPs (step 2d), and the lysate is transferred to the waste chamber (step 3). Two wash steps follow: each wash solution is transferred to the extraction chamber (steps 4a and 5a), and then moved to waste (steps 4c and 5c). Finally, elution buffer is transferred to the extraction chamber (step 6a), mixed, and incubated with MNPs (step 6b). The rotation speed is increased to capture MNPs (step 6c), and purified RNA is transferred to a clean tube (step 7).

(B) ddPCR Assay: To initiate the cDNA synthesis and ddPCR protocol, a 2 μ L aliquot of the eluted RNA is introduced into the PCR mix chamber of the ddPCR cartridge, containing the RT-ddPCR master mix. Two cartridges, each capable of performing a single fourplex ddPCR reaction, operate in parallel to detect the 8-gene classifier. The automated sequence begins with the transfer of fluorinated oil into the droplet imaging chamber (step 1), followed by emulsification of the RT-ddPCR master mix in the droplet generation chamber (step 2). This emulsification is achieved by applying positive pressure to the ports of the PCR mix chamber, pushing liquid through the resistive serpentine channel into an array of nozzles that merge into a deep reservoir in the droplet generation chamber. Once droplet generation is complete, rotation speed is reduced, and positive pressure is applied to transfer the emulsion off-chip into the PCR tube on the platform heater (step 3). After thermal cycling, the emulsion is transferred back onto the chip by applying positive pressure (step 4). The droplets are then arranged into a monolayer suitable for imaging by applying low negative pressure at the oil reservoir ports (step 5). This process withdraws the fluorinated oil into the reservoir and gently lowers the droplets into the shallower portion of the chamber. The pressure is gradually reduced to 0 psi until monolayer formation is complete, at which point the rotor stops, allowing fluorescence images to be captured for subsequent analysis.

A

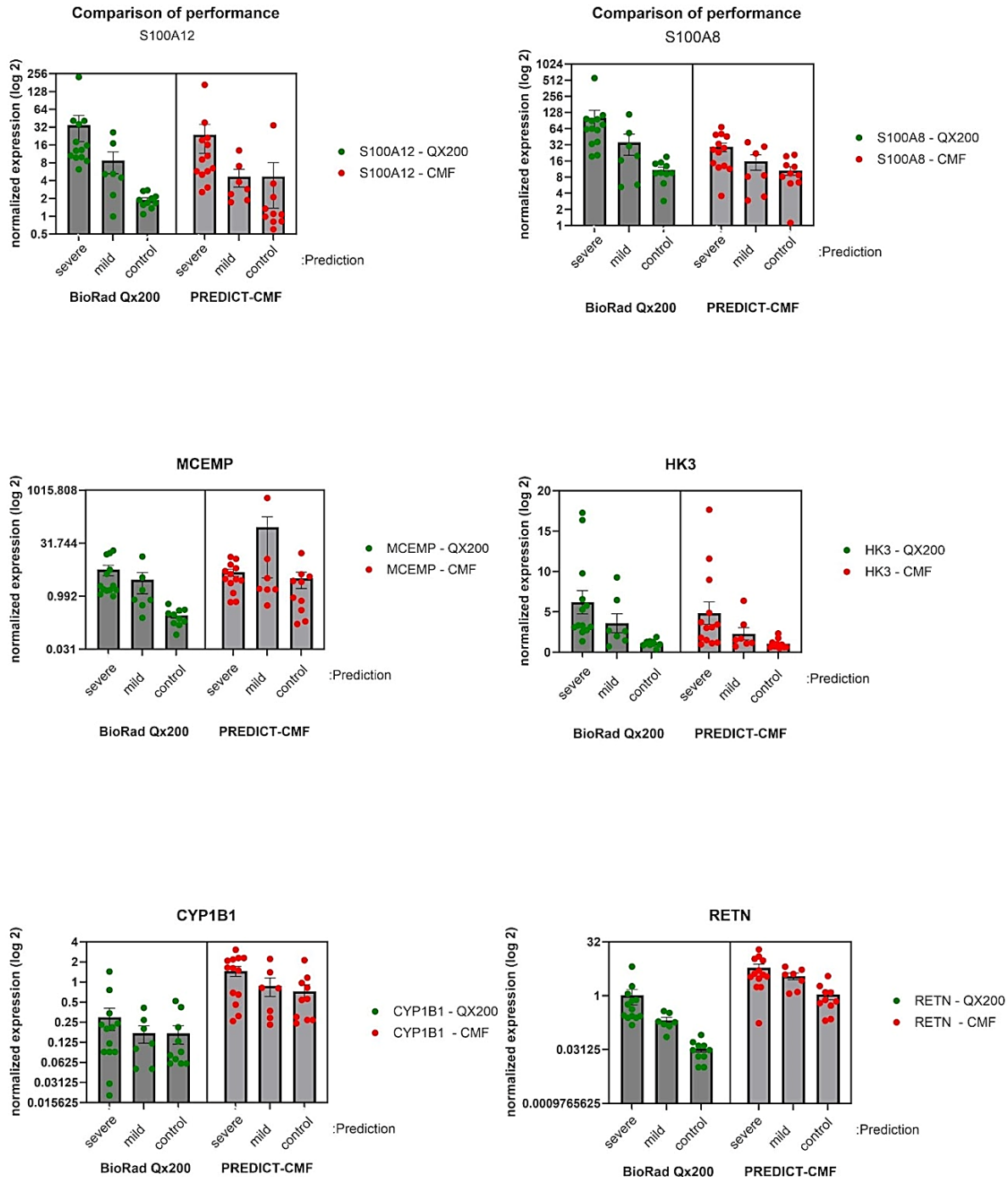


B



Supplemental Figure 6: Performance of the PREDICT system compared to BioRad QX200 ddPCR.

RNA from the whole blood of ICU patients or healthy controls was isolated, reverse transcribed and the two platforms (BioRad QX200 (green) and PREDICT (red)) quantitated (copies/ μ L) the expression of the (6) signature genes of interest (GOI) and (2) housekeeping genes (see Supplemental Data 2). The raw counts of the GOI (labelled at top of plots) were normalized to the GEOMEAN of the HKG [PTP4A2, CHTOP] and were plotted on $\log_2$ scale. ICU patient samples were classified as “severe or mild” based on their SOFA scores (>2 , severe and <2 , mild) with the healthy controls.

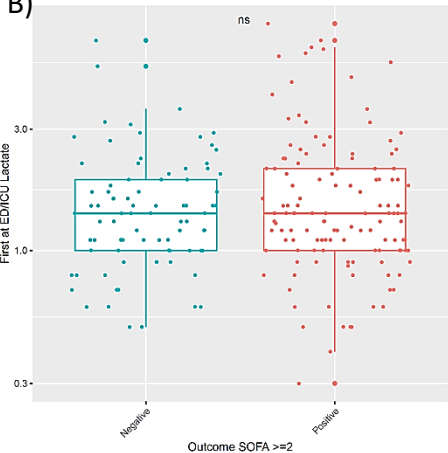


Supplemental Figure 7: Performance analysis with lactate levels vs. SepsetER and the combination (3 models; data from GSE185263). A) Table of samples that were used to investigate how the Sepset classifier compares to lactate in predicting clinical deterioration within 24 hours. B) Lactate levels at first presentation show no significant difference (Wilcoxon rank-sum test) between sepsis (Positive) vs. non-sepsis (Negative) cohorts. C) AUC curves for 2 models validated with Hancock emergency department samples: Lactate model showed AUC=0.5 which indicated no predicative power for lactate levels. Sepset^{ER} model showed AUC=0.75 suggesting 3 out of 4 predictions would be correct. Lactate is usually used as a marker for mortality. It is possible that lactate would have higher prediction value for the development of sepsis later in hospitalization, but in these ER patients it had no discriminatory ability.

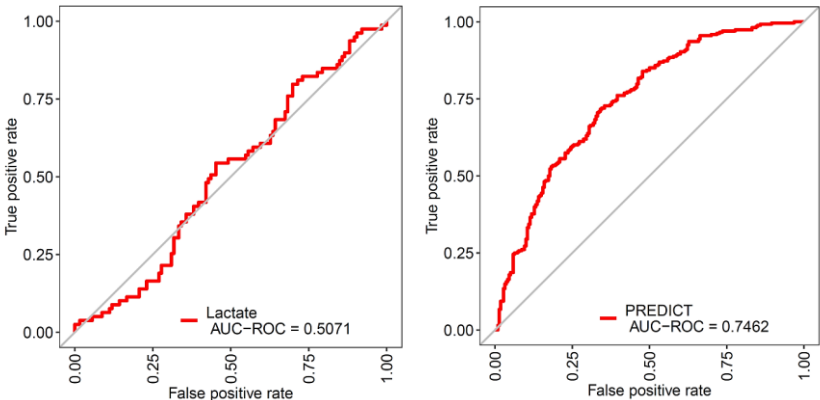
A)

Model	Training set	Sepsis (N)	Non-sepsis (N)	Total (N)	Predictors
Lactate	Hancock lactate	126	79	205	Lactate level
Sepset ^{ER}	Hancock	268	220	488	CR6

B)



C)



Supplemental Tables

Supplemental Table 1: Data sets used for derivation, validation and testing of classifier

Dataset group	Number	Patient location	Platform	Positive class (N)	Negative class (N)	Reference and Accession Number
Training & Cross Validation	586	Australia, Colombia, Netherland, USA, Canada	RNA-Seq	24hr SOFA ³ 2 (N=271)	24hr SOFA <2 (N=243)	Baghela 2022; GSE185263
Testing/ Validation	359	Quebec, Canada	RNA-Seq	SOFA ³ 2 (N=234)	SOFA <2 (N=125)	Baghela 2023; GSE185263/222393
	34	USA	RNA-Seq	Sepsis (N=12)	Non-sepsis (N=22)	Arunachalam 2020; GSE152418
	221	USA	RNA-Seq	Sepsis (N=129)	Non-sepsis (N=92)	Kalantar 2022; GSE189400
	802	Netherland, UK	Microarray	Sepsis (N=760)	Non-sepsis (N=42)	Scicluna 2017; GSE65682
	371	UK	Micro-array	SRS 1 (N=145)	SRS 2 (N=226)	Davenport 2016; E-MTAB-4421, -4451
	327	UK	Micro-array	SRS 1 (N=118)	SRS 2 (N=209)	Burnham 2016 E-MTAB-5273, -5274
	201	USA	RNA-Seq	Sepsis (N=24)	Non-sepsis (N=177)	McClain 2021; GSE161731
	280	USA	Micro-array	Sepsis (N=73)	Non-sepsis (N=207)	Tsalik 2016; GSE63990
	138	Thailand	Micro-array	Sepsis (N=83)	Non-sepsis (N=55)	Pankla 2009; GSE69528
In house data sets for RT-PCR, ddPCR and POC device validation	238	Canada	qPCR	24hr SOFA ³ 2 (N=170)	24hr SOFA <2 (N=68)	qPCR
	30	Canada	ddPCR	24hr SOFA ³ 2 (N=12)	24hr SOFA <2 (N=18)	ddPCR
	30	Canada	POC device	24hr SOFA ³ 2 (N=12)	24hr SOFA <2 (N=18)	POC device
Negative Control datasets						
Cardiogenic shock	33	Taiwan	Micro-array	Deceased (N=16)	Survived (N=17)	Yang 2022; GSE93101
Cancer	1755	Multiple countries	RNA-Seq	Stage 3 or 4 tumors (N=496)	Stage 1 or 2 tumors (N=1259)	TCGA ^a ; https://www.cancer.gov/ccg/research/genome-sequencing/tcga
IBD	1030	USA	RNA-Seq	IBD (N=819)	Healthy (N=211)	Argmann 2023; GSE186507
CAD	353	USA	RNA-Seq	High/Mid CAD (N=189)	Low CAD (N=164)	McCaffrey 2023, GSE180083/221911
Bacterial infection	170	UK	Micro-array	Infected (N=69)	Non-infected (N=101)	Smith 2014; GSE25504
Viral infection	64	Japan	RNA-Seq	Infected (N=54)	Non-infected (N=10)	Dapat 2020; GSE155925

Supplemental Table 2: Selection of genes for the Sepset signature. Presented are analyses performed per gene on 443 patient samples comprising 213 with SOFA scores ≤ 2 and 230 with SOFA scores ≥ 2 at 24 hours post-presentation. Shown are the final selection of 6 genes based on specificity, fold change and adjusted p-values (padj). Increased FCs were associated with eventual worsening SOFA score (≥ 5). All genes were previously reported as part of the 99-gene cellular reprogramming/endotoxin tolerance signature.

HGNC symbol	Fold Change SOFA ≥ 2	Specificity	P-value adjusted	Fold Change SOFA ≥ 5
RETN	4.28	0.77	1.46E-13	11.39
MCEMP1	2.23	0.69	1.40E-05	5.50
HK3	2.16	0.69	1.45E-09	3.77
S100A8	2.49	0.65	5.76E-09	6.57
CYP1B1	2.08	0.64	4.39E-08	2.03
S100A12	2.49	0.63	5.01E-07	6.01

Supplemental Table 3: Putative function of genes in the Sepset signature: The HUGO Gene Nomenclature Committee (HGNC) official gene symbol. Gene name and putative function in GeneCards the Human Gene database and joint functional prediction in Enrichr (Chen, E.Y., Tan, C.M., Kou, Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013). <https://doi.org/10.1186/1471-2105-14-128>) for which the adjusted p-value and odds ratio describe the likelihood that functional enrichment prediction using REACTOME or GO Molecular function was identified by chance alone.

HGNC symbol	Gene Name	Function
RETN	Resistin	Promotes chemotaxis in myeloid (immune) cells. Antimicrobial activity in low salt buffers.
MCEMP1	Mast Cell Expressed Membrane Protein 1	Based on expression pattern speculated to be involved in regulating mast cell differentiation or immune responses
HK3	Hexokinase 3	Catalyzes the phosphorylation of hexose, such as D-glucose and D-fructose, to hexose 6-phosphate. Initial step in glycolysis which is involved in immune functions.
S100A8	S100 Calcium Binding Protein A8	Calcium- and zinc-binding protein that plays a prominent role in the regulation of inflammatory processes and the immune response. It can induce neutrophil chemotaxis and adhesion.
CYP1B1	Cytochrome P450 Family 1 subfamily B member 1	Monooxygenase involved in the metabolism of various endogenous substrates, including fatty acids, steroid hormones and vitamins. Involved in arachidonic metabolism and plays a role in immunity.
S100A12	S100 Calcium Binding Protein A12	Calcium-, zinc- and copper-binding protein that plays prominent role in the regulation of inflammatory processes and the immune response. Pro-inflammatory activity involves recruitment of leukocytes, promotion of cytokine and chemokine production, and regulation of leukocyte adhesion and migration. Acts as an alarmin.
All 6 genes together	Sepset signature	REACTOME pathway: Neutrophil degranulation R-HSA-6798695 (adjusted p value = 0.00023; odds ratio 84). GO Molecular Function RAGE receptor binding GO:0050786 (adj. p=0.000038, odds ratio = 1666)

Supplemental Table 4: Baseline demographic and clinical characteristics and correlation with clinical traits, interventions and outcomes at recorded at 24 and 72 h. Data were collected from consenting adult patients (> 18 years of age) with ethics approval, who presented with prospective sepsis, within the first two hours of emergency room (ER) or within 24 h of intensive care unit (ICU) admission at baseline, 24 h and 72 h. Abbreviations: ICU (intensive care unit), SOFA (sequential organ failure score), WBC (white blood cell count), P/F ratio (ratio of the partial pressure of oxygen over fraction of inspired oxygen), IQR (interquartile range). A significant correlation between features and Sepsis is denoted by $p \leq 0.05$ using χ^2 or Fisher exact test for

Demographics and Characteristics	Baseline Median IQR	24 hrs	72 hrs
Age	52.5 ± 22.75	ns	1.60E-05
Sex, Female	45.50%	ns	2.66E-02
Number of Comorbidities	2 ± 4.75	4.50E-03	9.99E-09
Location ICU	31%	5.90E-03	8.80E-05
SOFA score	4 ± 4.5	4.70E-05	3.00E-08
Glasgo Coma Scale (GCS)	13 ± 2	ns	2.90E-04
Heart Rate (HR, beats per minute)	82.5 ± 23.5	ns	ns
Mean Arterial Pressure (mmHg)	81 ± 20	ns	ns
Temperature (Degrees Centigrade)	36.2 ± 0.77	ns	1.10E-02
Oxygen Saturation (Percent)	96 ± 4.25	ns	4.90E-02
Total Serum Bilirubin (μmol/L)	10 ± 4.5	ns	9.70E-06
WBC Count (10 ⁹ /L)	8.58 ± 7.92	ns	8.10E-11
Platelets Count (10 ⁹ /L)	207 ± 127.5	ns	1.40E-03
Serum Creatinine (μmol/L)	68 ± 136.25	ns	3.30E-03
P/F Ratio	227 ± 118.5	ns	ns
D-dimers (μg/L)	1,198 ± 949	ns	ns
C-Reactive Protein (mg/L)	97.3 ± 157.4	ns	ns
Serum Lactate (mmol/L)	1.7 ± 1.4	ns	1.20E-03
Interventions and Outcome Measures	Baseline Median IQR	24 hrs	72 hrs
Antibiotics given	74%	1.08E-02	8.39E-03
MV initiated	28%	2.16E-02	4.71E-05
Fluid resuscitation received	84%	4.10E-02	2.38E-02
Oxygen Therapy received	89%	ns	2.20E-03
Urine output at presentation	92%	ns	ns
Inotropic Support (days)	0 ± 7	ns	ns
Length of ICU Stay (days)	4 ± 9	5.90E-03	8.80E-05
Hospital Length of Stay (days)	9 ± 2	4.10E-09	4.10E-09
Suspected site of infection Lung	62%	ns	1.60E-05
Suspected site of infection Abdomen	33%	ns	5.77E-02
Suspected site of infection CNS	0%	ns	ns
Blood Culture Pathogen	18%	ns	5.89E-06
28-Day Mortality	17%	ns	2.62E-07
90-Day Mortality	2%	ns	ns

categorical variables and analysis of variance for continuous variables as appropriate with Benjamini and Hochberg correction; ns = not significant.

Supplemental Table 5: Primers and Probes used for ddPCR

Gene	Primer/Probe	Sequence (5' à 3')
CYP1B1	Forward	GGC TGG ATT TGG AGA ACG TA
	Reverse	CAT CAG GAT ACC TGG TGA AGA G
	Probe	/56-FAM/ACT ATC ACT /ZEN/GAC ATC TTC GGC GCC /3IABkFQ/
PTP4A2	Forward	TTA TAA GAC AAA AAA GAA GGG GAG C
	Reverse	CCT TCG TTT ACA TTT CCT TCT ACT G
	Probe	/5HEX/CCG ACC TAA /ZEN/GAT GCG ATT ACG CTT CAG A/3IABkFQ/
MCEMP1	Forward	GGG AGC TTT GGA ATG TCT CA
	Reverse	GCC TGC TAA TGT CGT CTC TAA T
	Probe	/5TexRd-XN/TTC CGT TCA GCA GAG CAT CAC CAT /3IAbRQSp/
S100A12	Forward	GCT TGC AAA CAC CAT CAA GAA TA
	Reverse	CAA TGG CTA CCA GGG ATA TGA A
	Probe	/5Cy5/CCA AGG CCT /TAO/GGA TGC TAA TCA AGA TGA /3IAbRQSp/
CHTOP	Forward	TGT CTC TAA ATG AGC GCT TTA CT
	Reverse	TCT TCT GTT TCT GGC ACT GG
	Probe	/56-FAM/ACA AAC AGC /ZEN/CGA CGC CAG TGA ATA /3IABkFQ/
HK3	Forward	GGT GAC TCT AAC TGG CAT TGA
	Reverse	GCA AAG TCA AAG AGC TGC TG
	Probe	/5HEX/ATC ACA AAC /ZEN/TCC TGG CTT CTG GGC /3IABkFQ/
RETN	Forward	CCT AAT ATT TAG GGC AAT AAG CAG
	Reverse	GGG CAA GTA GCC AGG TC
	Probe	/5TexRd-XN/CAT TGG CCT GGA GTG CCA GA/3IAbRQSp/
S100A8	Forward	GTG TCC TCA GTA TAT CAG GAA AAA GG
	Reverse	CTG CCA CGC CCA TCT TTA TC
	Probe	/5Cy5/GTG CAG ACG /TAO/TCT GGT TCA AAG AGT TGG A/3IAbRQSp/

Supplemental Table 6: Characteristics of reservoirs implemented in the RNA extraction cartridge fluidic design and corresponding reagents

Reservoir #	Description	Location on cartridge	Capacity (µL)	Reagent	Reagent volume (µL)
1	RNA extraction	Top side	980	Sample: PBS (1:1)	100
2	Proteinase K	Bottom side	70	Proteinase K	20
3	Lysis/binding	Bottom side	1,125	Lysis/binding buffer	400
4	Waste	Bottom side	3,500	–	–
5	Wash 1	Top side	1,020	Wash 1 buffer	550
6	Wash 2	Top side	1,020	Wash 2 buffer	550
7	Elution	Top side	70	Elution buffer	25

Supplemental Table 7: Characteristics of reservoirs implemented in the ddPCR cartridge fluidic design and corresponding reagents

Reservoir #	Description	Capacity (μL)	Reagent	Reagent volume (μL)
1	Oil	305	Fluorinated oil	100
2	Droplet imaging*	83 + 15	Imaged emulsion	15
3	PCR mix	53	PCR master Mix + Eluted RNA	50
4	Droplet generation	75	Emulsion (Oil + droplets)	70 (20 + 50)

Table 8: Microfluidic implementation of the automated RNA extraction protocol

Step	Operation	Active ports (#)	Applied pressure	Rotation speed (rpm)	Duration
			(psi) ^a		
1a	Transfer of Proteinase K	3	1	600	15 s
1b	Bubble mixing of sample and Proteinase K	2	1	600	0.5 s (10×)
1c	Incubation	–	–	400	1 min
2a	Transfer of Binding buffer	8	1	600	15 s
2b	Bubble mixing of sample and MNPs	2	1	600	0.5 s (10×)
2c	Incubation at 55 C	–	–	400	10 min
2d	Capture of MNPs	–	–	800	5 min
3	Transfer of lysate to waste chamber	2	–0.4 to –1.4	800	1 min
4a	Transfer of wash buffer 1	7	2.5	500	15 s
4b	Incubation	–	–	400	5 min
4c	Transfer of wash buffer 1 to waste chamber	2	–0.4 to –1.4	800	1 min
5a	Transfer of wash 2 buffer	6	2.5	500	20 s
5b	Incubation	–	–	400	2 min
5c	Transfer of wash buffer 2 to waste chamber	2	–0.4 to –1.4	800	1 min
5d	Air dry of MNP pellet	2	–1.5	800	1 min
6a	Transfer of elution buffer	4	1.9	700	5 s
6b	Bubble mixing of elution buffer and MNPs	2	1	700	0.2 s (30×)
6c	Capture of MNPs	–	–	800	5 min
7	Transfer of eluted RNA sample to an external tube	1	–1.5	600	30 s

^a Values are relative to atmospheric pressure.

Supplemental Table 9: Microfluidic implementation of the automated ddPCR protocol

Step	Operation	Active ports (#)	Applied pressure	Rotation speed (rpm)	Duration
			(psi) ^a		
1	Fill the imaging chamber with oil	8	1 to 2	400	30 s
2	Droplet generation	3,4	3	600	30 min
3	Transfer droplets to external PCR tube	1,2	1 to 3	300	1 min
4	Thermal cycling	–	–	400	2 h
5	Transfer droplets from external PCR tube to cartridge	6,7,8	–1 to –2.5	400	1 min
6	Monolayer droplet formation in imaging chamber	7,8	–0.3	300	3 min

^a Values are relative to atmospheric pressure.