Supplemental Digital Content

# HemoSpec study: Leukocyte activation profile assessed by Raman spectroscopy helps diagnosing infection and sepsis

**Anuradha Ramoji PhD[1#], Daniel Thomas-Rüddel MD[1,2,#] Oleg Rybachkov PhD[3,4 #], Michael Bauer MD[1,2], Natalie Arend MSc[1,3], Evangelos J. Giamarellos-Bourboulis MD[5], Jesper Eugen-Olsen PhD[6], Michael Kiehntopf MD[1,7], Thomas Bocklitz PhD [3,4], Jürgen Popp PhD[1,3,4], Frank Bloos MD, PhD[1,2*], Ute Neugebauer PhD, med. habil.[1 *]**

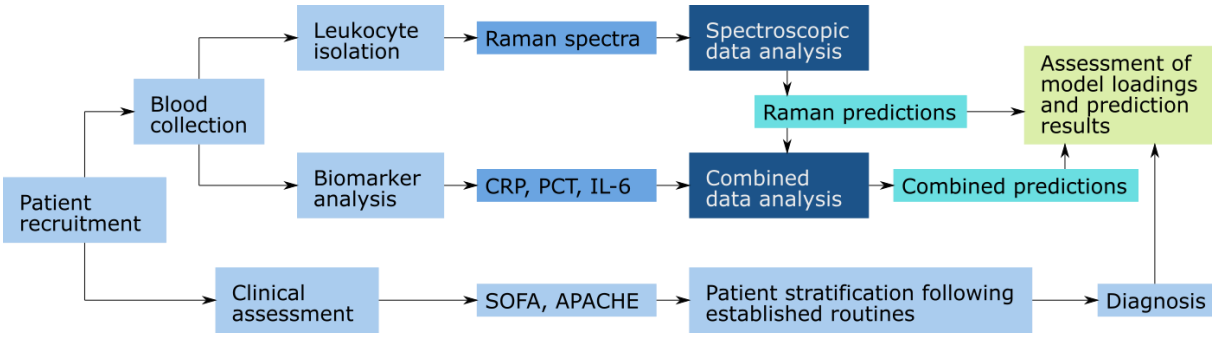
*# equal first author contribution*

**shared corresponding authors*

**Content:**

*Analysis done within the HemoSpec study*

Patients were enrolled (between July 2014 and December 2017) at the Jena University Hospital. The study was conducted according to the recommendations concerning human research that are contained in the Declaration of Helsinki. The general study flow from patient recruitment to evaluation is summarized in Supplemental Figure S1. Figure 4, main manuscript, shows the patient recruitment during the HemoSpec study.




***Supplemental Figure S1:*** **Study flow chart showing the main steps from patient recruitment to evaluation.**

As only a small amount of blood (< 500 µl) is needed to perform the outlined Raman analysis and the result can be available in short times (< 3 hours), Raman screening of peripheral leukocytes could be used as a first screening tool for the treating physician. The physician will gain insight if the patient is suffering from an infection or sterile inflammation. Among the patients with suspected infection, further analysis can be carried out in agreement with the workflow suggested by Prescott and Iwashyna (1). The potential to differentiate bacterial and fungal (and maybe viral) infections based on the host response has to be proven in further studies. With this information, the treating physician can select the appropriate further in-depth diagnostic tools, e.g. PCR-based assays or also new phage-based approaches (2) to further analyze and identify the specific cause of infection.

*Biomarker analysis in blood samples*

Blood samples were collected using standard protocols using ethylenediaminetetraacetic acid (EDTA) containing tubes (3x 2.7 ml, Sarstedt, Nuembrecht, Germany) and lithium-heparin tubes (1x 2.7 ml and 1x 4.9 ml, Sarstedt, Nuembrecht, Germany).

Concentrations of the three most frequently studied biomarkers (3) procalcitonin (PCT), C-reactive protein (CRP) and interleukin 6 (IL-6) were determined in blood samples of patients following established protocols as provided by the respective manufacture's kit. The concentration of PCT was determined in the serum samples (~50 µl). The BRAHMS PCT sensitive KRYPTOR (ThermoScientific, Germany) kit, a homogeneous sandwich immunoassay was used. The measuring principle is based on Time-Resolved Amplified Cryptate Emission (TRACE) technology. For CRP biomarker a latex immunoassay based on immunoturbidimetric was used (MULTIGENT CRP Vario, Germany). The CRP concentration was determined in EDTA plasma (~500 µl). The IL-6 concentration was determined in EDTA plasma samples (~30 µl) using the electrochemical luminescence Immunoassay "ECLIA" carried out on Elecsys and cobas e immunoassay systems (Diagnostics Roche, Germany). The suPAR concentration was measured in serum using an enzyme immunosorbent assay provided by ViroGates, Denmark.

For statistical data analysis, the concentrations of biomarkers (CRP, PCT and IL-6) were scaled to their standard deviation.

*Leukocytes isolation for Raman spectroscopic analysis*

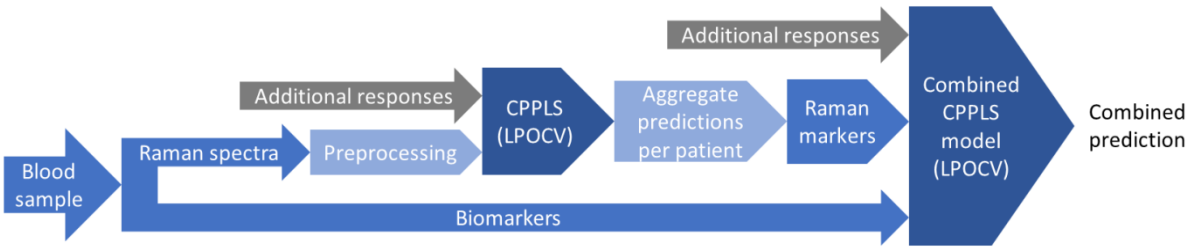
Blood collection protocols and leukocyte isolation procedure have been published previously (4) and are summarized here briefly. Leukocytes were isolated by removing erythrocytes (RBC) by cytolysis method using $NH_4Cl$ solution (VWR, Germany). For this, 500µl EDTA blood was transferred into a falcon tube containing 20 ml of home-made $NH_4Cl$ solution (41.5 g $NH_4Cl$ + 5.0 g $KHCO_3$ + 0.179 g EDTA + 500 ml $H_2O$ with pH 7.1 to 7.4) for the isolation of the leukocytes. After 10 min of incubation at RT, the leukocytes were centrifuged (300 rcf, 5 min, RT). The supernatant was discarded, and the leukocytes were washed and resuspended in 200 µl of PBS. Immediately after isolation, the leukocytes were chemically fixed with 1 ml of formaldehyde-based fixative solution for 10 min at RT (Roth, Germany). Afterwards, the cells were washed successively with PBS. Leukocytes were investigated either immediately after preparation or stored at 4 $^{o}$C until further use (maximum 2 h).

For the Raman measurements, leukocytes were washed with 0.9 % NaCl solution and re-suspended in 0.9 % NaCl with cell concentration of 1 x $10^6$ leukocytes/100 µl. The cells were coated onto $CaF_2$ slides by means of cytospin (Shandon Cytospin 3 Cytocentrifuge, ThermoScientific, Waltham, USA, 6 min, 300 g). To assure immobilization of the leukocytes the $CaF_2$ slides have been pre-coated with 0.2 % gelatin (Sigma-Aldrich, Germany) for 10 min. Gelatin has no significant contribution in the Raman spectra.

For Raman spectroscopic analysis, a commercial upright micro-Raman set-up (CRM 300, WITec GmbH, Germany) was used, which was equipped with a 300 g/mm grating and a back-illuminated deep depletion charge-coupled device camera: DU401 BR-DD, ANDOR, 1024 x 127 pixels cooled down to -60 °C. Single cell measurements of the cells coated on $CaF_2$ slides were carried out as described previously (4).

*Raman spectra pre-processing and statistical analysis of patient's leukocytes*

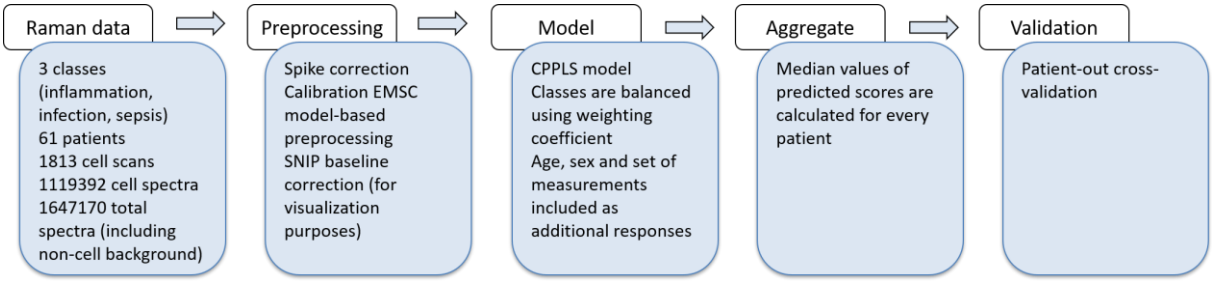
Data analysis was performed using in-house written code in R programming language (5). Prior to the statistical analysis, Raman spectra were preprocessed by a set of operations aimed to suppress corrupting effects. First, the cosmic ray noise was removed, based on one-dimensional discrete Laplace operator response with an automatic threshold (6). Subsequently, Raman spectra were calibrated according to the standard spectra of 4-acetamidophenol (7) and approximated to the same wavenumber axis (from 400 to 3050 $cm^{-1}$ with a step size of 2 $cm^{-1}$). After calibration, we excluded the spectra in which the median intensity of the CH-stretching bands (2850-3000 $cm^{-1}$) was lower than the standard deviation within the silent region (1900-2600 $cm^{-1}$). After excluding noisy spectra, averaged spectra for each cell captured within Raman spectral maps were calculated and standardized by extended multiplicative signal correction (EMSC) algorithm (8). The advantage of the method is that the spectra do not require further normalization and it allows standardizing spectra obtained from different sources using the same reference. As the reference for EMSC, a median spectrum of all cells was used. After EMSC correction, the baseline estimated from the reference spectrum by automated sensitive non-linear iterative peak-clipping (SNIP) algorithm (9, 10) was subtracted from all spectra for visualization purposes. Both the reference and the reference baseline were stored in case if the same preprocessing needs to be applied to a new test data set. Finally, the silent region (1750-2700 $cm^{-1}$) of Raman spectra was excluded from further data analysis.




***Supplemental Figure S2:*** Diagram showing the data processing flow used for the combined prediction based on multiple types of data. CPPLS models were trained using pre-processed Raman data of single leukocytes. For combined CPPLS analysis, the median of the predicted Raman scores per patient was used together with biomarker values in a subsequent combined CPPLS classification model. To suppress model bias, demographic data (age and gender) were used as additional responses during CPPLS model training, but not for obtaining the predictions. The issue of Raman device relocation to the new laboratory during the study was addressed by extending the additional responses with information about two sets of measurements: prior and after the relocation.

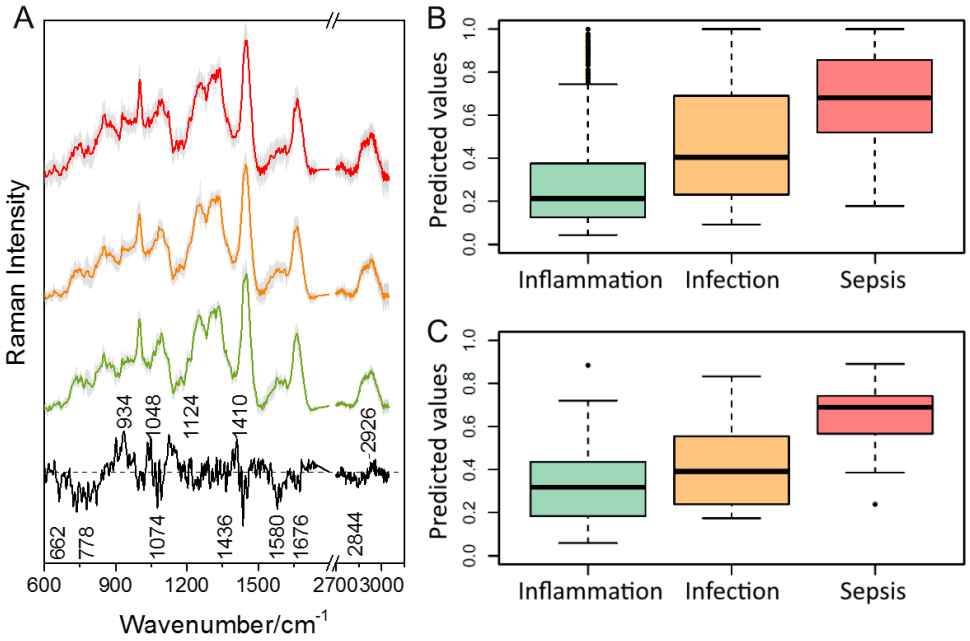
Concentrations of biomarkers (CRP, PCT and IL-6) serving as input values for the combined CPPLS model (Supplemental Figure S2) were scaled to their standard deviation. The scaling

was performed in order to balance the contributions of different biomarkers during the model training and the scaling coefficients were stored in case if model needs to be applied to a new test data set. Demographic information (age and sex) which were fed into the model as additional responses (Supplemental Figure S2) were used without any kind of pre-treatment.



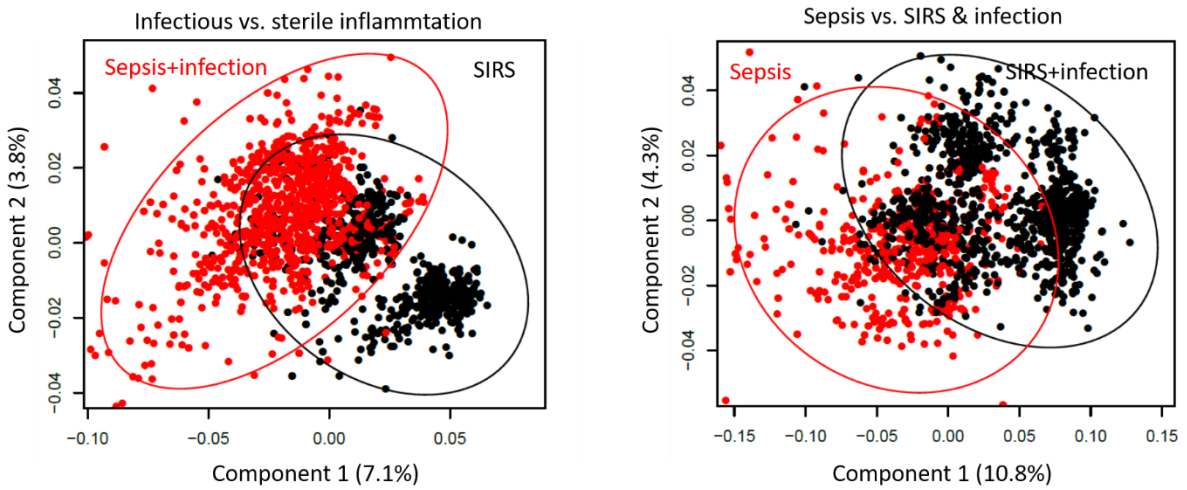

***Supplemental Figure S3*** Workflow showing the analysis of the Raman spectra of leukocytes from the HemoSpec study when using CPPLS model. The results are depicted in Figure 3 of the main manuscript.

The values for balanced accuracy of sepsis and infection detection are presented along with the upper and lower margins of the 95 % credible intervals (CI). These margins were obtained from the posterior distribution for the balanced accuracy. To obtain such distribution, a convolution of posterior distributions for sensitivity and specificity, divided by 2, was defined. According to Bayes theorem, such posterior distributions can be represented by beta distributions with Jeffrey's priors for binomial distribution ($\alpha = \beta = 1/2$) (13).

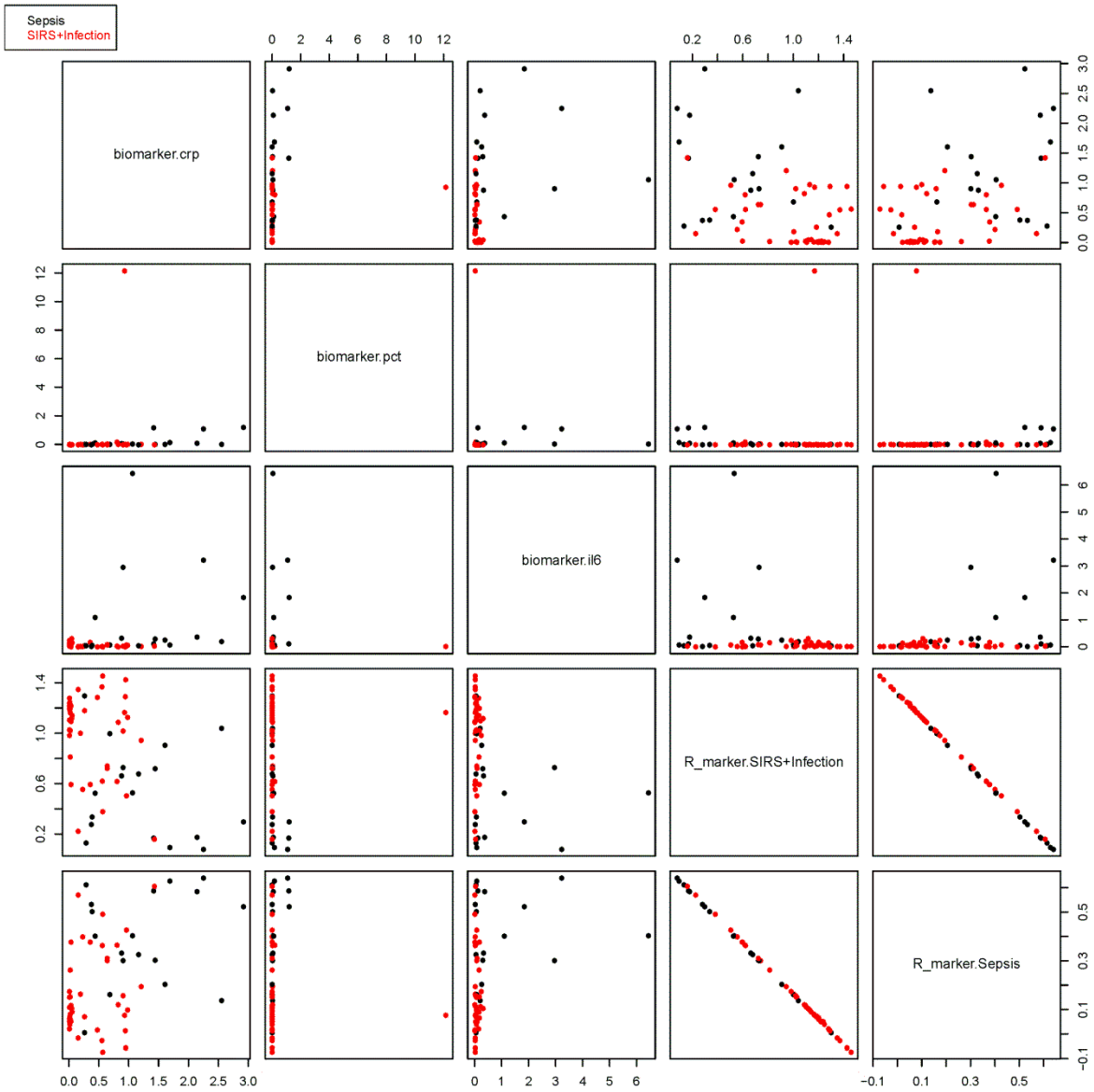



***Supplemental Figure S4*:** Raman spectroscopic data overview using logistic regression analysis to stratify patients. A) Mean Raman spectra together with standard deviation (grey shadow) per patient group (from top to bottom): sepsis (red), infection without organ failure (orange), sterile inflammation (green). The black spectrum at the bottom is the regression coefficient obtained from the Raman spectral model depicted in panel B. Spectra are shifted on the y-axis for better visualization. B) Box plot showing cross-validated regression predictions per cell obtained from the Raman spectral model using logistic regression. C) Predictions from panel B, averaged per patient.

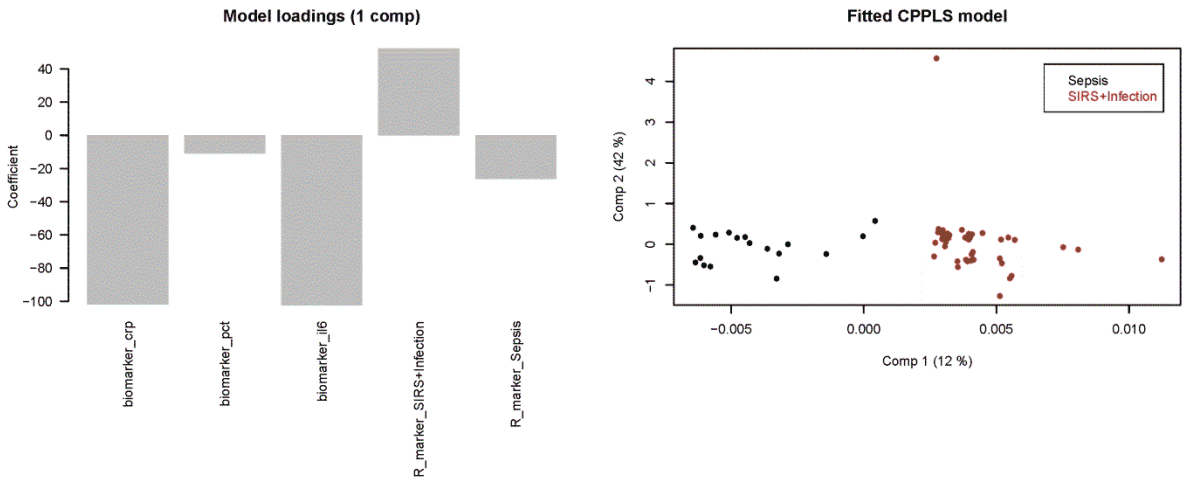
Logistic regression was used for a first visualization of the Raman data. To avoid overfitting, a dimensional reduction was performed prior to the regression by principal component analysis (PCA) and the optimal number of components (8 PC) was selected according to the minimum of root mean square error (RMSE) obtained within leave-one-patient-out cross-validation (LPOCV). As there are three groups in the data set, the values 0, 0.5, and 1 were assigned to the groups: (0) sterile inflammation, (0.5) infection without organ dysfunction, and (1) sepsis, respectively. The logistic regression made it possible to generate a model with a single predicted score, which visualizes the potential of the Raman approach for sepsis detection in a simple way. It also produces a single loading vector that makes it possible to determine which Raman bands are related to the patient's condition severity.



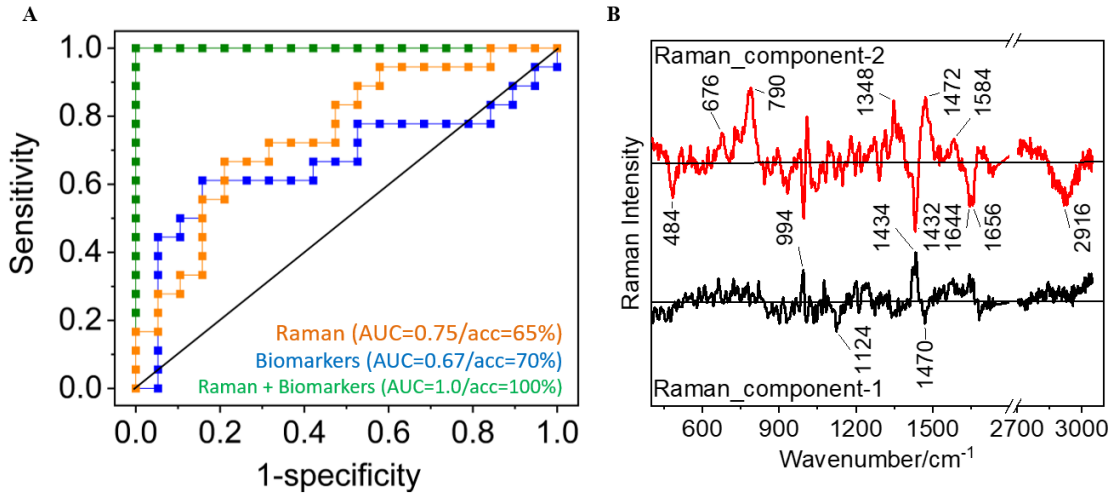

***Supplemental Figure S5*:** Scatter plots with two fitted CPPLS components for the Raman-based models presented in Figure 3 (main manuscript). The score plots demonstrate fitted scores for CPPLS components for individual leukocytes. The values in parenthesis at the axis labels specify the proportion of data variance explained by the respective component. Loading coefficients of component 1 and 2 of the two models are depicted in Figure 4A and 4B, respectively. The loadings of higher components are not depicted as they are more difficult for interpretation. The predicted values obtained from cross-validation of this model were used to obtain the predicted Raman scores. The predictions were obtained using multiple components (15 for infection detection and 5 for sepsis detection). The predicted scores aggregated per patient are depicted in Supplemental Figure S6 and used as input in the combined model (Supplemental Figure S7).




***Supplemental Figure S6*:** Scatter plots of the data used in the model for sepsis detection (see Figure 3B, main manuscript), based on Raman marker scores (R_marker) combined with biomarkers (CRP, PCT, and IL-6). Biomarker scores are normalized to their standard deviation and reflect the measured values for each patient. Raman scores are the aggregated (from all leukocytes per patient) predicted, cross-validated values. Both responses (for sterile inflammation+infection and for sepsis), obtained from cross-validated predictions of the Raman model are included in the combined model as Raman markers because they do not add up precisely to 1 and this slight variation can also improve the prediction, as seen from the loadings (see Figure S7).




***Supplemental Figure S7:*** Model loadings and a fitted scatter plot for the CPPLS model for sepsis detection based on Raman spectroscopic markers displayed in Figure 4B, main manuscript. The loadings depict coefficients only for the first component because a single component was used for the prediction. Two components of fitted CPPLS model are shown on the scatterplot for demonstration. The values in parenthesis at the scatterplot's axis labels specify the proportion of data variance explained by the respective component. The model is based on the two Raman marker scores (R_marker) predicted in a cross-validation loop from the Raman measurements of the leukocytes and combined with biomarkers. The scatter plots for the individual contributing scores which are combined in the CPPLS model are depicted in Supplemental Figure S6.




***Supplemental Figure S8*:** Added value of Raman spectroscopic scores for identification of sepsis patients within the infection cohort (N=37 patients) using canonical powered partial least square (CPPLS) analysis.

**A)** Receiver operating characteristics (ROC) curve to identify sepsis patients within the infection cohort using only the Raman spectroscopic data (orange), only biomarker information (IL-6, CRP, PCT) (blue) and using combined Raman spectroscopic scores and biomarker values (CRP, PCT and IL-6) (green). All models were validated using a leave-one-patient-out cross-validation approach. Predictions for Raman spectroscopic data were obtained on a single-cell level and aggregated to obtain a single value per patient. *AUC* is area under ROC curve and *acc* is balanced accuracy which is average of sensitivity and specificity.

**B)** CPPLS coefficients for the first two latent variables in the Raman spectral model (orange curve in Panel A) for the detection of sepsis patients within the infection cohort. Spectra are shifted on the y-axis for clarity. The solid horizontal lines (for each Raman component) marks zero.

***Supplemental Table S1***: Cause of infection in patients with infection and sepsis

| Patient group | Patient | Pathogen | Pathogen found in |
|---|---|---|---|
| Infection without organ dysfunction | INF 1 | No pathogen | |
| | INF 2 | Streptococcus, Candida, Hepatitis C | Blood |
| | INF 3 | *Klebsiella oxytoca, Acinetobacter* | Sputum |
| | INF 4 | *Escherichia coli, Enterococcus faecalis* | Blood |
| | INF 5 | *Staphylococcus aureus* | Blood |
| | INF 6 | No pathogen | |
| | INF 7 | No pathogen | |
| | INF 8 | *Escherichia coli* | Urine |
| | INF 9 | No pathogen | |
| | INF 10 | No pathogen | |
| | INF 11 | No pathogen | |
| | INF 12 | No pathogen | |
| | INF 13 | No pathogen | |
| | INF 14 | *Borrelia* | |
| | INF 15 | *Klebsiella oxytoca* | Urine |
| | INF 16 | No pathogen | |
| | INF 17 | *Enterobacter cloacae* | Blood and urine |
| | INF 18 | *Escherichia coli* | Sputum |
| | INF 19 | No pathogen | |
| Sepsis | SEP 1 | *Proteus, Staphylococcus aureus* (Methicillin-sensitive) | Tracheal secretion |
| | SEP 2 | *Staphylococcus aureus* | Blood |
| | SEP 3 | Not defined | |
| | SEP 4 | *Serratia, Pseudomonas, Stenotrophomonas, Chryseobacterium* | Tracheal secretion |
| | SEP 5 | *Escherichia coli, Enterococcus faecalis* | Abdomen and blood |
| | SEP 6 | *Staphylococcus aureus* (Methicillin-sensitive) | Blood and abscess |
| | SEP 7 | *Lactobacillus rhamnousus*<br>*Candida* | Blood<br>Abdomen |
| | SEP 8 | No pathogen | |
| | SEP 9 | Coagulase-negative *Staphylococci, Streptococcus* | Abscess |
| | SEP 10 | No pathogen | |
| | SEP 11 | *Proteus* | Tracheal secretion |
| | SEP 12 | *Proteus, Escherichia coli* | Blood and abdomen |
| | SEP 13 | *Klebsiella pneumoniae* | Lung |
| | SEP 14 | Not defined | |
| | SEP 15 | No pathogen | |
| | SEP 16 | *Escherichia coli* | Abdomen |
| | SEP 17 | *Escherichia coli* | Urine |
| | SEP 18 | *Enterococcus faecalis, Candida* | Abdomen |

***Supplemental Table S2***: Overview of number of leukocytes measured per disease group. Leukocyte subtype was determined by staining cells after Raman measurement with Kimura and investigating cellular morphology. The cells which did not belong to the top 3 subtypes, as well as cells which were not clearly identified or were not analyzed through Kimura staining due to technical issues, are summed up in the column "Others". This subtype information was not used in the statistical analysis.

| **Disease group** | **Total leukocytes** | **Neutrophils** | **Lymphocytes** | **Monocytes** | **Others** |
|---|---|---|---|---|---|
| Inflammation | **443** | 294 (66.37%) | 52 (11.74%) | 33 (7.45%) | 64 (14.44%) |
| Infection | **545** | 402 (73.76%) | 62 (11.38%) | 40 (7.34%) | 41 (7.52%) |
| Sepsis | **337** | 161 (47.77%) | 41 (12.17%) | 19 (5.64%) | 116 (34.42%) |
| **Total** | **1325** | **857** | **155** | **92** | **221** |

Ten microliters of Kimura staining solution (toluidine blue, 0.03% light green SF yellowish, saturated saponin and phosphate buffer, pH 6.7, all from PAA Chemical) were placed on the cells and allowed to stand for 5 min. The stained cells were dip-washed gently with distilled water and allowed to dry at room temperature. The Raman mapped cells were manually relocated, and cell type was assigned based on nuclear morphology by investigation through a microscope (Axio Imager Z1, Carl Zeiss micro). Cells that could not be assigned to one of the three subpopulations neutrophils, lymphocytes and monocytes were grouped as "others".

Although low numbers of cells were measured, the relative ratio of the leukocyte subpopulations agrees very well with the numbers obtained from routine clinical chemistry measurements (Table 1, main manuscript). As expected, the major leukocyte subpopulation was neutrophils. The subpopulation assignment of the remaining leukocytes based on Kimura-stained nuclear morphology is difficult in sepsis patients and a larger fraction of cells had to be assigned to the group "others". This fact was also reported by other studies (11-13).

Several scores and measures have been proposed recently to provide measureable insights into the complex host response. e.g., functional assessment of leukocytes (14), expression of HLA-DR expression in monocytes (15), monocyte distribution width (16) or the intensive care infection score (ICIS) which contains five blood-cell-derived parameters characterizing the early innate immune response (17). Among others, increased number of immature neutrophils and myeloid-derived suppressor cells are found in sepsis patients (18, 19). In our HemoSpec trial, we also find more cells in the sepsis population where it is difficult to assign the leukocyte subtype based on Kimura staining.

***Supplemental Table S3***: Overview of sensitivity and specificity obtained by using the canonical powered partial least square (CPPLS) method for all the models (*Sens* stands for sensitivity and *Spec* for specificity) in the HemoSpec trial.

| **Model** | **Raman scores** | | **Biomarker values** | | **Raman scores + Biomarker values** | |
|---|---|---|---|---|---|---|
| | *Sens* | *Spec* | *Sens* | *Spec* | *Sens* | *Spec* |
| Detection of sepsis among infected patients | 0,72 | 0,58 | 0,56 | 0,84 | 1,00 | 1,00 |
| Detection of infection | 0,95 | 0,58 | 0,78 | 0,88 | 0,95 | 0,92 |
| Detection of sepsis | 0,50 | 0,93 | 0,50 | 0,98 | 0,83 | 1,00 |

# References


1. Prescott HC, Iwashyna TJ. Improving Sepsis Treatment by Embracing Diagnostic Uncertainty. Ann Am Thorac Soc 2019;16(4):426-429.
2. De Plano LM, Fazio E, Rizzo MG, et al. Phage-based assay for rapid detection of bacterial pathogens in blood by Raman spectroscopy. Journal of Immunological Methods 2019;465:45-52.
3. Pierrakos C, Velissaris D, Bisdorff M, et al. Biomarkers of sepsis: time for a reappraisal. Critical Care 2020;24(1):287.
4. Ramoji A, Neugebauer U, Bocklitz T, et al. Toward a spectroscopic hemogram: Raman spectroscopic differentiation of the two most abundant leukocytes from peripheral blood. Anal Chem 2012;84(12):5335-5342.
5. Team RDC. R: A Language and Environment for Statistical computing. 2011.
6. Ryabchykov O, Bocklitz T, Ramoji A, et al. Automatization of spike correction in Raman spectra of biological samples. Chemometrics and Intelligent Laboratory Systems 2016;155:1-6.
7. Bocklitz TW, Dörfer T, Heinke R, et al. Spectrometer calibration protocol for Raman spectra recorded with different excitation wavelengths. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 2015;149:544-549.
8. Martens H, Stark E. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. Journal of pharmaceutical and biomedical analysis 1991;9(8):625-635.
9. C. G. Ryan EC, W .L. Griffin, S. H. Sie and D. R. Cousens. SNIP, A Statistics-sensitive background treatment for the quantitative analysis of pixe spectra in Geoscience applications. Nuclear Instruments and Methods in Physics Research 1988;B34:396-402.
10. Morhac M. Peaks: Peaks. 2012.
11. Roberts RE, Hallett MB. Neutrophil Cell Shape Change: Mechanism and Signalling during Cell Spreading and Phagocytosis. Int J Mol Sci 2019;20(6):1383.
12. Demaret J, Venet F, Friggeri A, et al. Marked alterations of neutrophil functions during sepsis-induced immunosuppression. J Leukoc Biol 2015;98(6):1081-1090.
13. Zonneveld R, Molema G, Plotz FB. Analyzing Neutrophil Morphology, Mechanics, and Motility in Sepsis: Options and Challenges for Novel Bedside Technologies. Critical care medicine 2016;44(1):218-228.
14. Albert-Vega C, Tawfik DM, Trouillet-Assant S, et al. Immune Functional Assays, From Custom to Standardized Tests for Precision Medicine. Frontiers in Immunology 2018;9.
15. Leijte GP, Rimmelé T, Kox M, et al. Monocytic HLA-DR expression kinetics in septic shock patients with different pathogens, sites of infection and adverse outcomes. Critical Care 2020;24(1):110.
16. Crouser ED, Parrillo JE, Seymour CW, et al. Monocyte Distribution Width: A Novel Indicator of Sepsis-2 and Sepsis-3 in High-Risk Emergency Department Patients*. Read Online: Critical Care Medicine | Society of Critical Care Medicine 2019;47(8):1018-1025.
17. van der Geest PJ, Mohseni M, Linssen J, et al. The intensive care infection score – a novel marker for the prediction of infection and its severity. Critical Care 2016;20(1):180.
18. Venet F, Demaret J, Gossez M, et al. Myeloid cells in sepsis-acquired immunodeficiency. Annals of the New York Academy of Sciences 2020.
19. Monneret G, Gossez M, Aghaeepour N, et al. How Clinical Flow Cytometry Rebooted Sepsis Immunology. Cytometry Part A : the journal of the International Society for Analytical Cytology 2019;95(4):431-441.