

25

26

27

Description of Graphic Abstract: Feature reduction and development of a gene classifier that predicts deterioration risk groups in patients starts with in-house RNA sequencing data from patient collected from a heterogeneous cohort of patients with suspected sepsis (top left) to reduce our original published gene signature down to 6-genes (Sepset), for which expression could be related to 2 housekeeping genes. Feature selection was performed using machine learning (ML) and AI and the classifier validated in samples from published transcriptomic studies. Molecular assay is then developed by designing and testing primer/probe sequences specific to the target genes using digital droplet PCR. In parallel, sample-to-answer microfluidic platform and cartridges are developed (bottom right) and analytical performance of multiplex quantitative assay is tested. Prognostic enrichment is obtained by analyzing the results using ML algorithm to determine the percent likelihood of significant clinical deterioration within the immediate next 24 h. The deployment of PREDICT platform (center) at the point-of-care is anticipated to aid in triage and management of prospective sepsis within the first 3 h of clinical presentation.

28 **Using machine learning and centrifugal microfluidics at the point-of-need to predict clinical**
29 **deterioration of patients with suspected sepsis within the first 24 h.**

30 Lidija Malic^{1,2,3‡}, Peter G.Y. Zhang^{4‡}, Pamela Plant^{5‡}, Liviu Clime¹, Christina Nassif¹, Dillon Da
31 Fonte¹, Evan E. Haney⁴, Byeong-Ui Moon¹, Victor Sit¹, Daniel Brassard¹, Maxence Mounier¹, Eryn
32 Churcher⁵, James T. Tsoporis⁵, Reza Falsafi⁶, Manjeet Bains⁶, Andrew Baker⁵, Uriel
33 Trahtemberg^{5,7,8}, Ljuboje Lukic¹, John C. Marshall⁵, Matthias Geissler¹, Robert E.W. Hancock^{4,6},
34 Teodor Veres^{1,2,2,9}, and Claudia C. dos Santos^{4,2,#}

- 35
36 1. Life Sciences Division, National Research Council of Canada, 75 de Mortagne Boulevard,
37 Boucherville, QC, J4B 6Y4, Canada.
38 2. Center for Research and Applications in Fluidic Technologies (CRAFT), University of Toronto, 5
39 King's College Rd. Toronto, ON, M5S 1A8, Canada.
40 3. Department of Biomedical Engineering, McGill University, 775 Rue University, Suite 316,
41 Montreal, QC, H3A 2B4, Canada.
42 4. Sepset Biosciences Inc., 420 – 730 View St, Victoria, BC, V8W 3S2, Canada.
43 5. Keenan Research Centre for Biomedical Science, St. Michael's Hospital, University of Toronto,
44 Critical Care Medicine, 30 Bond Street, Toronto, ON, M5G 1W8, Canada.
45 6. Centre for Microbial Diseases and Immunity Research, University of British Columbia, 232-2259
46 Lower Mall, Vancouver, BC, V6T 1Z4, Canada.
47 7. Department of Critical Care, Galilee Medical Center, Nahariya, Israel.
48 8. Medicine Faculty, Bar Ilan University, Zafed, Israel.
49 9. Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College
50 Road, Toronto, ON, M5S 3G8, Canada.

51
52

53 ‡Authors contributed equally to the manuscript

54
55 # Corresponding author: Claudia.santos@utoronto.ca; Claudia.dossantos@unityhealth.to
56 St. Michael's Hospital
57 30 Bond Street, Bond Wing Room 4-008,
58 Toronto, ON, M5B 1WB, Canada.
59 Tel: (416)-864-6060 (x3198)
60 Fax: (416)-864-6013

61
62
63

64 **Key words:** sepsis, prediction, prognostication, risk stratification, transcriptomics, biomarkers,
65 microfluidics, point-of-care

66

67 **Abstract**

68 Sepsis is the body's dysfunctional response to infection associated with organ failure. Delays in
69 diagnosis have a substantial impact on survival. Herein, samples from 586 in-house patients were
70 used in conjunction with machine learning and cross-validation to narrow a gene expression
71 signature of immune cell reprogramming to predict clinical deterioration in patients with
72 suspected sepsis within the first 24 hours (h) of clinical presentation using just six genes (Sepset).
73 The accuracy of the test (~90% in early intensive care unit (ICU) and 70% in emergency room
74 patients) was validated in 3,178 patients from existing independent cohorts. A real-time reverse
75 transcriptase polymerase chain reaction (RT-PCR)-based test was shown to have a 98% sensitivity
76 in >230 patients to predict worsening of the sequential organ failure scores or admission to the
77 ICU within the first 24 h following Sepset detection. A stand-alone centrifugal microfluidic
78 instrument that integrates the entire automated workflow for detection of the Sepset classifier
79 in whole blood using digital droplet PCR was developed and tested. This PREcision meDIcine for
80 CriTical care (PREDICT) system had a high sensitivity of 92%, specificity of 89%, and an overall
81 accuracy of 88% in identifying the risk of imminent clinical deterioration in patients with
82 suspected sepsis.

83 Introduction

84 Sepsis, a complex syndrome of organ dysfunction caused by a dysregulated host response to
85 infection,¹ has been declared a global emergency.² Estimates from 2017 are >48.9 million cases
86 and 11.0 million deaths per year,³ not including deaths from COVID-sepsis.⁴ Early individualized
87 interventions^{5,6} may significantly reduce mortality and morbidity, and prevent poor long-term
88 outcomes and disability.^{5,6,7} It has been shown that even short delays in appropriate treatment
89 can cause significant increases in mortality from sepsis.⁷ Despite decades of research, however,
90 patients are still triaged and treated on the basis of clinical symptoms.⁸ While these include
91 measures of overall severity, they are largely nonspecific and do not adequately assess
92 dysregulated responses to infection, align patients with appropriate pharmacotherapies, or
93 predict impending deterioration and the need for resuscitative level care. Biomarkers that
94 provide early prognostic and therapeutic enrichment are actively sought to achieve the
95 personalization necessary to improve research and care.^{2,9}

96 Importantly, while some diagnostic tests have moved to a 'distributed' model with
97 relevant tests performed at bedside (e.g., glucose monitoring for diabetes), for patients with
98 sepsis, the vast majority of biochemical and molecular tests are implemented in a 'centralized'
99 format, found only in well-equipped laboratories that require trained operators and specialized
100 instruments, making them less accessible (or even inaccessible) to some of our most vulnerable
101 and difficult to access populations.^{8,9,10,11} The need to send samples to the lab for analysis
102 significantly delays results (some more than the 6 h recommended by the surviving sepsis
103 guidelines)¹² and limits the ability to provide timely care to patients.^{6,13,14} This 'bottleneck' means
104 that testing standards fail one of the primary objectives of the World Health Organization (WHO)
105 "to promote health, ...and serve the vulnerable so everyone, everywhere can attain the highest
106 level of health."¹⁵

107 Lab-on-a-chip (LOC)^{16,17,18,19} technologies have the potential to advance care beyond
108 traditional 'syndromic' approaches, outside specialized centers, promising to democratize care
109 for millions of patients.^{20,21,22} In addition to performing assays in a compact, miniaturized format,
110 LOC devices promote portability and point-of-care (POC) testing, enabling minimally trained

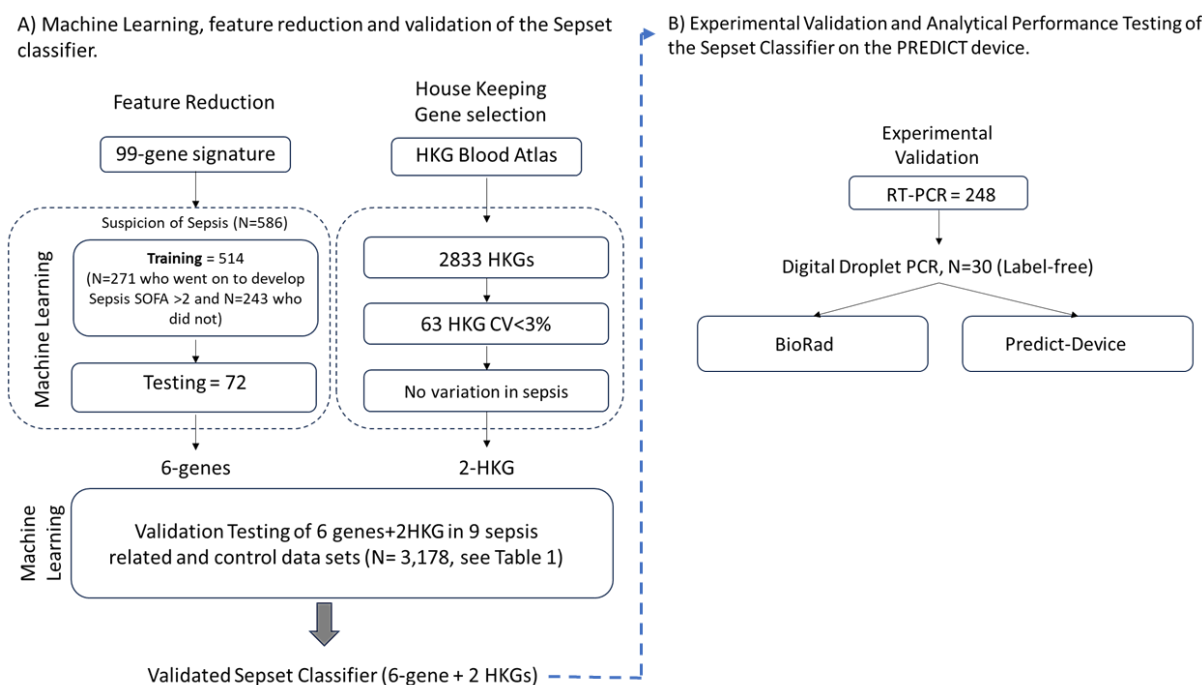
111 personnel to perform analytical procedures outside laboratory settings. We have previously
112 shown that extraction and detection of molecular markers from biological samples, such as
113 blood, can be fully automated using a centrifugal-based LOC system.²³ We have further
114 integrated bioanalytical assays for pathogen detection involving polymerase chain reaction
115 (PCR)²⁴ and loop-mediated isothermal amplification in a sample-to-answer format.²⁵ Combining
116 our LOC system with innovative RNA-based biomarkers that predict clinical
117 deterioration^{26,27,28,29,30} may significantly advance sepsis care in specialized and non-specialized
118 settings alike.

119 In parallel, our group discovered and validated a 99 gene signature, present within 2 h of
120 presentation to the emergency department (ED), able to predict clinical deterioration based on
121 the emergence of a cellular reprogramming profile associated, with the inability of cells to
122 respond to pathogens.^{31,32} Our original expression signature was pathogen agnostic, predicting
123 both all-cause and COVID sepsis and different immunological response endotypes. The signature
124 was also predictive of organ dysfunction and severity outcomes.^{30,33} Here, we hypothesized that
125 a reduced gene signature, would retain the ability to discriminate patients with suspected sepsis
126 at high risk of clinical deterioration. The ability of this signature to classify patients defined by
127 Sepsis-3 criteria¹ was validated in 3,178 patient samples from nine published transcriptomic
128 studies. Analytical performance of the primers and the ability of the signature to classify label-
129 free patients with suspected sepsis into deterioration risk groups, defined by a worsening
130 sequential organ failure assessment (SOFA) score and the need for ICU admission within the first
131 24 h, was determined in an independent cohort by routine semi-quantitative real-time reverse
132 transcriptase polymerase chain reaction (RT-PCR, N=248). We then developed the molecular
133 assay to perform quantitative one-step reverse transcriptase digital droplet PCR (ddPCR)
134 detection of the RNA-based classifier at the POC using our LOC centrifugal microfluidic (CMF)
135 system.^{23,25,34} A compact prototype instrument, the **PRE**cision **meDI**cine for **CriTi**cal care
136 (**PREDICT**) device, that detects the Sepset classifier in less than 3 h using 50 μ L of whole blood in
137 the near-patient-environment was developed. Finally, we used PREDICT to classify patients into
138 clinical deterioration risk groups using an independent cohort of label-free patients with
139 suspected sepsis (N=30).

140 **Results**

141 A high-level overview of the risk prediction classifier, platform development path and
 142 deployment workflow are shown in **Figure 1**. Three groups of independent cohorts
 143 (**Supplemental Table S1**) were used to reduce our original 99-gene signature down to 6 genes.
 144 We then validated, in a blinded fashion, the classifier’s ability to accurately identify amongst 248
 145 well characterized patients, those who went on to deteriorate clinically, as defined by worsening
 146 of the sequential organ failure assessment (SOFA ≥ 2) score and the need for ICU admission, 24
 147 h post-initial assessment (early for patients with prospective sepsis).

Figure 1: Overview of prediction classifier development, validation and analytical performance testing



148

149 Our own whole blood RNA sequencing (RNA seq) data from 586 samples from 514 individuals
 150 (176 ICU and 338 ER patients); including 392 previously published³⁰ and 194 new patient samples,
 151 were used to refine the gene expression signature able to predict clinical deterioration in patients
 152 with suspected sepsis.^{30,31,35} Briefly, these data were generated from whole blood tubes
 153 collected from consenting adult patients (> 18 years of age) with ethics approval, who presented

154 with prospective sepsis, within the first 2 h of emergency room (ER) or within 24 h of intensive
155 care unit (ICU) admission (see supplemental material) as previously published.^{30,31,35}

156 Each of the previously published 99 cellular reprogramming signature genes³⁰ were tested for
157 their ability to discriminate between 271 patients with suspected infections who went on to
158 record a SOFA score >2 in the first 24 h of presentation and the 243 who did not. We used 514
159 first samples (out of total of 586) for machine learning (ML) model training and the remaining 72
160 second samples (out of 586) from the same 514 individuals as well as samples from 50 healthy
161 individuals as validation samples. We selected six upregulated genes with the highest specificity,
162 fold-change (FC) and lowest adjusted p-value (adj p-value, **Supplemental Table 2**). Increased FCs
163 were associated with eventual worsening of SOFA scores as demonstrated in **Supplemental**
164 **Figure S1A** showing the relative expression values for each of the putative signature genes as a
165 function of the 24-hr SOFA score. This revealed a relationship between the relative gene
166 expression of each signature gene and SOFA score ($p < 0.01$ to $p < 0.0001$). **Supplemental Figure**
167 **S2B**, shows no relationship between lactate level and increased SOFA scores in sepsis samples
168 (SOFA ≥ 2) vs. non-sepsis samples (SOFA < 2). Scattering of expression values around the mean
169 implied that no single gene was discriminatory on its own, justifying the need for a multi-gene
170 risk-prediction expression signature. The six-gene signature arose from the larger immune
171 dysfunction (cellular reprogramming) signature.³¹ Genes comprising the signature and their
172 abbreviated putative biological functions are shown in **Supplemental Table 3**.

173 To ensure we did not need to include positive and negative controls in the prognostic assay, the
174 expression of each of our 6 genes was compared to two housekeeping genes (HKG) selected from
175 the list of the 2,833 HKGs expressed in blood cells.³⁶ Their expression variance (coefficient of
176 variation, CV) was analyzed in the entire data set and 63 genes showing CV < 3% across all of our
177 datasets were selected as candidate HKG. These were filtered for lack of variance according to
178 the clinical metadata considering sex, age, SOFA score, location (ICU vs ER), sepsis mortality,
179 prediction and endotype³⁰, as well as lack of variance in cancer patients (**Tables 1 and S1**). The
180 accuracy of prediction using ratios of individual genes to HKG was tested. While no single gene
181 expression ratio can discriminate those patients that are imminently at risk of clinical

182 deterioration from those that are not, the collective set of 6 genes, when compared to 2 HKGs,
 183 gave excellent performance (**Table 1**).

184 To determine the best ML model to predict clinical deterioration within the first 24 h following
 185 clinical presentation, the input transcriptomic and associated clinical outcome data were
 186 analyzed using an ensemble of ML approaches. In patients with suspected sepsis, a SOFA cut-off
 187 score of 2 was used to discriminate predicted sepsis from non-sepsis.¹ Using a change of SOFA
 188 score of >2 did not substantially change the results. Groups of patients were analyzed in a binary
 189 fashion (prospective sepsis cf. non-sepsis), using 18 machine learning algorithms, initially
 190 adopting a 10X cross-validation strategy. We chose to proceed with eXtreme Gradient Boosting
 191 (XGBoost)³⁷, a regularizing gradient boosting framework/library involving matrices of decision
 192 trees, since it gave us the best performance (**Table 1 and Supplemental Figure S2**). Negative
 193 control data sets were used to reduce the likelihood of identifying false positives - patients with
 194 non-septic pro-inflammatory, shock, malignancy and infectious (bacterial or virus non-septic)
 195 conditions.

RNA-Seq Dataset	AUC	Sensitivity	Specificity	Precision	NPV	Accuracy	Reference
COVID-19 Sepsis (N=359)	0.85	85%	72%	85%	71%	79%	An ³⁸
ICU Validation (N=176) ¹	0.9	92%	82%	88%	88%	87%	This study
ER Validation (N=338) ¹	0.69	70%	59%	63%	67%	65%	This study
Scicluna (N=802)	0.99	97%	90%	99%	61%	94%	Scicluna ²⁶
Davenport (N=371)	0.83	86%	62%	59%	88%	74%	Davenport ²⁹
Burnham (N=327)	0.92	87%	83%	74%	92%	85%	Burnham ³⁹
Kalantar (N=152)	0.81	83%	65%	61%	86%	74%	Kalantar ⁴⁰
McClain (N=201)	0.96	92%	91%	58%	99%	91%	McClain ^{41,42}
Tsalik (N=280)	0.86	75%	88%	70%	91%	82%	Tsalik ⁴³
Pankla (N=138)	0.97	94%	100%	100%	92%	97%	Pankla ⁴⁴
Arunachalam (N=34)	0.89	83%	95%	91%	91%	89%	Arunachalam ⁴⁵
Aggregated Data (N = 3,178)	0.88	87%	78%	80%	83%	83%	

Negative Control Data Sets							
Stage 3/4 Cancer (n=1,755)	0.5	0%	100%	-	63%	50%	TCGA/GDC ⁴⁶
Cardiogenic shock (n=33)	0.53	0%	100%	-	52%	50%	Yang 2022 ⁴⁷
Coronary Artery Disease (N= 353) ²	0.51	3%	100%	100%	51%	51%	McCaffrey ⁴⁸
Inflammatory Bowel Disease (N=1030) ³	0.53	0%	100%	-	21%	51%	Argmann ⁴⁹
Bacterial infection (N= 170) ²	0.39	83%	9%	38%	43%	46%	Smith ⁵⁰
Viral infection (N=64) ³	0.48	0%	93%	0%	83%	46%	Dapat ⁵¹

196 ¹ These datasets were distinct subsets of the training set and performance criteria might be affected by overfitting

197 ² McCaffrey⁴⁸; 353 patients with coronary artery disease comparing 189 positives (MID/MID+/CAD) to 164 negatives
198 (LOW CAD).

199 ³ Argmann⁴⁹; comparing 819 inactive/Mild/Moderate/Severe IBD blood-based transcriptomes to 211 healthy
200 controls

201

202 **Table 1: Summary Performance for Sepset signature and classifier.** Performance of the Sepset signature was
203 evaluated in different validation and testing datasets comprised of transcriptomic analyses of samples collected from
204 various regions and settings. Aggregated results include analysis of patient samples subjected to transcriptomic
205 analysis and available in the literature (see specific references). The internal data sets, samples indicated as “This
206 Study”, were from an independent subset of patients from the Biomarkers of Lung Injury study (ClinicalTrials.gov ID
207 NCT04747782), a prospective, observational, clinical study that collected whole blood from patients who presented
208 to the ICU or ward with acute respiratory distress and suspected sepsis. Definitions: NPV = Negative Predictive Value;
209 ICU, intensive care unit; ER emergency room.

210 ***A six-gene signature classified patients with sepsis in independent cohorts and was robust***
211 ***over time and a range of disease severity***

212 We determined the ability of the six-gene compared to two-HKGs signature to accurately classify
213 patients into pre-specified sepsis groups using nine sepsis-related gene expression datasets
214 (N=3,178 (all samples) – 514 (training, or ICU + ER) = 2,664) downloaded from the Gene
215 Expression Omnibus (GEO) repository. In these independent external cohorts, the reduced
216 signature was able to discriminate between septic versus non-septic patients (Sepsis-3 criteria¹),
217 with a sensitivity ranging from 83-97% and balanced accuracy of 74-94%, a combined mean
218 sensitivity of 85%, a specificity of 76%, and an accuracy of 81% (**Table 1**). We then analyzed
219 samples from patients that were seen in the ER with suspicion of sepsis (N=338) versus those
220 admitted directly to the ICU from the training cohort (N=176); this yielded sensitivity assessments
221 of 70% and 92% respectively.³⁰ To determine if the signature was robust over time and a range

222 of disease severity, we assessed its performance in a validation cohort from the Biobanque
223 Québécoise de la COVID-19 (BQC-19; Quebec COVID-19 Biobank) comprised of 359 blood
224 samples taken at various times from 133 patients hospitalized with COVID-sepsis and exhibiting
225 a range of severity.^{33,38} The ability of the reduced signature to discriminate between samples
226 from patients that went on to clinically deteriorate (increased SOFA >2 or ICU admission) within
227 the first 24 h from initial assessment, compared to those that did not was very good, with
228 sensitivity and balanced accuracy of 85% and 79%, respectively. The set of six genes and two
229 HKGs comprising the signature, we called here **Sepset**, was able to discriminate patients with
230 suspicion of sepsis at risk of clinical deterioration.

231
232 ***Sepset RNA amplification using RT-PCR reliably identifies patients that clinically deteriorated***
233 ***24 h after initial assessment***

234 To validate changes in gene expression using RT-PCR, we designed sets of PCR primers that
235 spanned intron-exon boundaries (i.e., only contained in RNA not DNA), that amplified specific
236 regions for each of the mRNA isoforms selected by the ML analysis (>90% of mRNAs), avoiding
237 common single nucleotide polymorphisms (SNPs), for each of the 6 signature and 2 HKGs. Control
238 experiments demonstrated that our primers successfully amplified mRNAs of interest extracted
239 from human blood in an RNA copy number dependent manner, amplified as few as 10-20 copies
240 of the gene and gave Ct curves in <25 cycles with blood mRNA. The PCR primers were then
241 combined in triplex reactions and tested using whole blood RNA extracts from healthy donors.
242 Reaction efficiencies were 100% with as little as 0.2 ng of input RNA (**Supplemental Figure S3**).
243 We then tested the Sepset signature in a blinded fashion, using label-free whole blood samples
244 from our in-house cohort of patients with prospective sepsis not included in the training set. As
245 shown in **Table 1**, in 248 patients, Sepset was able to identify patients who went on to deteriorate
246 with worsening SOFA score (fulfilling sepsis diagnosis based on SOFA ≥ 2 and suspicion of
247 infection) within the first 24 h after sampling for RT-qPCR analysis, to those who did not. In these
248 independent samples, Sepset demonstrated an AUC of 0.88 and a sensitivity of 94%.

249

250 ***Association between Sepset with clinical features of clinical deterioration 24 h after initial***
251 ***assessment***

252 In our 238 patients, we looked at the association between presence of Sepset with demographic
253 characteristics, clinical features of sepsis, severity, and outcome measures at baseline, at 24 h
254 (day 1) and 72 h (day 3) post-initial assessment (**Supplemental Table 4**). In the first 24 h, the
255 signature was correlated with the number of comorbidities, SOFA score, the use of antibiotics,
256 mechanical ventilation, and fluid resuscitation, but not oxygen or vasopressor therapy.
257 Interestingly, we found no association with specific clinical parameters of organ dysfunction until
258 day 3 (72 h) post-initial assessment. At 72 h, culture positivity and 28-day mortality correlated
259 with the presence of the signature. At both 24 h and 72 h post-initial assessment, the signature
260 was associated with important clinical outcome parameters such as the need for ICU admission,
261 initiation of mechanical ventilation, use of fluid resuscitation, and initiation of antibiotics, as well
262 as outcome parameters such as ICU length of stay. Unfortunately, we did not have data on
263 dialysis requirements, but while presence of urine output at baseline was not associated, an
264 increase in serum creatinine correlated with Sepset expression.

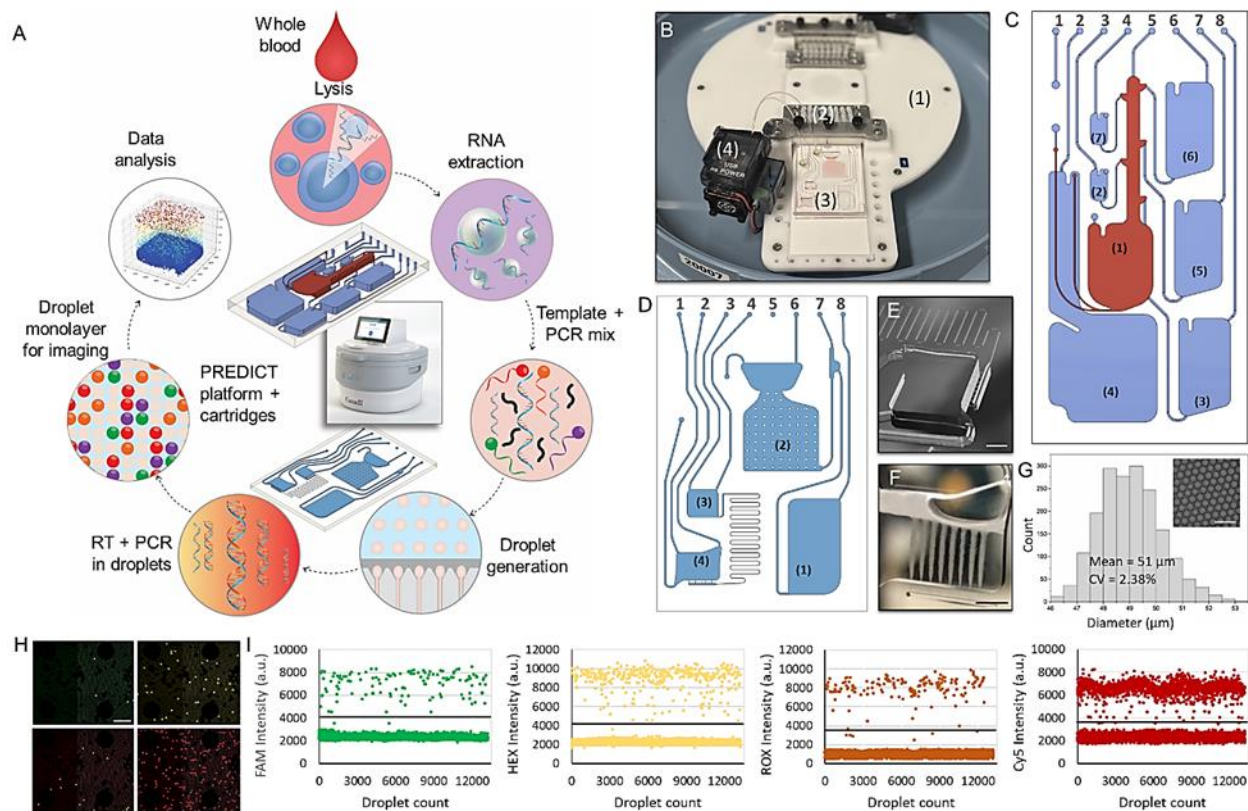
265
266 ***Digital droplet PCR detected the Sepset classifier in 0.5 ng of patient-derived whole blood RNA***

267 To migrate precision RNA-mark detection onto the LOC device, we designed the microfluidic
268 platform to measure RNA expression using ddPCR. Herein, primers that spanned exon-exon
269 junctions, were designed for each of the 8 genes (6 Sepset and 2 HKG) in the signature. The
270 performance of each primer was benchmarked using the BioRad AutoDG/QX200 reader. Serial
271 dilutions of human cDNA plasmids obtained from the mammalian genome collection⁵² were used
272 to determine limits of detection (1 copy/ μ l) and dynamic range of the probes for each of the
273 genes of interest (**Supplemental Figure S4A** left panel). Verification of primer performance was
274 conducted using universal human cDNA (**Supplemental Figure S4A** middle panel) and 500 pg of
275 patient cDNA (**Supplemental Figure S4A** right panels). A total of 500 pg of patient cDNA resulted
276 in optimized partitioning of the positive and negative droplets and quantitation of the Sepset
277 signature.

278 We then established the conditions for multiplexing using FAM and HEX labeled probes at final
279 concentration of 5 μ M (**Supplemental Figure S4B** left panel). Representative 2D tracing
280 (**Supplemental Figure S4B** bottom right panel) showed partitioning of double negative (bottom
281 left quadrant), FAM only (upper left quadrant), HEX only (lower right quadrant) and double
282 positive (upper right quadrant) droplets. Probes alone (no cDNA) as a control rendered only
283 double negative droplets (bottom right panel). Counts (in copies/ μ L) for each gene were obtained
284 in triplicate and used to further refine the training of the predictive algorithm. Primers and probes
285 used for ddPCR are presented in **Supplemental Table S5**.

286 ***PREDICT platform detected the Sepset classifier at the point-of-need***

287 Having validated the ability of the Sepset classifier to identify patients with ‘prospective’ sepsis,
288 i.e. patients that will go on to clinically deteriorate within the first 24 h of clinical presentation,
289 we sought to design and build an innovative POC device to perform the risk prediction at the
290 bedside. Our previously published microfluidic technology^{23,34} was adapted to perform the
291 molecular assay required to quantitate specific mRNAs using ddPCR at POC. The PREDICT system
292 was developed as a stand-alone instrument that integrates the entire automated workflow
293 (**Figure 2A**) for detection of the Sepset classifier directly from whole blood – without any pre-
294 processing requirement. The analytical process was split between two microfluidic cartridges
295 that when connected enable RNA isolation and ddPCR biomarker detection (see Supplemental
296 materials for details including video 1 and video 2) to be performed in sequence automatically.
297 Microfluidic cartridges are operated on a centrifugal platform (**Figure 2B**).^{25,53,54} Pneumatically-
298 induced pressure imbalances allowed for transfer, valving, flow switching, inward pumping, and
299 on-demand bubble-based mixing. The combination of centrifugation and pneumatic actuation
300 rendered manipulation of liquids independent of wetting properties, which allowed for
301 automation of the analytical protocol. The sample preparation cartridge (**Figure 2C**) contained
302 buffers and reagents (**Supplemental Table S6**) to isolate RNA from whole blood, which was then
303 used as a template for the ddPCR assay. The detection cartridge was adapted for monodisperse
304 droplet generation and visualization (**Figure 2D–G, Supplemental Table S7**).



305

306 **Figure 2: Microfluidic platform and cartridges.** A, Microfluidic workflow for automation of RNA extraction and
 307 downstream ddPCR on the PREDICT platform and related cartridges. B, PREDICT platform showing rotor (1),
 308 pneumatic manifold (2), cartridge installed on the rotor (3) and connected to the external PCR tube using world-to-
 309 chip tubing inserted into the wirelessly controlled platform heater (4). C, RNA extraction cartridge design. D, Image
 310 of the ddPCR cartridge design. E, Micrograph showing close-up view of droplet generation chamber and nozzles. The
 311 scale bar is 2.4 mm. F, Micrograph showing close-up view of droplet streams generated using the ddPCR cartridge.
 312 The scale bar is 2.4 mm. G Size distribution of droplet diameter. The inset shows optical micrograph of droplet
 313 monolayer using ddPCR cartridge. The scale bar is 150 μm. H Example of acquired fluorescence images showing
 314 droplet monolayer within a region of the imaging chamber. For clarity, only a zoomed-in portion of the imaging
 315 chamber region is shown to increase visibility of droplets. The scale bar is 400 μm. I Intensity maps for different
 316 fluorophores. Horizontal lines denote the threshold for positive and negative counts.

317

318 The analytical workflow was conducted through a timed sequence of centrifugation and
 319 pneumatic actuation steps for which operational parameters are provided in **Supplemental**
 320 **Tables S8 and S9. Supplemental Figure S5** illustrates the sample preparation and detection
 321 cartridges during different stages of the process. The PREDICT system used a miniature epi-
 322 fluorescence imaging module for readout of droplets which comprises single-color excitation
 323 LEDs and proprietary optical filters for detecting probes labelled with FAM, HEX, ROX and Cy5
 324 fluorophores. The instrument was adapted for recording fluorescence micrographs (**Figure 2H**)

325 for each channel in sequence using an embedded monochrome CCD camera equipped with a 2×
326 objective that had a sufficiently large field of view (e.g., 8 mm × 8 mm) to capture more than
327 10,000 droplets per image. The proprietary PREDICT software performed image analysis and
328 sample quantification using droplet scatter intensity map for each fluorophore (**Figure 2I**) and
329 ‘definetherain’ thresholding algorithm.^{55,56} Comparison of the performance of the PREDICT and
330 BioRad QX200 ddPCR is presented in **Supplemental Figure S6**. A proprietary algorithm considered
331 Sepset and HKG expression values and provided results as a probability of clinical deterioration
332 within the next 24h.

333 We tested and validated the performance of the PREDICT prototype using label-free real-world
334 samples from clinically annotated patients with prospective sepsis (N=30, Research Ethics Board
335 # 20-078). These analyses were performed in a blinded fashion using a subset of the 248
336 technically and clinically benchmarked patients. The PREDICT system was able to predict clinical
337 deterioration at presentation, as determined by worsening of SOFA score or need for ICU
338 admission, providing a risk assessment to physicians regarding the consequent need for initiation
339 of appropriate therapeutic (antibiotic, anti-viral, immune modulatory and/or monoclonal
340 antibody) or supportive (ICU entry, mechanical ventilation, fluids) measures post-presentation.
341 The sensitivity was 92%, specificity 89%, and overall accuracy 88% (**Table 2**), consistent with the
342 RT-qPCR assays described above.

343

RNA-Seq Dataset	AUC	Sensitivity	Specificity	Precision	NPV	Accuracy	Cohorts
RT-qPCR (N=248)	0.88	94%	72%	84%	88%	83%	This study
ddPCR (N=30)	0.91	100%	83%	80%	100%	92%	This study
PREDICT (N=30)	0.88	92%	83%	79%	94%	88%	This study

344

345 **Table 2: Label-free performance validation of Sepset classifier.** The PREDICT system was able to predict clinical
346 deterioration at presentation, as determined by worsening of SOFA score or need for ICU admission compared to
347 RT-qPCR and ddPCR. Definitions: RT-qPCR, reverse transcription quantitative polymerase chain reaction; ddPCR,
348 droplet digital RT-PCR.

349 Discussion

350

351 We have demonstrated the successful creation of a predictive sepsis gene expression signature
352 and its migration to a novel POC platform PREDICT. This proof-of-concept demonstrates the
353 feasibility of developing a rapid POC device that will enable physicians to conduct a prospective
354 sepsis risk assessment at the point-of-need (in as little as 1-2 h after clinical presentation). In
355 patients with suspicion of/or early sepsis, a positive result indicates a high probability that the
356 patient will clinically deteriorate in the next 24 h (worsening SOFA score ≥ 2 points and need for
357 ICU admission), guiding physicians as to the need for ICU entry (in ER patients) and interventions
358 like mechanical ventilation. Currently, the device provides a risk assessment in < 3 h, well below
359 the 6 'golden h' recommended by the surviving sepsis campaign guidelines.¹² In patients with
360 suspected sepsis, PREDICT efficiently performs the entire workflow of RNA signature-based
361 classifier detection, from sample-to-answer using just 50 μ L of whole blood, without requiring
362 pre-preparation of sample, and once transfer of eluate from the first to the second cartridge is
363 automated, with no involvement from the clinical laboratory. Using real-world label-free patient
364 samples, we showed the device has a high sensitivity of 92%, specificity of 89%, and an overall
365 accuracy of 88% to identify the risk of imminent clinical deterioration in previously benchmarked
366 patients with known outcomes.

367

368 There are two other established immune signature-based platforms, but neither appear to be
369 intended for use like the Sepset signature. Septicyte RAPID test⁵⁷ discriminates sepsis from non-
370 infectious systemic inflammation and the original SeptiCyt²⁷. This test, requires prior sample
371 preparation, including RNA extraction from blood, separates patients into 4 bands with varying
372 accuracy for prediction of sepsis cf. sudden SIRS, with only 55.8% being assigned to the highest
373 and lowest bands and AUCs of 0.81-0.85⁵⁷. In comparative studies, using Sepsis-3 definitions (96%
374 of tested patients had sepsis), the AUC was 0.81, below that for all cohorts with the Sepset
375 signature, except for ER patients not considered in the Septicyte study. Larger (29-31 genes)
376 signatures that predict bacterial vs. viral infections (IMX-BVN3) and sepsis mortality (IMX-SEV),
377 have been devised and again separate patients into 3 or 4 bands⁵⁸. A recent performance

378 study^{58,59} showed that the IMX-BVN3 signature adjudicated that, amongst patients with
379 confirmed bacterial infections (64% of patients), 31% were possible and 6% were unlikely.
380 Critically, however it is clear that both bacteria and specific viruses like COVID-19 can cause
381 mechanistically almost-identical sepsis, and a positive cultures is only found in around 50% of
382 patients^{60,61} even when using the latest molecular detection methods. Thus, discriminating
383 bacterial from viral infections at the host response level may be useful with regards to indicating
384 treatment options but not for identifying sepsis or identifying patients with suspicion of sepsis at
385 risk of clinical deterioration. Conversely the sepsis mortality signature IMX-SEV⁵⁸ shows a good
386 AUC of 0.81 amongst ICU patients but this is separate from the Sepset signature that does not
387 detect eventual mortality (Supplemental **Table S4**), in contrast to a separate signature we
388 devised³³. Overall, Sepset provides a simple binary risk assessment with high accuracy in very
389 large numbers of patients (**Table 1**).

390
391 In our clinical study, the lower sensitivity of the test in ER patients with suspected sepsis possibly
392 reflects the potential for early antibiotics to prevent development of organ dysfunction (74%
393 received antibiotics Supplemental Table S4).⁶² We found no statistically significant association of
394 Sepset prediction with presentation systolic or diastolic blood pressure, heart rate, respiratory
395 rate, or temperature, but a modest association with altered mental status (as defined by the
396 Glasgow Coma Scale Score <14, $p=0.00099$) and quick SOFA (qSOFA, $p=0.013$). We found the
397 predictive function of the Sepset signature used in the PREDICT device is independent of initial
398 clinical markers of severity (e.g., presentation lactate), and the association with parameters of
399 deterioration was sustained for up to day 3 (72 h post-presentation) in ICU-admitted patients
400 ($p<0.005$). These findings strongly suggest the Sepset classifier is not simply detecting a time
401 window of severity; instead, it identifies a specific mechanism-based signature. This signature
402 provides information about immunological cell reprogramming which impairs immune cells'
403 ability to respond to pathogens. This impairment is associated with future clinical deterioration,
404 independent from initial clinical markers of severity. In keeping with this hypothesis, we found
405 that the addition of other clinical or laboratory information, outside of the Sepset classifier, is not
406 required for its performance in predicting which patients are likely to go on deteriorate over the

407 next 24 h. This means that, other than suspicion of sepsis, no other information should be needed
408 for accurate prediction of the clinical course over the first 24 h following initial assessment. We
409 have prospectively performed validation of this finding in the COVID cohort that was studied
410 longitudinally,³⁸ which suggests that, the PREDICT device may be used in low resource or remote
411 settings, by non-expert personnel, to make immediate risk assessments about level of care.

412 Currently, there are no universally accepted diagnostic or prediction tools routinely used at the
413 bedside for patients with suspected sepsis; especially during the critical early stages of disease
414 where the risk of clinical deterioration may not be clear. Despite early adequate treatment, many
415 patients fail to improve; highlighting the heterogeneity of the syndrome and the need for
416 pathogen-agnostic risk stratification as well as novel therapeutics.^{63,64,65} The condition remains
417 misdiagnosed in approximately 30% of patients.⁶⁶ Poor specificity of current syndromic
418 diagnostic approaches results in unnecessary use of drugs (e.g., antibiotics and steroids) and
419 exposure to harm (iatrogenicity) associated with hospital and ICU-level care. They also fail to
420 predict, for those patients who are admitted to the ward or are sent home, whether and/or how
421 patients deteriorate and need higher levels of care,⁶⁷ placing a considerable financial and human
422 burden on both hospitals and patients, directly affecting prognosis. The primary purpose of the
423 SOFA score is to objectively describe organ (dys)function rather than to predict outcome, so no
424 associated equation developed for prediction is currently in use.⁶⁸ There is evidence from a range
425 of observational studies however, that even a small change in SOFA score is associated with a
426 persistent trend in mortality.⁶⁹ Recently, Seymour et al.⁷⁰ investigated the validity of the SOFA
427 score to predict in-hospital mortality in patients with suspected infections and found it to have
428 an AUC-ROC of 0.74. Therefore, predicting what the SOFA score will be 24 h post-presentation
429 may have important clinical value.

430
431 Unfortunately, there are important limitations to the calculation and utility of the SOFA score.
432 For example, it requires input from healthcare personnel, and samples need to be processed in
433 a clinical laboratory which may prove challenging when access to clinicians/nurses and/or the
434 laboratory is limited. While a qSOFA has been proposed as a potential solution in low resource
435 settings, there are some intrinsic problems with its utility in guiding early management and

436 qSOFA show poor accuracy (42-46% in 2 meta analyses^{71,72}). One of the advantages of PREDICT
437 is its potential to provide information about the risk of clinical deterioration independent of the
438 initial SOFA/qSOFA.

439
440 The advantages of Sepset and PREDICT include no pre-preparation of the sample, requirement
441 of very small volumes of blood and the accuracy afforded by the use of ddPCR. Importantly,
442 PREDICT is intended for prediction of risk of clinical deterioration in patients suspected of sepsis,
443 so this technology differs fundamentally from most other available tests. Other studies have used
444 retrospective electronic transcriptional data to predict 30-day mortality in adults.²⁸ The
445 advantage of using a small-sized gene classifier, and the ability to use highly accurate multiplex
446 PCR, needs to be balanced with the need for precision, although we have observed no diminution
447 of effectiveness of the 6 gene signature compared to the 99-gene cellular reprogramming
448 signature from which it was devised.³¹

449
450 An important limitation of our study is that the PREDICT and Sepset classifier are yet to be tested
451 prospectively to determine if its implementation would lead to improved management and
452 outcomes. Although we were able to show high overall accuracy in studies comprising 2,559
453 patients, there is a paucity of existing cohorts with high-quality (especially RNA-Seq)
454 transcriptomic data for validation, despite the high frequency of sepsis (49M cases per year).
455 Very few studies have collected early sepsis and/or longitudinal data with accompanying
456 biological samples linked to granular metadata. Since the pandemic, the importance of collecting
457 clinically annotated biological samples for translational research has significantly improved.
458 While showing that we can predict clinical deterioration in patients with bacterial and COVID-19
459 sepsis, the study should be expanded to other viral causes of sepsis and to important subsets of
460 bacterial sepsis (e.g., caused by particular bacteria). Importantly, we did not have information
461 regarding longer-term outcomes, such as 90-day and hospital mortality for the entire cohort. Our
462 data comes from observational studies, so it precludes us from determining whether the current
463 classifier is informative regarding response to therapy. In contrast, the classifier can be tested in
464 any other cohort even if similar clinical data have not been collected, as long as clear indicators

465 of clinical deterioration, such as a SOFA scores are available. Given that the PREDICT-based test
466 is conducted in the immediate near-patient environment, the expectation is that there will be no
467 need for storage, transport, or handling of tubes, simplifying its use at the bedside.

468

469 In summary, the current study describes the development of a molecular risk classifier for clinical
470 deterioration and onset of sepsis, and a novel POC device to measure this at the bedside, as well
471 as the proof-of-concept demonstration using real-world patient samples. The important feature
472 of the classifier and the technology is that no additional information, other than suspicion of
473 sepsis, is required to obtain a risk assessment that can be used in a clinically actionable fashion
474 at the bedside. Moreover, in the future, no expert personnel or equipment are required to
475 prepare the sample or interpret the results. Prospective testing of the device and the classifier
476 will be fundamental in moving forward to determine the clinical utility of the tool. However, the
477 technology can be adapted to measure virtually any nucleic acid-based marker, making it
478 modular, and therefore adaptable to easily measure both existing and emerging markers. This
479 adaptability could potentially enhance the predictive performance of the classifier.

480 **Acknowledgements:** We acknowledge our membership in the Canadian Critical Care Trials
481 (CCCTG), Canadian Critical Care Translational Biology Group (CCCTBG, CCDS) and Sepsis Canada
482 (RH and CCDS) and are grateful for their insightful review of the manuscript (Drs. A. Fox-
483 Robichaud, and Keith Walley). We are also indebted to Drs. Laurent Brochard and Arthur Slutsky
484 for the insightful comments. We gratefully acknowledge funding from the Collaborative R&D
485 Initiative Pandemic Response Challenge Program Grant Application from the National Research
486 Council of Canada to CCDS, RH, TV and LM (2020); the Canadian Institutes of Health Research
487 (CIHR) Collaborative Health Research Projects (CHRP, Natural Sciences and Engineering Research
488 Council of Canada partnered, FDN-433426 to CCDS, TV, LM, MG and DB); CIHR to CCDS (FDN-
489 420463); CIHR to AB and CCDS (Emerging COVID-19 Research Gaps & Priorities, FDN-466806);
490 CIHR COVID-19 Rapid Research Funding and CIHR FDN-154287 to RH. RH holds the University of
491 British Columbia Killam Professorship and previously held a Canada Research Chair. CCDS holds
492 the Robert and Dorothy Pitts Research Chair in Acute and Emergency Medicine and AB holds the
493 Cara Phelan Chair of Critical Care Medicine at the University of Toronto. The COVID cohort studies
494 were made possible through open sharing of data and samples from the Biobanque Québécoise
495 COVID-19, funded by the Fonds de recherche du Québec - Santé, Génome Québec and the Public
496 Health Agency of Canada, the University of Toronto COVID Biobank funded by the Canadian
497 Foundation for Innovation (CFI); and the CFI John R. Evans Leaders Fund (CFI-JELF, 2020) and St
498 Michael's Hospital Foundation (2020) for the creation of the PREDICT-biobank at Unity Health
499 Toronto.

500 **Conflicts of Interest**

501 REWH is an inventor of the Sepset signature that has been patented in 17 countries (e.g. US
502 patent 11,851,717 issued Dec 26, 2023) and is CEO and a shareholder of Asep Medical and its
503 subsidiary Sepset BioSciences Inc. that have licensed in these patents and are actively
504 commercially developing sepsis diagnostics. REWH also has a contract from Sepset Biosciences
505 for development of diagnostic assays for adult sepsis. PGYZ and EFH are employees of Sepset
506 Biosciences Inc. and/or Asep Medical.

507 **References**

- 508 1. Singer, M. *et al.* The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3).
509 *JAMA* **315**, 801 (2016).
- 510 2. Reinhart, K. *et al.* Recognizing Sepsis as a Global Health Priority - A WHO Resolution. *N Engl J Med*
511 **377**, 414–417 (2017).
- 512 3. Rudd, K. E. *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis
513 for the Global Burden of Disease Study. *The Lancet* **395**, 200–211 (2020).
- 514 4. Coronavirus (COVID-19). *Sepsis Alliance* <https://www.sepsis.org/sepsisand/coronavirus-covid-19/>.
- 515 5. Liu, R., Hunold, K. M., Caterino, J. M. & Zhang, P. Estimating treatment effects for time-to-treatment
516 antibiotic stewardship in sepsis. *Nat Mach Intell* **5**, 421–431 (2023).
- 517 6. Seymour, C. W. *et al.* Time to Treatment and Mortality during Mandated Emergency Care for Sepsis.
518 *N Engl J Med* **376**, 2235–2244 (2017).
- 519 7. Frost, R., Newsham, H., Parmar, S. & Gonzalez-Ruiz, A. Impact of delayed antimicrobial therapy in
520 septic ITU patients. *Crit Care* **14**, P20 (2010).
- 521 8. Lo, V. C. K. *et al.* Management of Patients With Sepsis in Canadian Community Emergency
522 Departments: A Retrospective Multicenter Observational Study. *Health Serv Res Manag Epidemiol* **7**,
523 (2020).
- 524 9. Rudd, K. E. *et al.* The global burden of sepsis: barriers and potential solutions. *Crit Care* **22**, 232
525 (2018).
- 526 10. Plopper, G. E., Sciarretta, K. L. & Buchman, T. G. Disparities in Sepsis Outcomes May Be Attributable
527 to Access to Care*. *Critical Care Medicine* **49**, 1358 (2021).
- 528 11. Sheikh, F., Douglas, W., Catenacci, V., Machon, C. & Fox-Robichaud, A. E. Social Determinants of
529 Health Associated With the Development of Sepsis in Adults: A Scoping Review. *Crit Care Explor* **4**,
530 e0731 (2022).

- 531 12. Dellinger, R. P. *et al.* Surviving sepsis campaign: international guidelines for management of severe
532 sepsis and septic shock: 2012. *Crit Care Med* **41**, 580–637 (2013).
- 533 13. Asner, S. A., Desgranges, F., Schrijver, I. T. & Calandra, T. Impact of the timeliness of antibiotic
534 therapy on the outcome of patients with sepsis and septic shock. *Journal of Infection* **82**, 125–134
535 (2021).
- 536 14. Im, Y. *et al.* Time-to-antibiotics and clinical outcomes in patients with sepsis and septic shock: a
537 prospective nationwide multicenter cohort study. *Critical Care* **26**, 19 (2022).
- 538 15. About WHO. <https://www.who.int/about>.
- 539 16. Janasek, D., Franzke, J. & Manz, A. Scaling and the design of miniaturized chemical-analysis systems.
540 *Nature* **442**, 374–380 (2006).
- 541 17. Wu, J., Dong, M., Rigatto, C., Liu, Y. & Lin, F. Lab-on-chip technology for chronic disease diagnosis.
542 *npj Digital Med* **1**, 1–11 (2018).
- 543 18. Haeberle, S. & Zengerle, R. Microfluidic platforms for lab-on-a-chip applications. *Lab Chip* **7**, 1094–
544 1110 (2007).
- 545 19. Mark, D., Haeberle, S., Roth, G., Stetten, F. von & Zengerle, R. Microfluidic lab-on-a-chip platforms:
546 requirements, characteristics and applications. *Chem. Soc. Rev.* **39**, 1153–1182 (2010).
- 547 20. Reddy, B. *et al.* Point-of-care sensors for the management of sepsis. *Nature Biomedical Engineering*
548 **2**, 640–648 (2018).
- 549 21. Gupta, E. *et al.* Fast Track Diagnostic Tools for Clinical Management of Sepsis: Paradigm Shift from
550 Conventional to Advanced Methods. *Diagnostics* **13**, 277 (2023).
- 551 22. Oeschger, T., McCloskey, D., Koppa, V., Singh, A. & Erickson, D. Point of Care Technologies for
552 Sepsis Diagnosis and Treatment. *Lab Chip* **19**, 728–737 (2019).

- 553 23. Brassard, D. *et al.* Extraction of nucleic acids from blood: unveiling the potential of active pneumatic
554 pumping in centrifugal microfluidics for integration and automation of sample preparation
555 processes. *Lab Chip* **19**, 1941–1952 (2019).
- 556 24. Geissler, M. *et al.* Centrifugal microfluidic lab-on-a-chip system with automated sample lysis, DNA
557 amplification and microarray hybridization for identification of enterohemorrhagic *Escherichia coli*
558 culture isolates. *Analyst* **145**, 6831–6845 (2020).
- 559 25. Malic, L. *et al.* Automated sample-to-answer centrifugal microfluidic system for rapid molecular
560 diagnostics of SARS-CoV-2. *Lab Chip* **22**, 3157–3171 (2022).
- 561 26. Scicluna, B. P. *et al.* Classification of patients with sepsis according to blood genomic endotype: a
562 prospective cohort study. *Lancet Respir Med* **5**, 816–826 (2017).
- 563 27. McHugh, L. *et al.* A Molecular Host Response Assay to Discriminate Between Sepsis and Infection-
564 Negative Systemic Inflammation in Critically Ill Patients: Discovery and Validation in Independent
565 Cohorts. *PLOS Medicine* **12**, e1001916 (2015).
- 566 28. Sweeney, T. E. *et al.* A community approach to mortality prediction in sepsis via gene expression
567 analysis. *Nat Commun* **9**, 694 (2018).
- 568 29. Davenport, E. E. *et al.* Genomic landscape of the individual host response and outcomes in sepsis: a
569 prospective cohort study. *Lancet Respir Med* **4**, 259–271 (2016).
- 570 30. Baghela, A. *et al.* Predicting sepsis severity at first clinical presentation: The role of endotypes and
571 mechanistic signatures. *EBioMedicine* **75**, 103776 (2022).
- 572 31. Pena, O. M. *et al.* An Endotoxin Tolerance Signature Predicts Sepsis and Organ Dysfunction at Initial
573 Clinical Presentation. *EBioMedicine* **1**, 64–71 (2014).
- 574 32. Pena, O. M., Pisticic, J., Raj, D., Fjell, C. D. & Hancock, R. E. W. Endotoxin Tolerance Represents a
575 Distinctive State of Alternative Polarization (M2) in Human Mononuclear Cells. *The Journal of*
576 *Immunology* **186**, 7243–7254 (2011).

- 577 33. Baghela, A. *et al.* Predicting severity in COVID-19 disease using sepsis blood gene expression
578 signatures. *Sci Rep* **13**, 1247 (2023).
- 579 34. Geissler, M. *et al.* Centrifugal microfluidic system for colorimetric sample-to-answer detection of
580 viral pathogens. *Lab Chip* **24**, 668–679 (2024).
- 581 35. An, A. Y. *et al.* Severe COVID-19 and non-COVID-19 severe sepsis converge transcriptionally after a
582 week in the intensive care unit, indicating common disease mechanisms. *Frontiers in Immunology*
583 **14**, (2023).
- 584 36. Hounkpe, B. W., Chenou, F., de Lima, F. & De Paula, E. V. HRT Atlas v1.0 database: redefining human
585 and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq
586 datasets. *Nucleic Acids Res* **49**, D947–D955 (2020).
- 587 37. <https://xgboost.ai/>.
- 588 38. An, A. Y. *et al.* Dynamic gene expression analysis reveals distinct severity phases of immune and
589 cellular dysregulation in COVID-19. 2023.11.04.565404 Preprint at
590 <https://doi.org/10.1101/2023.11.04.565404> (2023).
- 591 39. Burnham, K. L. *et al.* Shared and Distinct Aspects of the Sepsis Transcriptomic Response to Fecal
592 Peritonitis and Pneumonia. *Am J Respir Crit Care Med* **196**, 328–339 (2017).
- 593 40. Kalantar, K. L. *et al.* Integrated host-microbe plasma metagenomics for sepsis diagnosis in a
594 prospective cohort of critically ill adults. *Nat Microbiol* **7**, 1805–1816 (2022).
- 595 41. Cani, E. *et al.* Immunothrombosis Biomarkers for Distinguishing Coronavirus Disease 2019 Patients
596 From Noncoronavirus Disease Septic Patients With Pneumonia and for Predicting ICU Mortality. *Crit*
597 *Care Explor* **3**, e0588 (2021).
- 598 42. Wang, L. *et al.* An atlas connecting shared genetic architecture of human diseases and molecular
599 phenotypes provides insight into COVID-19 susceptibility. *medRxiv* 2020.12.20.20248572 (2020)
600 doi:10.1101/2020.12.20.20248572.

- 601 43. Tsalik, E. L. *et al.* Host gene expression classifiers diagnose acute respiratory illness etiology. *Sci*
602 *Transl Med* **8**, 322ra11 (2016).
- 603 44. Pankla, R. *et al.* Genomic transcriptional profiling identifies a candidate blood biomarker signature
604 for the diagnosis of septicemic melioidosis. *Genome Biol* **10**, R127 (2009).
- 605 45. Arunachalam, P. S. *et al.* Systems biological assessment of immunity to mild versus severe COVID-19
606 infection in humans. *Science* **369**, 1210–1220 (2020).
- 607 46. The Cancer Genome Atlas Program (TCGA) - NCI. [https://www.cancer.gov/ccg/research/genome-](https://www.cancer.gov/ccg/research/genome-sequencing/tcga)
608 [sequencing/tcga](https://www.cancer.gov/ccg/research/genome-sequencing/tcga) (2022).
- 609 47. GEO Accession viewer. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93101>.
- 610 48. McCaffrey, T. A. *et al.* RNA sequencing of blood in coronary artery disease: involvement of
611 regulatory T cell imbalance. *BMC Medical Genomics* **14**, 216 (2021).
- 612 49. Argmann, C. *et al.* Biopsy and blood-based molecular biomarker of inflammation in IBD. *Gut* **72**,
613 1271–1287 (2023).
- 614 50. Smith, C. L. *et al.* Identification of a human neonatal immune-metabolic network associated with
615 bacterial infection. *Nat Commun* **5**, 4649 (2014).
- 616 51. Dapat, C. *et al.* Gene signature of children with severe respiratory syncytial virus infection. *Pediatr*
617 *Res* **89**, 1664–1672 (2021).
- 618 52. Mammalian Gene Collection | National Institute of Biomedical Imaging and Bioengineering.
619 <https://www.nibib.nih.gov/content/mammalian-gene-collection>.
- 620 53. Clime, L., Brassard, D., Geissler, M. & Veres, T. Active pneumatic control of centrifugal microfluidic
621 flows for lab-on-a-chip applications. *Lab Chip* **15**, 2400–2411 (2015).
- 622 54. Brassard, D. *et al.* Extraction of nucleic acids from blood: unveiling the potential of active pneumatic
623 pumping in centrifugal microfluidics for integration and automation of sample preparation
624 processes. *Lab Chip* **19**, 1941–1952 (2019).

- 625 55. Jones, M. *et al.* Low copy target detection by Droplet Digital PCR through application of a novel
626 open access bioinformatic pipeline, 'definetherain'. *Journal of Virological Methods* **202**, 46–53
627 (2014).
- 628 56. Malic, L. *et al.* Epigenetic subtyping of white blood cells using a thermoplastic elastomer-based
629 microfluidic emulsification device for multiplexed, methylation-specific digital droplet PCR. *Analyst*
630 **144**, 6541–6553 (2019).
- 631 57. Balk, R. *et al.* Validation of SeptiCyte RAPID to Discriminate Sepsis from Non-Infectious Systemic
632 Inflammation. *J Clin Med* **13**, 1194 (2024).
- 633 58. Brakenridge, S. C. *et al.* Evaluation of a Multivalent Transcriptomic Metric for Diagnosing Surgical
634 Sepsis and Estimating Mortality Among Critically Ill Patients. *JAMA Netw Open* **5**, e2221520 (2022).
- 635 59. Brakenridge Scott, Uan-I Chen, Oliver Liesenfeld, Natalie N Whitfield, Timothy E Sweeney, and Lyle
636 Moldawer. The IMX-BVN-3 Classifier detects bacterial infections regardless of anatomical
637 localization in surgical ICU patients. in (Elsevier Inc., Nashville, TN, 2022).
- 638 60. Phua, J. *et al.* Characteristics and outcomes of culture-negative versus culture-positive severe sepsis.
639 *Crit Care* **17**, R202 (2013).
- 640 61. Tsalik, E. L. *et al.* Multiplex PCR To Diagnose Bloodstream Infections in Patients Admitted from the
641 Emergency Department with Sepsis. *J Clin Microbiol* **48**, 26–33 (2010).
- 642 62. Seymour, C. W. *et al.* Murine sepsis phenotypes and differential treatment effects in a randomized
643 trial of prompt antibiotics and fluids. *Critical Care* **23**, 384 (2019).
- 644 63. Shankar-Hari, M. *et al.* Reframing sepsis immunobiology for translation: towards informative
645 subtyping and targeted immunomodulatory therapies. *Lancet Respir Med* S2213-2600(23)00468-X
646 (2024) doi:10.1016/S2213-2600(23)00468-X.
- 647 64. van der Poll, T., Shankar-Hari, M. & Wiersinga, W. J. The immunology of sepsis. *Immunity* **54**, 2450–
648 2464 (2021).

- 649 65. Maslove, D. M. *et al.* Redefining critical illness. *Nat Med* **28**, 1141–1148 (2022).
- 650 66. Kumar, A. *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the
651 critical determinant of survival in human septic shock. *Crit Care Med* **34**, 1589–1596 (2006).
- 652 67. Fernando, S. M. *et al.* Emergency Department disposition decisions and associated mortality and
653 costs in ICU patients with suspected infection. *Crit Care* **22**, 172 (2018).
- 654 68. Moreno, R. *et al.* The Sequential Organ Failure Assessment (SOFA) Score: has the time come for an
655 update? *Critical Care* **27**, 15 (2023).
- 656 69. Lambden, S., Laterre, P. F., Levy, M. M. & Francois, B. The SOFA score-development, utility and
657 challenges of accurate assessment in clinical trials. *Crit Care* **23**, 374 (2019).
- 658 70. Seymour, C. W. *et al.* Assessment of Clinical Criteria for Sepsis: For the Third International
659 Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **315**, 762 (2016).
- 660 71. Jiang, J., Yang, J., Mei, J., Jin, Y. & Lu, Y. Head-to-head comparison of qSOFA and SIRS criteria in
661 predicting the mortality of infected patients in the emergency department: a meta-analysis. *Scand J*
662 *Trauma Resusc Emerg Med* **26**, 56 (2018).
- 663 72. Wang, C., Xu, R., Zeng, Y., Zhao, Y. & Hu, X. A comparison of qSOFA, SIRS and NEWS in predicting the
664 accuracy of mortality in patients with suspected sepsis: A meta-analysis. *PLoS One* **17**, e0266755
665 (2022).
- 666

667 **Supplemental Materials**

668 **Discovery dataset**

669 The whole transcriptome (RNA-seq) data from 586 whole blood samples from different
670 countries and continents comprised our patient cohort. 514 samples were collected and used
671 for discovery analyses (i.e., the discovery cohort). The remaining 72 samples were secondary
672 samples collected from 72 individuals, which were excluded from discovery analyses (to
673 prevent same-individual artifacts) and used as a validation cohort. The sepsis severity
674 associated with the discovery cohort (514) was based on the SOFA score of the patient at 24 h
675 after the first sample collection: 271 samples with $\text{SOFA} \geq 2$ were sepsis, and 243 samples with
676 $\text{SOFA} < 2$ were non-sepsis.

677 **Sepsis signature and housekeeping gene selection**

678 We tested 99 cellular reprogramming (CR) genes as potential sepsis markers³¹. We used DESeq2
679 to perform differential gene expression analyses and chose the genes that had the highest up-
680 regulation (positive fold-change) in high severity samples. We also estimated the predictive
681 accuracy of each CR gene by setting the sensitivity to 75%. We picked six genes (RETN, S100A8,
682 MCEMP1, S100A12, CYP1B1 and HK3) that had the best results in both analyses for the
683 SepsetER model.

684 We analyzed 2,833 housekeeping genes³⁶ in our discovery cohort of 514 samples to set a
685 baseline for RNA quantity and sequencing depth. We selected housekeeping genes (HKG) with
686 high and consistent expression across all samples based on mean and variance. We then
687 examined the expression variance of the top 20 HKG candidates across key clinical factors such
688 as age group, gender, sepsis severity, patient location, mortality, etc. The two housekeeping
689 genes (PTP4A2 and CHTOP) with the lowest variances were used to set a baseline for the
690 SepsetER model.

691 **ML algorithm construction and testing**

692 Our own RNA-Seq data from 873 patient samples³⁰, was used for feature (gene) reduction using
693 ML. An additional 1241 transcriptomes from patients were used for testing the derived

694 signature. Three major groups of datasets were used for biomarker development – training,
695 validation and testing. The discovery data set (N=586) was first tested by 10X cross validation
696 and randomly divided into a training (90% of the samples) used for the construction of the
697 models (10,000+ models) and a test dataset (10% of the samples) to assess the best model
698 (**Supplemental Figure S2**). We trained 18 different machine learning algorithms on the
699 transcriptomic profiles of the 514 discovery cohort samples. The algorithms were: K-Nearest
700 Neighbors (KNN), Ridge Regression (RR), Lasso regression (LR), Elastic Net (EN), Partial Least
701 Square (PLS), Linear Discriminant Analysis (LDA), Regularized Discriminant Analysis (RDA),
702 Quadratic Discriminant Analysis (QDA), Bayesian Generalized Linear Model (BL), Naïve Bayes
703 (NB), Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), Adaptive
704 Boosting (AB), Stochastic Gradient Boosting Model (GBM), Extreme Gradient Boosting (XGB),
705 Neural Network (NN), and Multilayer Perceptron (MLP). We tested each model with different
706 parameters and chose the best one based on the AUC-ROC using 10-fold cross validation,
707 repeated 10 times. We then validated the performance of each model with the additional 72
708 validation samples (that were not in the training dataset).

709 We tested the SepsetER model with multiple methods. We used various sepsis transcriptome
710 datasets (from microarray and RNA-seq platforms) with over 3,000 sepsis and healthy samples
711 to evaluate the SepsetER sepsis classification model. We also trained other published sepsis
712 gene-signatures with our training dataset and compared them with SepsetER. The SepsetER
713 model, using the Extreme Gradient Boosting (XGB) algorithm, performed better than all other
714 signatures, with a median AUC-ROC of 0.85 in all testing datasets.

715 **Design of primers and probes**

716 The expression of 6 top genes of interest were assessed based on the selection of the highest
717 fold changes with respect to severity of disease in the ICU cohort. These genes are HK3
718 (108bp), RETN (78bp), S100A12 (122bp), S100A8 (122bp), MCEMP1 (131bp), CYP1B1 (114bp).
719 The housekeeping genes were also selected based on stable expression: PTP4A2 (138bp) and
720 CHTOP (113bp).

721 Primers (IDT) for these genes have been designed to span the exon-exon junction to avoid
722 amplification of genomic DNA and to cover the different isoforms. The amplicons' sizes range
723 between 78bp and 138bp. The probes were synthesized (IDT) with either FAM, HEX, ROX or Cy5
724 fluorescent labels and a ZEN/3' Iowa Black FQ (IABkFQ) double quencher when possible to
725 reduce background noise.

726 Two fourplex reactions were designed to include 3 genes of interest and 1 HKG for
727 normalization. As such one reaction targeted: CYP1B1, MCEMP1, S100A12 and PTP4A2, and the
728 other: HK3, RETN, S100A8 and CHTOP. The sequences of the primers and probes are provided
729 and described in the supplementary S.

730 Specificity and sensitivity of the primers and probes were first assessed by performing qPCR
731 standard curves of the individual targeted genes from human universal cDNA (P/N 637223,
732 Clontech/TaKaRa Bio) and comparing the efficiency with the multiplex reaction. The results
733 obtained were then used to design the multiplex ddPCR reactions in order to ensure
734 appropriate amplification of differentially expressed genes and avoid amplification bias.

735 **Sample collection, RNA isolation and cDNA conversion**

736 Patient samples were collected in Pax Gene tubes and total RNA was isolated using standard
737 protocol for Qiagen RNeasy mini kit. RNA was assessed first using NanoDrop One
738 spectrophotometer (Thermo Scientific) and A260/A280 values were between 1.8 and 2.2, with
739 typical yields in the range of 6–8 µg total.

740 RNA Integrity Number (RIN) was determined using the Agilent 2100 Bioanalyzer (Agilent
741 Technologies). Following the standard Nanochip protocol, samples with RIN values >7.0 were
742 used for conversion to cDNA. Input volumes for reverse transcription were calculated using the
743 concentration from the bioanalyzer (approximately 500 ng total was used per sample) and a
744 High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) was used following standard
745 protocols.

746 **Commercial duplex ddPCR**

747 For optimal results, recommendations made in the Droplet Digital PCR Applications Guide
748 (bulletin 6407) were followed. We used equal concentrations of cDNA for droplet generation
749 following the protocol for ddPCR supermix for probes (bio-rad.com/web/ddPCRsmxProbes).
750 Briefly, a 20 μ L reaction set-up consists of 2X supermix, 20X probes (a duplex of FAM and HEX),
751 equal concentrations of the patient samples (500 pg), and RNase-free water. The bulk solution
752 (in a 96-well plate) is applied to the AutoDG (automated droplet generator) where the solution
753 is partitioned into 10,000 individual water-in-oil droplets. The 96 well plate is foil sealed and put
754 into the C1000 thermal cycler (BioRad) where the individual droplets are subjected to the
755 following conditions: 10 min at 95 °C, 40 cycles of 30 s at 94 °C and 1 min at 60 °C, followed by
756 10 min at 98 °C and a 4 °C hold. Subsequently, the droplets were read in the QX200 Droplet
757 Reader using FAM and HEX channel readout in the QuantaSoft software. After data acquisition,
758 the QC of the samples was assessed (ensuring equal droplet numbers generated) and samples
759 were selected in the well selector tool under the Analyze tab. Samples were all manually
760 thresholded using the values from probe alone readout and confirmed in 2D tracings of the
761 duplexed reaction. Samples were tested in duplicate. The concentration reported is “copies/ng
762 DNA” of the final 1X ddPCR reaction.

763 **Microfluidic assay implementation**

764 **Total RNA extraction from whole blood**

765 Total RNA was extracted from 50 μ L of whole blood collected in PAXgene tubes using custom
766 Galenvs Total RNA kit (Galenvs) following manufacturer’s recommendations. Briefly, whole
767 blood aliquot is mixed with 50 μ L PBS and introduced onto the cartridge for automated
768 protocol or processed manually for extraction in tubes. The mixture was first combined with 20
769 μ L Proteinase K, and mixed. Lysis/binding buffer was then added to the solution and incubated
770 at 55 °C for 10 min. For manual extraction in tubes, a DynaMag magnetic rack (Thermo Fisher
771 Scientific) was used to capture magnetic nanoparticles. Following the capture of the RNA bound
772 to the beads, two consecutive wash steps are performed. Elution was performed in 25 μ L of
773 nuclease-free water (Sigma-Aldrich). On-chip extraction of total RNA was performed using the
774 automated protocol implemented on the centrifugal platform with the same reagents and

775 volumes as for the manual extraction. For the on-chip capture of MNPs, the external magnetic
776 field was provided by a nickel-plated neodymium alloy disk magnet (D201, 1/8" in diameter,
777 1/32" in thickness; K&J Magnetics) which remained inserted in the designated area on the
778 cartridge for the entire duration of the automated assay. The extracted RNA was subsequently
779 used in downstream RT-qPCR for assessment of RNA extraction efficiency as well as in on-chip
780 ddPCR for determination of transcript copy number.

781 **qPCR**

782 cDNA obtained from different patients were analyzed in a multiplexed qPCR using primer-probe
783 sequences for genes of interest and housekeeping genes as internal controls for normalization.
784 Each qPCR reaction consisted of 5 μ L 10X PCR Buffer (Qiagen), 8 μ L HotStar Taq Plus DNA
785 Polymerase, 3 μ L 25 mM MgCl₂, 1 μ L dNTPs, 5 μ L 10X primer-probe mix (final concentration of
786 1 μ M and 0.5 μ M, respectively), 2 μ L template, and 26 μ L nuclease-free water (Sigma-Aldrich),
787 for a total volume of 50 μ L. Samples were tested in duplicate. A no-template control (NTC)
788 reaction was included to assess for contamination. Thermal cycling was performed according to
789 the manufacturer's recommended protocol in a Bio-Rad CFX96 Touch Real-Time PCR Detection
790 System (Bio-Rad). To quantify the copies of genes of interest each qPCR run included serial
791 dilutions of cDNA (Takara) generating as such a standard curve. Cq values were plotted against
792 the log concentration and linear regression was used to determine standard curves. The
793 efficiency of each assay was 100 \pm 10% and the R^2 of each standard curve was >0.98.

794 **RT-ddPCR**

795 The ddPCR reaction master mix comprised of 5 μ L 10X PCR Buffer (Qiagen), 8 μ L HotStar Taq
796 Plus DNA Polymerase, 8 μ L 100X QuantiTect Virus RT Mix (Qiagen), 3 μ L 25 mM MgCl₂, 1 μ L
797 dNTPs, 5 μ L 10X primer-probe mix (final concentration of 1 μ M and 0.5 μ M respectively), 2 μ L
798 template, and 18 μ L nuclease-free water (Sigma-Aldrich), for a total volume of 50 μ L. Template
799 input was 2 μ L of cDNA, RNA, or nuclease-free water for NTC samples. On-chip ddPCR assay was
800 performed using an automated protocol implemented on the centrifugal platform. Briefly,
801 droplets containing template input in ddPCR reaction master mix were generated on-chip in
802 fluorinated carrier oil (5% 00-8 FluoroSurfactant in HFE7500) (RAN Biotechnologies). The

803 resultant emulsion was then transferred to the platform heater and cycled following
804 manufacturer's recommended protocol (20 min at 50 °C, followed by 5 min at 95 °C and 40
805 cycles of 15 s at 95 °C and 45 s at 60 °C, with ramp rate of 1 °C/s). Following thermal cycling, the
806 emulsion was transferred to the chip for fluorescence imaging and data analysis. All
807 experiments were performed in duplicate (no significant differences).

Supplemental Table S1: Data sets used for derivation, validation and testing of classifier						
Dataset group	Size	Patient location	Platform	Positive class (N)	Negative class (N)	Reference
Training and Cross Validation	586	Australia, Colombia, Netherlands, USA, Canada	RNA-Seq	24hr SOFA ³ 2 (N=271)	24hr SOFA <2 (N=243)	Baghela 2022; GSE185263 + in-house (this study)
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	Netherlands, UK	Micro array	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 used 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

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
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
^a Sourced from https://www.cancer.gov/ccg/research/genome-sequencing/tcga						

808

809
810
811
812
813
814
815
816
817
818
819
820
821
822

HGNC symbol	Fold Change SOFA ≥ 2	Specificity	P-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 S2: 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	Gene Name	Functional Prediction	Source
823	RETN	Resistin	Antibacterial activity against both Gram positive and Gram negative bacteria	https://www.genecards.org/cgi-bin/carddisp.pl?gene=RETN&keywords=MCEMP1
824	MCEMP1	Mast Cell Expressed Membrane Protein 1	Involved in regulating innate immune responses	https://www.genecards.org/cgi-bin/carddisp.pl?gene=MCEMP1&keywords=MCEMP1
825	HK3	Hexokinase 3	phosphorylate glucose to produce glucose-6-phosphate, the first step in most glucose metabolism pathways	https://www.genecards.org/cgi-bin/carddisp.pl?gene=HK3&keywords=HK3
826	S100A8	S100 Calcium Binding Protein A8	Intracellular activity: facilitates leukocyte arachidonic acid trafficking and metabolism, modulation of tubulin-dependent cytoskeleton during migration of phagocytes and activation of the neutrophilic NADPH-oxidase. Extracellular: acts as an alarmin, antimicrobial, oxidant-scavenging and apoptosis-inducing activities	https://www.genecards.org/Search/Keyword?queryString=%20S100A8
827	CYP1B1	Cytochrome P450 Family 1 Subfamily B Member 1	Monoxygenase which catalyze reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.	https://www.genecards.org/cgi-bin/carddisp.pl?gene=CYP1B1&keywords=CYP1B1
828	S100A12	S100 Calcium Binding Protein A12	Acts as an alarmin, activity involves recruitment of leukocytes, promotion of cytokine and chemokine production, and regulation of leukocyte adhesion and migration.	https://www.genecards.org/cgi-bin/carddisp.pl?gene=S100A12&keywords=S100A12
	All 6 genes together (RETN, MCEMP1, HK3, S100A8, CYP1B1, S100A12)	Sepset genes	REACTOME Pathway: Neutrophil Degranulation R-HSA-6798695 (adjusted p= 0.0002310, odds ratio= 84.18) GO Molecular Function: RAGE Receptor Binding GO:0050786 (adjusted p=0.00003777, odds ratio=1665.67)	https://maayanlab.cloud/Enrichr/enrich

Supplemental Table S3: 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, adjusted p-value and odds ratio describe likelihood that functional enrichment prediction using REACTOME or GO Molecular function was identified by chance alone.

Supplemental Table 4: Baseline demographic, clinical characteristics and adjusted p-values for the association with the classifier at 24 and 72 hrs post-initial clinical assessment (N=248)			
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
------------------	----	----	----

829

830 **Supplemental Table S4: Baseline demographic and clinical characteristics and correlation with clinical traits,**
831 **interventions and outcomes at recorded at 24 and 72 h.** Data were collected from consenting adult patients (> 18
832 years of age) with ethics approval, who presented with prospective sepsis, within the first two hours of emergency
833 room (ER) or within 24 h of intensive care unit (ICU) admission at baseline, 24 h and 72 h. Abbreviations: ICU
834 (intensive care unit), SOFA (sequential organ failure score), WBC (white blood cell count), P/F ratio (ratio of the
835 partial pressure of oxygen over fraction of inspired oxygen), IQR (interquartile range). A significant correlation
836 between features and Sepset is denoted by $p \leq 0.05$ using χ^2 /Fisher exact test for categorical variables and analysis
837 of variance for continuous variables as appropriate with Benjamini and Hochberg correction.

838

839

Supplemental Table S5. Primers and Probes used for ddPCR			
Gene	Primer/Probe	Sequence (5' → 3')	Source
CYP1B1	Forward	GGC TGG ATT TGG AGA ACG TA	IDT
	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	IDT
	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	IDT
	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	IDT
	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	IDT
	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	IDT
	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	IDT
	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	IDT
	Reverse	CTG CCA CGC CCA TCT TTA TC	
	Probe	/5Cy5/GTG CAG ACG /TAO/TCT GGT TCA AAG AGT TGG A/3IAbRQSp/	

840

841

Supplemental Table S6. 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

842

843

Supplemental Table S7. 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)

844

Table S8 Microfluidic implementation of the automated RNA extraction protocol

Step	Operation	Active ports (#)	Applied pressure (psi) ^a	Rotation speed (rpm)	Duration
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 \times)
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 \times)
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 \times)
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.

845

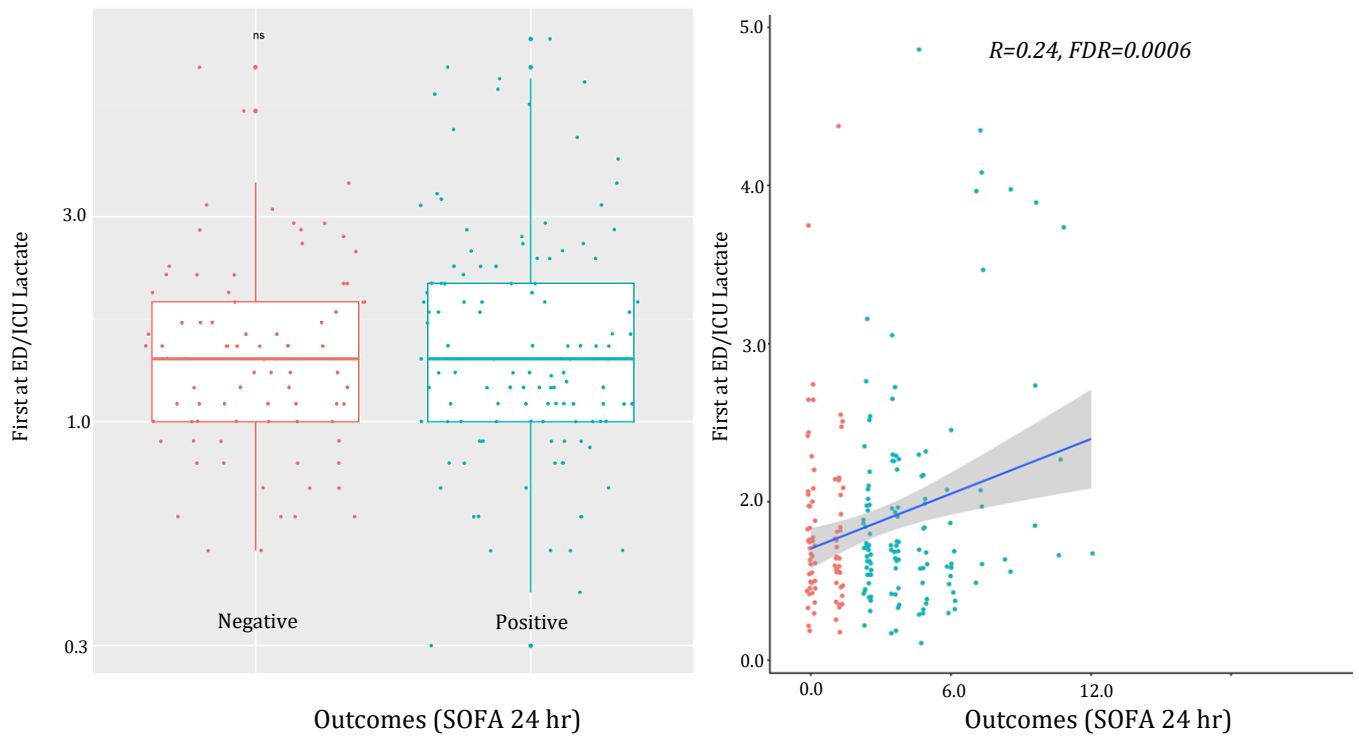
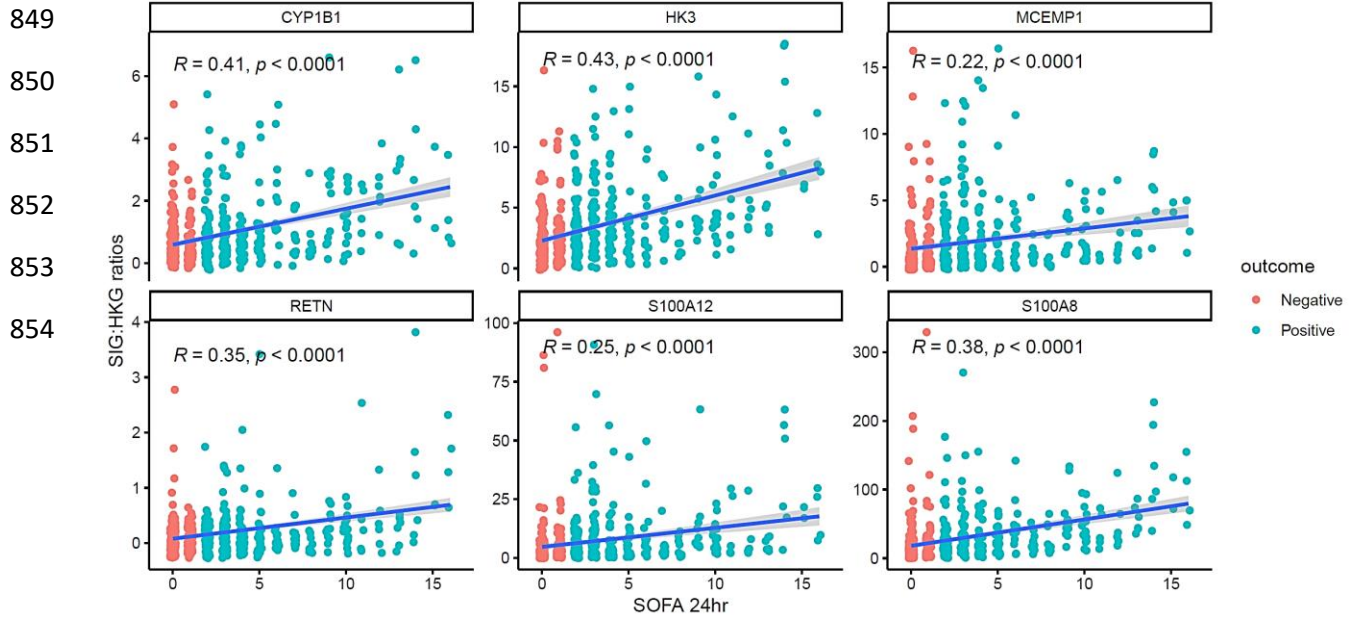
846

Supplemental Table S9. Microfluidic implementation of the automated ddPCR protocol					
Step	Operation	Active ports (#)	Applied pressure (psi)^a	Rotation speed (rpm)	Duration
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.

847

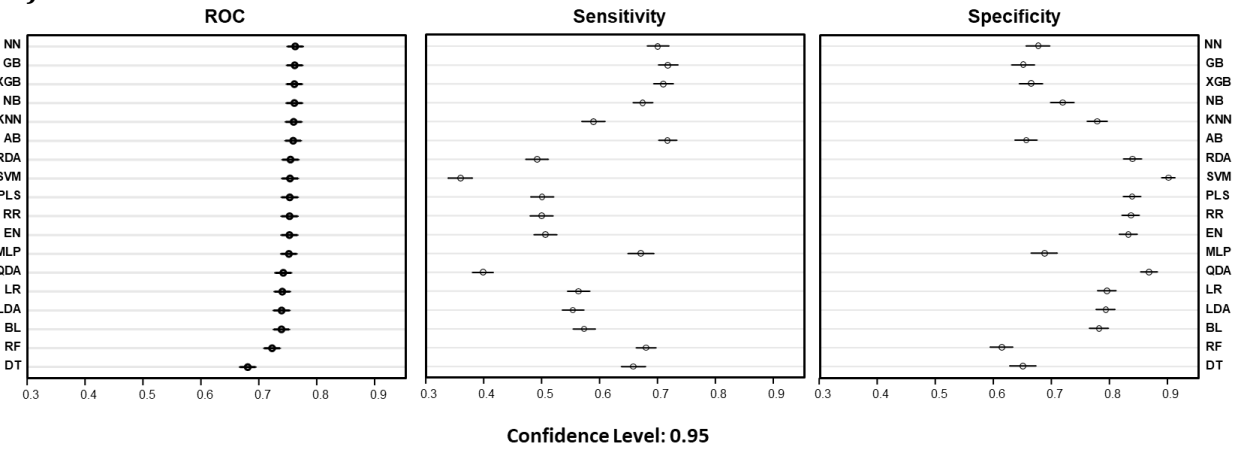
848



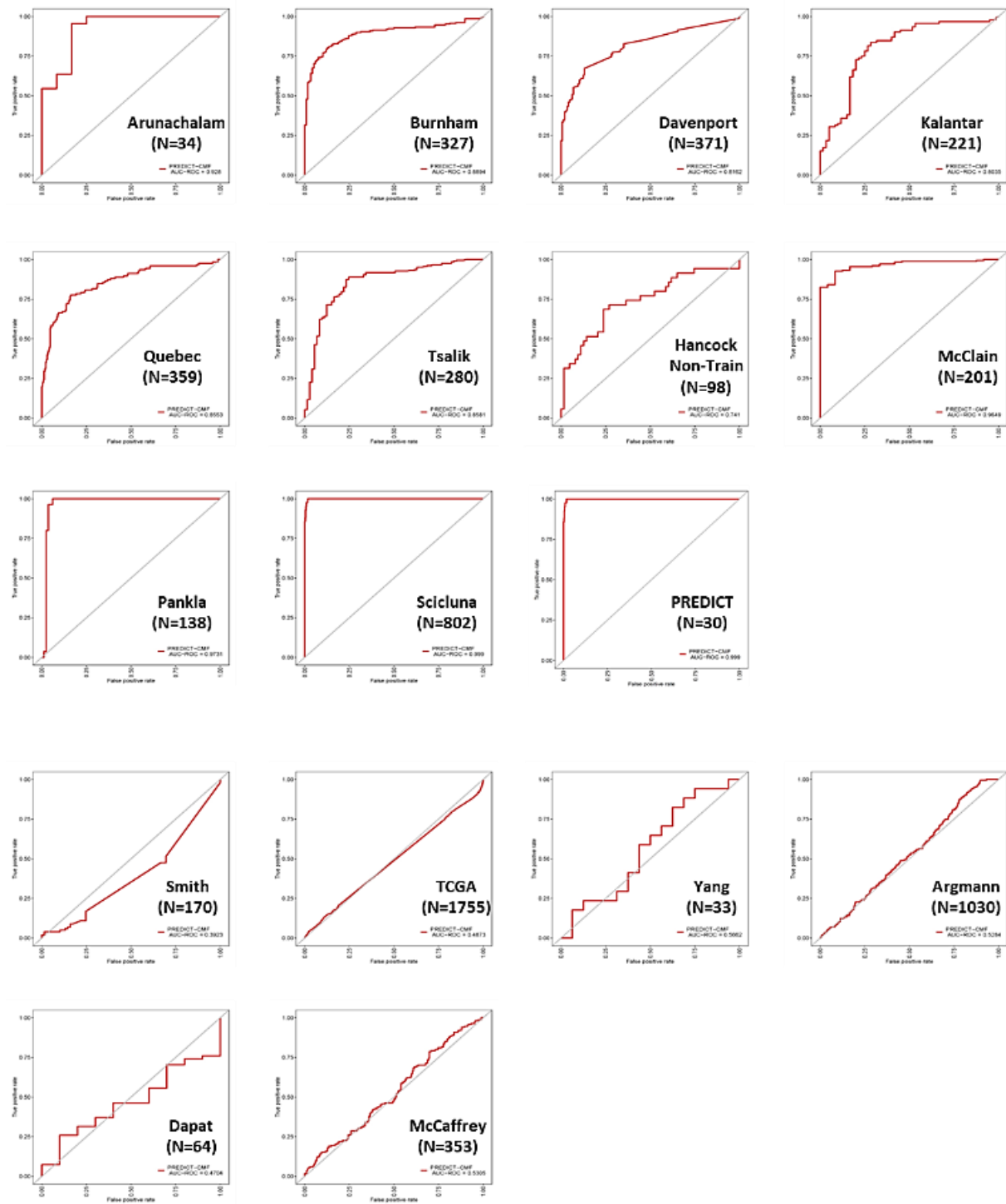
855 **Supplemental Figure S1: A)** Relationship between 24-hour SOFA score and the relative gene expression of each of the Sepset-signature genes demonstrating a relative increase in expression as a function of 24-hour SOFA score. Red = Sepsis negative SOFA <2; Blue = sepsis positive SOFA \geq 2. **B)** Relationship between baseline SOFA and serum lactate with clinical deterioration. Boxplots shows no difference between lactate level between sepsis samples (positive = SOFA \geq 2) vs. non-sepsis samples (negative = SOFA < 2). The scatterplot shows poor correlation between lactate level vs. SOFA score (R = 0.24, suggesting lactate – as a marker of clinical severity - can only account for ~5% variation in SOFA scores).

856
857
858
859
860
861
862

A)



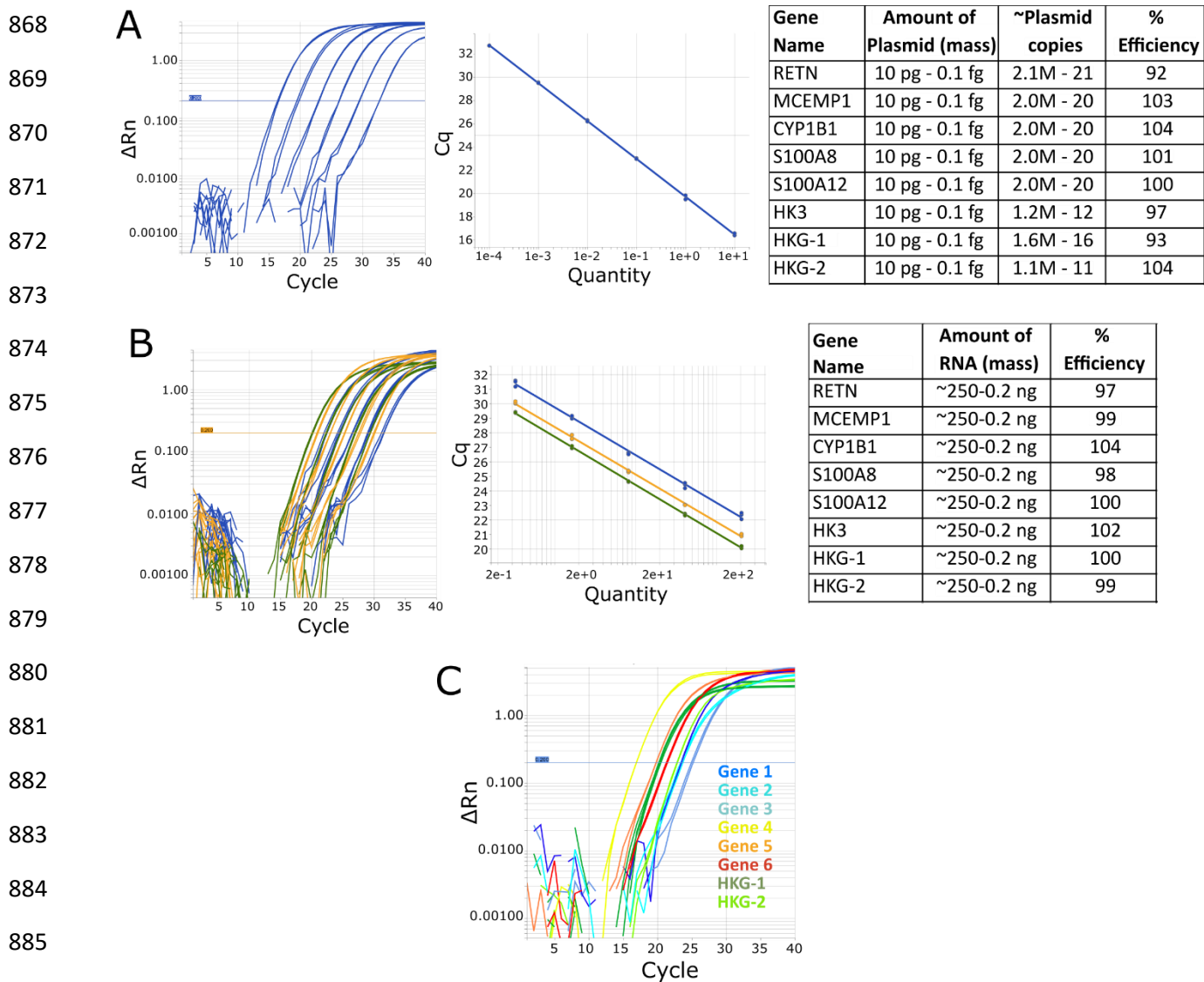
863 **B)**



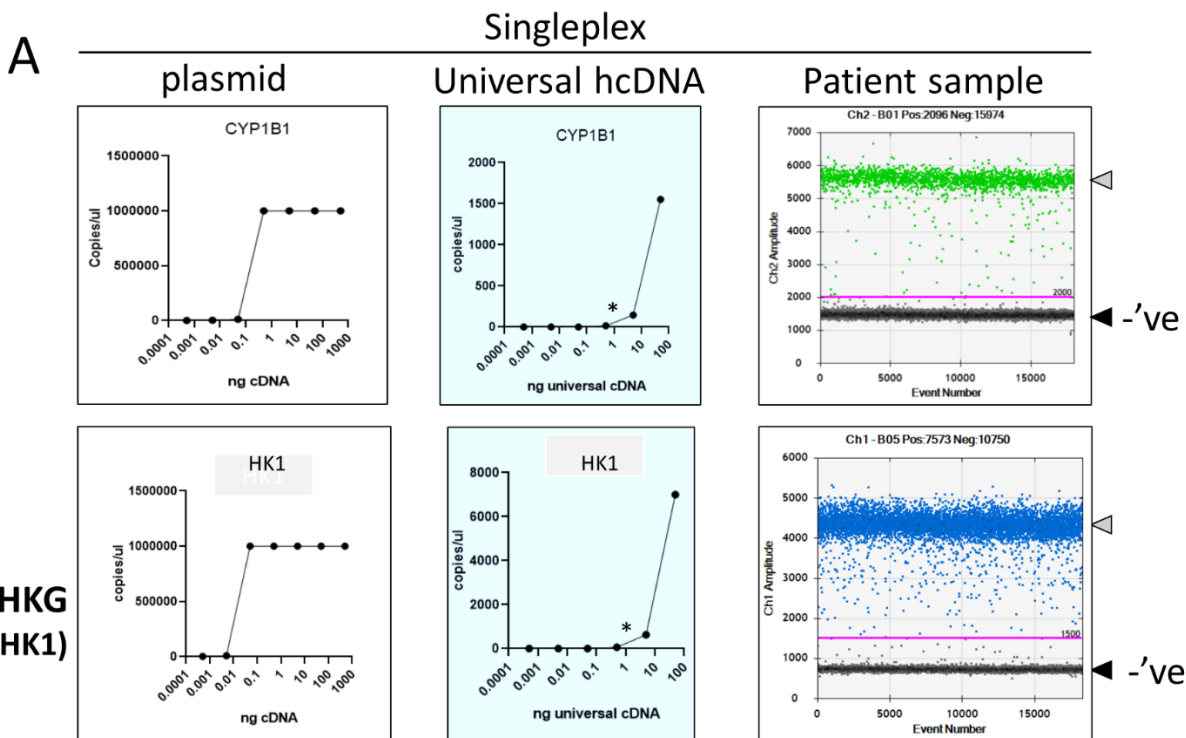
864

865

866 **Supplemental Figure S2 Selection of the Sepset (and 2 HKG) gene expression signature.** Schematic of the patient
867 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).



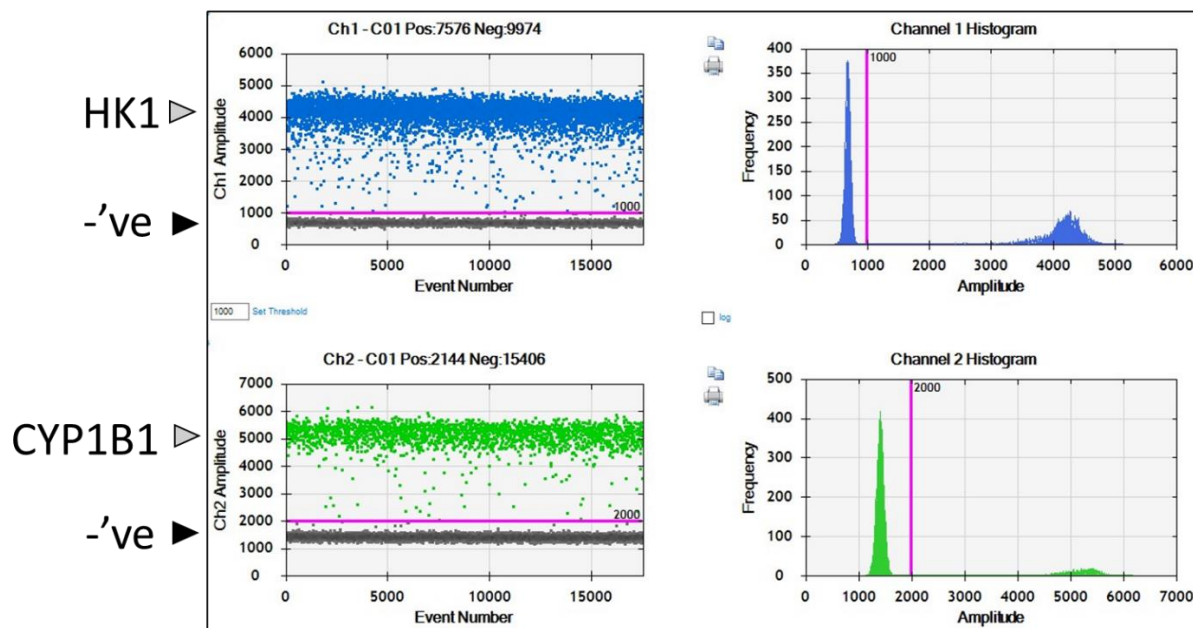
Supplemental Figure S3 – Development and performance summary of RT-qPCR primers and probes used in the Sepset signature. PCR primers were developed for each sepsis gene signature gene as well as two HKGs and their reaction efficiency was evaluated in a single-plex reaction using gene specific plasmids as the reaction template (A). Shown is the representative amplification curves and corresponding standard curve for the MCCEMP1 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).



896

897

B Multiplex on patient sample



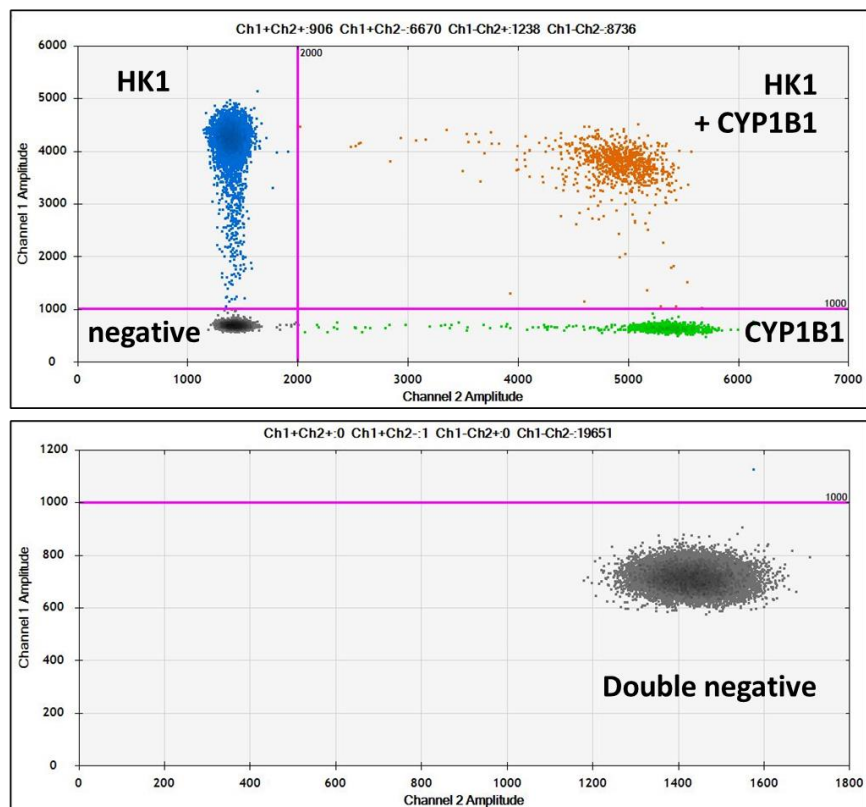
898

899

900

C

HK1 +
CYP1B1



Probe alone

901

902

903

904

905

906

907

908

909

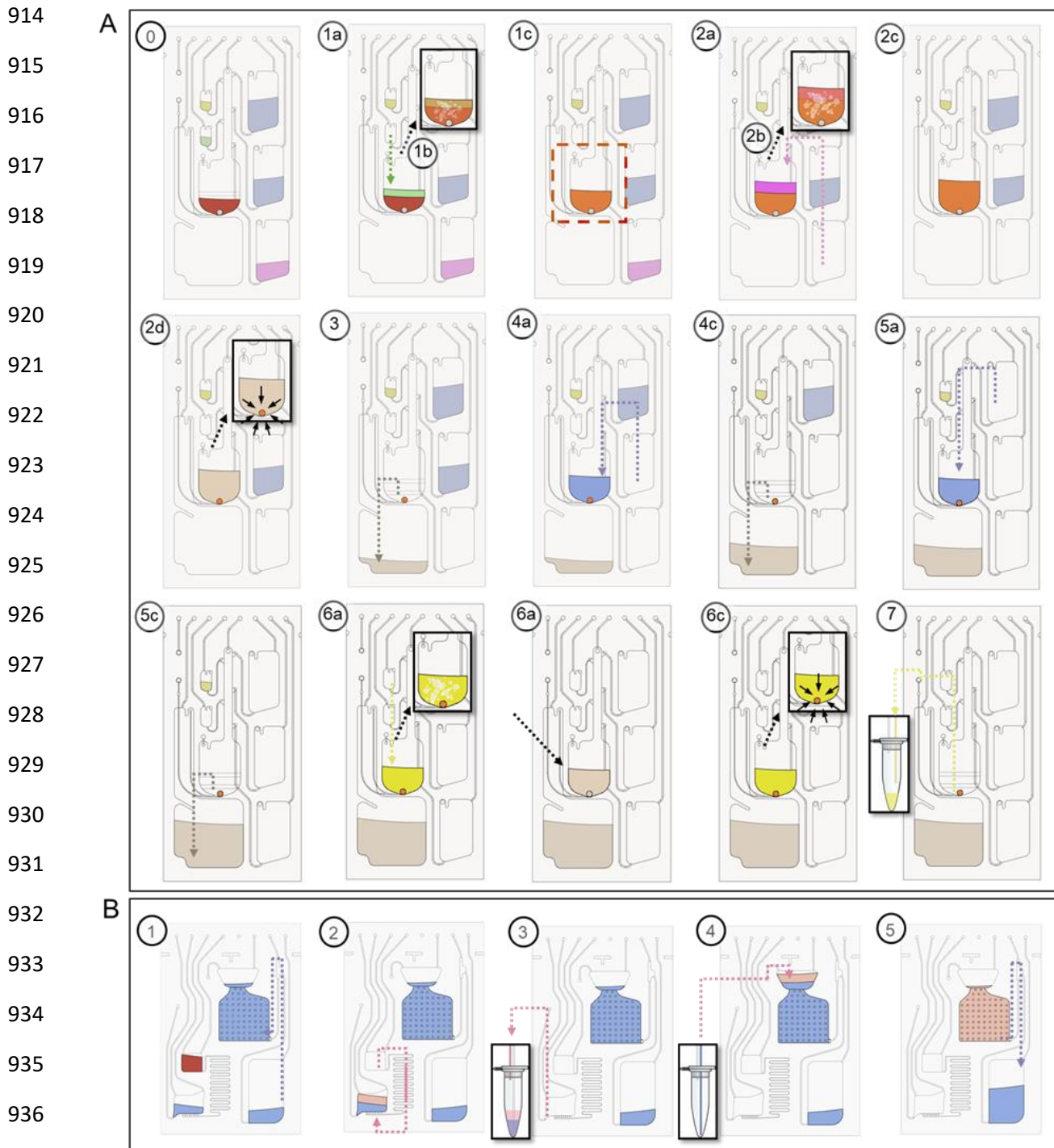
910

911

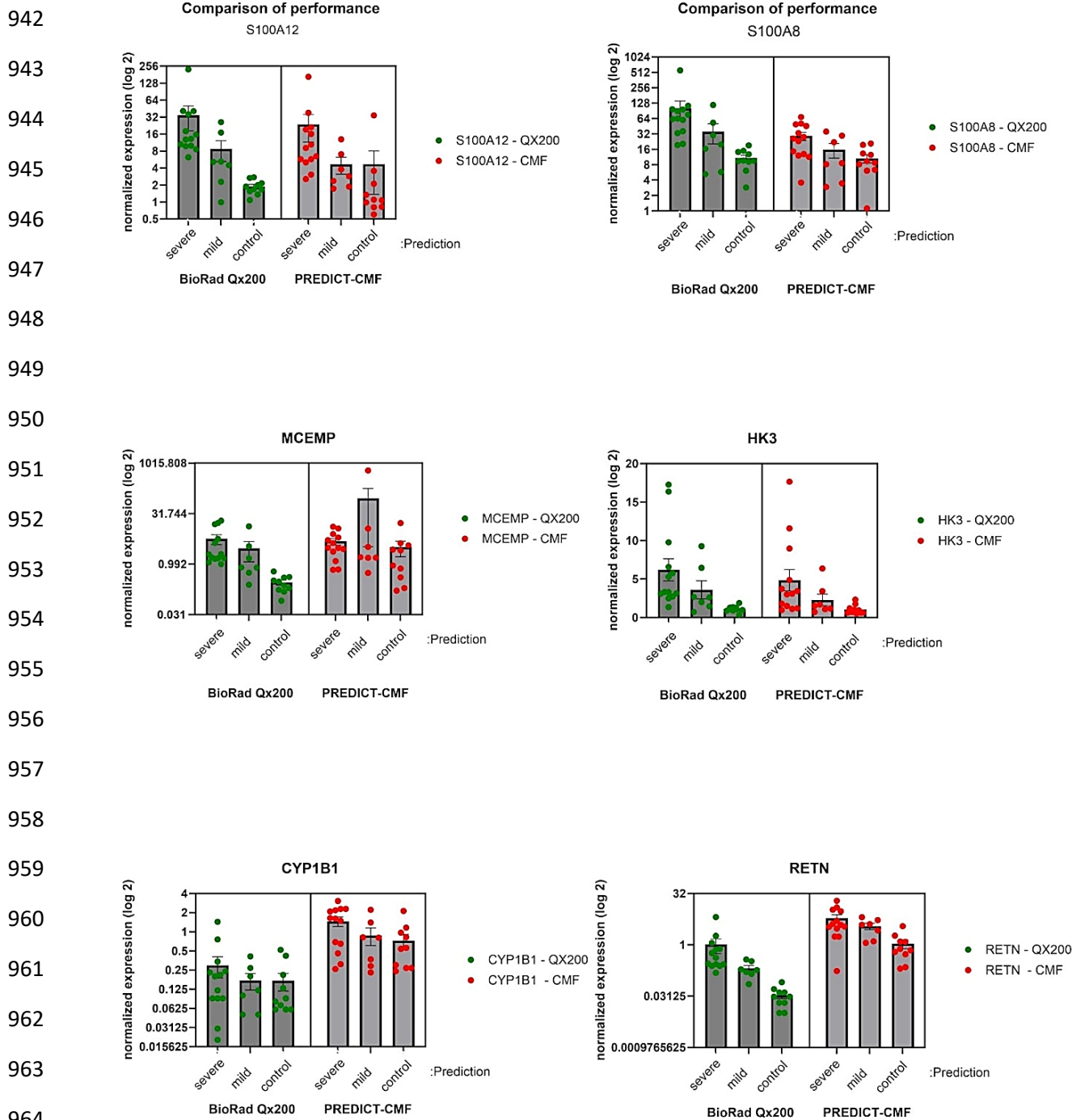
912

913

Supplemental Figure S4. 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 PTP4A2 as representative housekeeping gene (HKG, 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 PTP4A2) 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 (B. top right panel) and clear demarcation of double negative droplets (bottom left quadrant), PTP4A2 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 (B. bottom right panel).



Supplemental Figure S5. Microfluidic implementation. (A) RNA extraction and (B) ddPCR assays. The platform executes a predetermined sequence of spinning and pneumatic actuation steps to perform the entire analytical protocol in an automated fashion on the cartridge. These steps are represented by the sequence of schematic images showing the status of the cartridge at different steps during the automated RNA extraction protocol and subsequent ddPCR (see text and Table S2 and Table S3 for details). The arrows point in the direction of fluid flow for each step. The activated ports at each step are highlighted by green and red circles, representing the application of positive and negative pressures at that port, respectively.



Supplemental Figure S6: 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-CMF (red)) quantitated (copies/ μ L) the expression of the (6) signature genes of interest (GOI) and (2) housekeeper genes (PTP4A2 and CHTOP). 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 log₂ scale. ICU patient samples were classified as “severe or mild” based on their SOFA scores (>2, severe and <2, mild) with the healthy controls.