



OPEN

## Peptidome profiling for the immunological stratification in sepsis: a proof of concept study

Martín Ledesma<sup>1,7</sup>, María Florencia Todero<sup>2</sup>, Lautaro Maceira<sup>2</sup>, Mónica Prieto<sup>3</sup>, Carlos Vay<sup>1</sup>, Marcelo Galas<sup>4</sup>, Beatriz López<sup>5</sup>, Noemí Yokobori<sup>6,7</sup> & Bárbara Rearte<sup>2,7</sup>✉

Sepsis has been called *the graveyard of pharmaceutical companies* due to the numerous failed clinical trials. The lack of tools to monitor the immunological status in sepsis constrains the development of therapies. Here, we evaluated a test based on whole plasma peptidome acquired by MALDI-TOF-mass spectrometer and machine-learning algorithms to discriminate two lipopolysaccharide-(LPS) induced murine models emulating the pro- and anti-inflammatory/immunosuppression environments that can be found during sepsis. The LPS group was inoculated with a single high dose of LPS and the IS group was subjected to increasing doses of LPS, to induce proinflammatory and anti-inflammatory/immunosuppression profiles respectively. The LPS group showed leukopenia and higher levels of cytokines and tissue damage markers, and the IS group showed neutrophilia, lymphopenia and decreased humoral response. Principal component analysis of the plasma peptidomes formed discrete clusters that mostly coincided with the experimental groups. In addition, machine-learning algorithms discriminated the different experimental groups with a sensitivity of 95.7% and specificity of 90.9%. Data reveal the potential of plasma fingerprints analysis by MALDI-TOF-mass spectrometry as a simple, speedy and readily transferrable method for sepsis patient stratification that would contribute to therapeutic decision-making based on their immunological status.

Sepsis constitutes one the major causes of death in intensive care units<sup>1,2</sup> with an incidence of 31.5 million cases and 7–9 million deaths annually<sup>3</sup>. The impact of sepsis in developing regions such as Latin America is elusive due to the limited number of prevalence studies, but the high incidence rate of infectious diseases and the inequities in the access to health care indicates that sepsis is an alarming health problem in the region<sup>4–7</sup>.

The redefinition by the Sepsis-3 consensus highlights the need to integrate the information on the etiological agent and the host response status and strongly emphasizes that the dysregulated host response as a key component of sepsis<sup>8</sup>.

From the immunological point of view, sepsis is characterized by an unbalanced response in which both hyper-inflammatory and anti-inflammatory/immunosuppressive responses occur concurrently contributing in different degrees to the sepsis immunopathology in each patient<sup>9–12</sup>. Sepsis has earned the epithet of “the graveyard of pharmaceutical companies” due to the numerous failed clinical trials and this could be related to an inappropriate timing for the application of immunomodulatory interventions<sup>11</sup>. The lack of tools to monitor the host immune status constrains the evaluation of novel immunomodulatory therapeutic approaches. Despite the efforts put in biomarker search in the past decades, only a handful of molecules, such as procalcitonin, proved to be useful in the clinical setting under specific conditions, and the fact that efficient biomarkers not necessarily correspond to key mediators adds further complexity<sup>13</sup>. Moreover, none of them determines the quickly changing immunological status of the host<sup>9–11,14–19</sup>.

<sup>1</sup>Laboratorio de Bacteriología, Departamento de Bioquímica Clínica, Hospital de Clínicas “José de San Martín”, Facultad de Farmacia y Bioquímica, UBA, Av. Córdoba 2351, C1120 CABA, Argentina. <sup>2</sup>Instituto de Medicina Experimental (IMEX) - CONICET - Academia Nacional de Medicina, Pacheco de Melo 3081, C1425AUM CABA, Argentina. <sup>3</sup>Servicio de Bacteriología Especial. Instituto Nacional de Enfermedades Infecciosas (INEI), ANLIS “Dr. C. G. Malbrán”, Av. Vélez Sarsfield 563, C1282AFF CABA, Argentina. <sup>4</sup>Special Program of AMR, Communicable Diseases and Environmental Determinants of Health Department, Pan-American Health Organization, 525 23rd St NW, Washington, DC 20037, USA. <sup>5</sup>Departamento de Bacteriología. INEI, ANLIS “Dr. C. G. Malbrán”, Av. Vélez Sarsfield 563, C1282AFF CABA, Argentina. <sup>6</sup>Servicio de Micobacterias INEI, ANLIS “Dr. C. G. Malbrán”, Av. Vélez Sarsfield 563, C1282AFF CABA, Argentina. <sup>7</sup>Consejo Nacional de Investigaciones Científicas Y Técnicas (CONICET), Godoy Cruz 2290, C1425FQB CABA, Argentina. ✉email: barbararearte@yahoo.com.ar

With the advent of artificial intelligence algorithms including machine learning (ML)<sup>20</sup>, top-down analysis of unlabeled biological samples is gaining ground. This is the case of matrix-assisted light desorption ionization-time of flight-mass spectrometry (MALDI-TOF-MS), a powerful tool that is currently spreading in the clinical setting for microbiological biotyping and in biomedical research for the study of biological samples in several diseases including sepsis<sup>21,22</sup>.

Taking into account that around 50% of all sepsis cases are caused by Gram-negative bacteria<sup>23,24</sup>, we<sup>25–31</sup> and other authors<sup>32,33</sup> demonstrated that both inflammatory and anti-inflammatory/immunosuppression processes that occur during sepsis can be emulated by the administration of lipopolysaccharides (LPS) in murine models. These models proved to be useful because they represent in a discrete manner two immunological states that are believed to underlie simultaneously the complex immunopathogenesis of sepsis.

Herein, we hypothesized that the proinflammatory and anti-inflammatory/immunosuppression environments can be discriminated through the analysis of plasma peptidome profiles generated by MALDI-TOF-MS. Thus, in this study, we developed and evaluated the performance of predictive models based on ML algorithms to discriminate the pro- and anti-inflammatory/immunosuppression states in two LPS-induced in murine models.

## Results

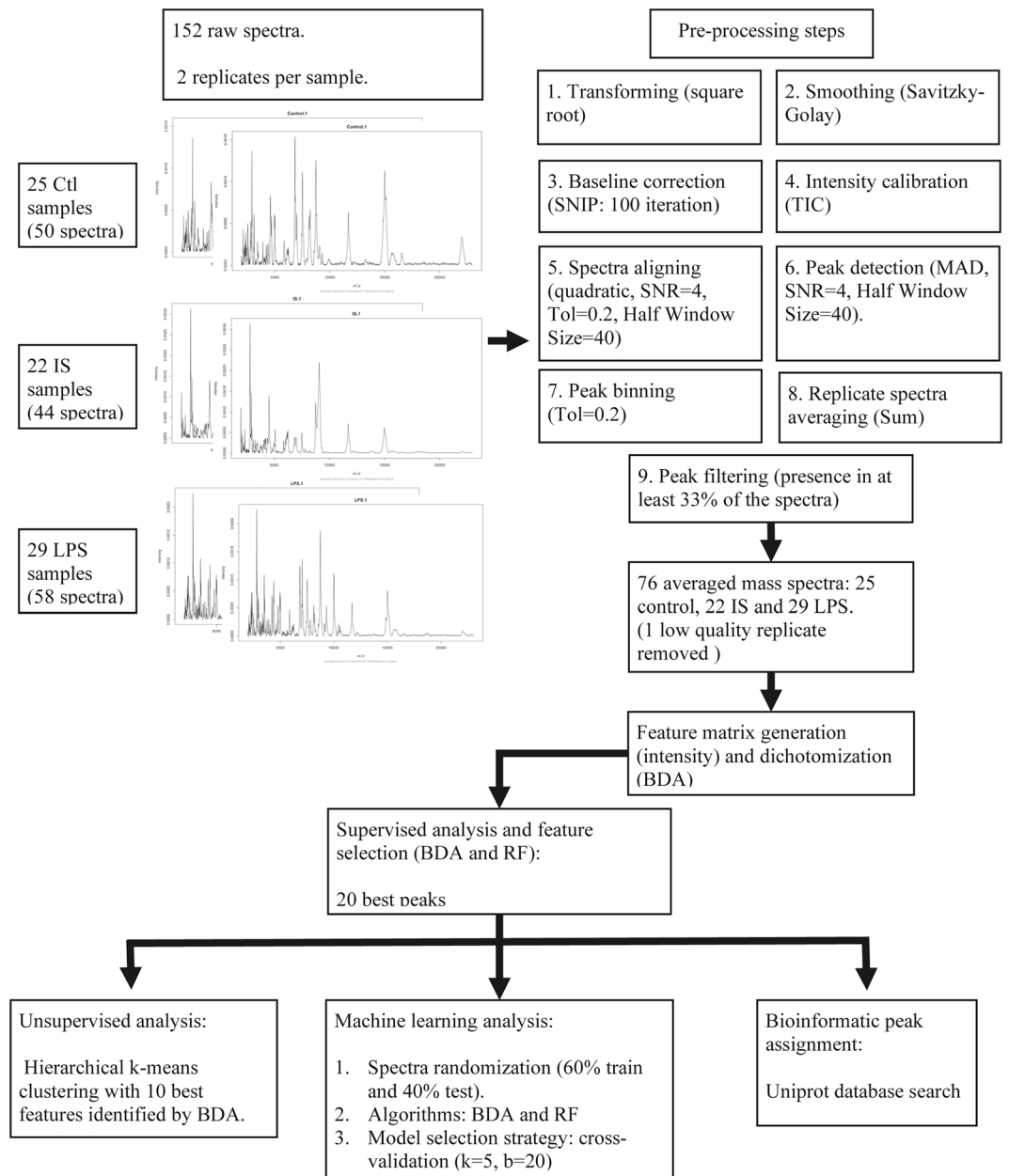
**Simulation of pro- and anti-inflammatory states.** The proinflammatory state was induced with a high dose of LPS (LPS group), whereas the anti-inflammatory/immunosuppression state (IS group) was reached with a scheme of increasing doses of LPS (Supplementary Fig. S1). In order to characterize the two models, immunological, inflammatory and tissue damage markers were studied in the two experimental groups as well as in the control group (CTL group). In accordance with our previous results, the LPS group was characterized by a marked leukopenia (Supplementary Fig. S2a), high plasma levels of pro-inflammatory cytokines such as TNF- $\alpha$  but also of the anti-inflammatory cytokines IL-10 and TGF- $\beta$  (Supplementary Fig. S3a, b). Tissue enzymes indicative of damage were also elevated (Supplementary Fig. S4a–c).

On the other hand, the IS group showed a leukocytosis due to increased numbers of circulating granulocytes, particularly of polymorphonuclear neutrophils (Supplementary Fig. S2b)<sup>25</sup>, along with high plasma levels of TGF- $\beta$  and low levels of pro-inflammatory cytokines (Supplementary Fig. S3c). No signs of tissue damage were observed in this group (Supplementary Fig. S4d–f). As previously described<sup>28,30,31</sup>, IS mice had a profound immunological impairment, accompanied by a marked lymphopenia and an increased number of neutrophils in the spleen. In addition, splenic myeloid CD11b cells expressed higher levels of the inhibitory receptor PDL-1 (Supplementary Fig. S5a, b). Moreover, the IS group showed significantly lower antibody titers upon immunization with SRBCs (Supplementary Fig. S5c). Collectively, these results indicate that the pro- and anti-inflammatory/immunosuppression states were emulated in our murine models.

**MALDI-TOF MS data analysis.** Plasma samples of the CTL, IS and LPS groups were analyzed by MALDI-TOF MS in the 2000 – 20,000 Da range. Representative spectra of each experimental group are shown in Supplementary Fig. S6 and the analysis pipeline is detailed in Fig. 1. A statistical analysis set was implemented to test if MALDI-TOF MS data were useful to discriminate the three experimental groups. First, through a supervised classification algorithm we looked for peaks that best differentiated the CTL, LPS, and IS groups. The 20 best discriminant peaks identified by binary discriminant analysis (BDA) and random forest (RF) algorithms are shown in the Supplementary Fig. S7a and b respectively, which partially overlapped. Of the 20 best peaks selected by the BDA algorithm, the 10 most discriminant peaks were chosen for an unsupervised statistical analysis to plot the hierarchical k-means clustering-PCA clusters (Fig. 2a). Clusters appeared as three non-superimposed groups, and the first two principal components explained 76.6% of the variation with good intra cluster homogeneity. Specifically, cluster 2 achieved 100% (16/16) of homogeneity for spectra corresponding to CTL plasma and 64% (16/25) of CTL spectra fell in this cluster, while for cluster 1 an 85% (29/34) homogeneity was achieved for spectra corresponding to the LPS group and 100% (29/29) of LPS spectra fell in this cluster (Fig. 2b). Lastly, homogeneity of cluster 3 was 77% (20/26) mostly represented by spectra of the IS group, with 91% of the IS spectra (20/22) included in this cluster. This result shows that the immunosuppression/anti-inflammatory and pro-inflammatory states have distinctive plasma peptidome signatures that allow their discrimination.

**Machine learning analysis.** ML algorithms were used to develop a predictive model based on the plasma fingerprints, using a training data set (14 CTL, 13 IS and 15 LPS samples) challenged with a test data set (11 CTL, 9 IS and 14 LPS samples; Fig. 1). To find the most refined model, a dimensionality reduction was performed through the selection of different numbers of distinctive peaks in the training set to train the algorithms, and the performances on the test set were computed for each condition. Specifically, the top 5, 10, 15, and 20 discriminant peaks for BDA and RF algorithms were evaluated by cross-validation (CV; 5 folds and 20 repetitions) and percentages of accuracy, sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) in the test data set were obtained (Table 1).

Between models, RF using the best 20 peaks showed the most consistent values, with sensitivity and specificity values above 90%. The BDA model trained with the best five peaks showed the best performance in terms of detection of positive cases, with a sensitivity above 95% (Table 1); this strategy constitutes the simplest method in terms of both the number of required peaks and computational demands. These results display the potential of plasma peptidome fingerprints to develop predictive models.



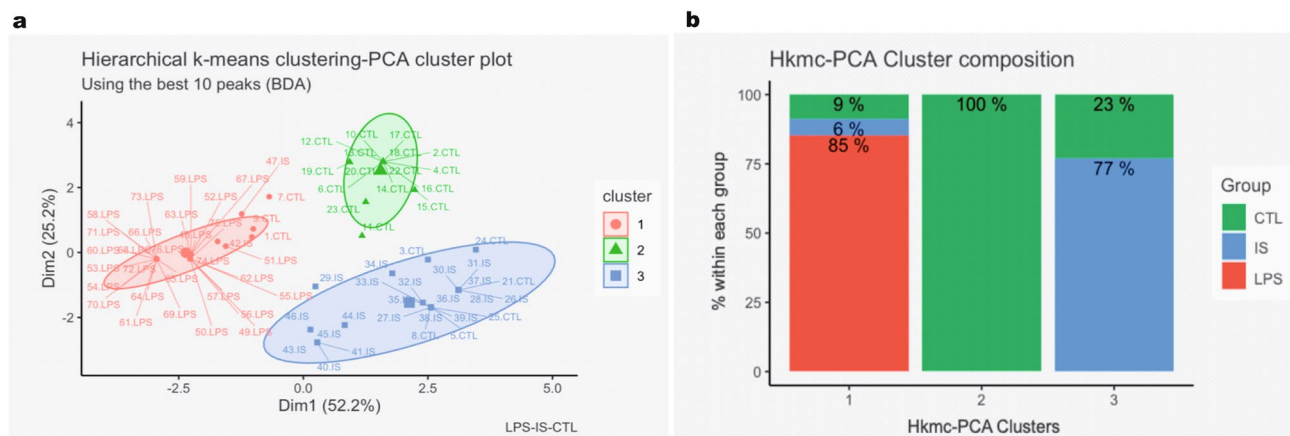
**Figure 1.** MALDI-TOF-MS data analysis pipeline. CTL, control group; IS, immunosuppression/anti-inflammatory group; LPS, pro-inflammatory group; BDA, binary discriminant analysis; RF, random forest.

## Discussion

Sepsis constitutes a highly heterogeneous syndrome of complex pathophysiology. The underlying immunological alterations that include pro-inflammatory, anti-inflammatory and immunosuppressive components<sup>11</sup> are reflected in septic plasma protein composition and varies over time according to the disease severity<sup>34</sup>.

In the last decades, several circulating markers with recognized functions in sepsis were identified. However, none of these markers showed the specificity and sensitivity required for its clinical use<sup>16,19</sup>. Furthermore, methods that rely on immunolabeling such as ELISA or flow cytometry are highly specific and sensitive for the detection of low abundance markers, but require costly equipment and consumables, as well as specific technical knowhow. In addition, these techniques are time consuming, limiting their clinical usefulness for early decision-making.

Regarding biomarker discovery, proteomics proved to be an important source of information. However, in the classical approaches the plasma samples are subjected to depletion of abundant proteins<sup>22</sup> and fractionation to identify individual biomarkers<sup>34,35</sup> which requires more analysis time, specific equipment and higher costs. These constraints led us to explore a top-down strategy that could overcome these difficulties. Thus, having in mind that we aimed to design a readily transferrable method, our approach was meant to be as simple as possible and unlabeled whole plasma were directly read by MALDI-TOF-MS.



**Figure 2.** Unsupervised statistical analysis. **(a)** Hierarchical k-means clustering (Hkmc)-Principal Component Analysis (PCA) cluster plot using the top ten peaks selected by the binary discriminant analysis (BDA) algorithm. Labels contain the mice ID and the experimental groups. PC1 (Dim1, x-axis) and 2 (Dim2, y-axis) are depicted. Spectra were clustered into three groups using the Hkmc algorithm, which are represented with three different colors. 95% confidence ellipses were added around cluster means, assuming a multivariate normal distribution. **(b)** Hkmc-PCA cluster composition. The green color represents the CTL mice group, the blue represents the IS group, and the red color represents the LPS group.

Peaks	Algorithm	A (%)	S (%)	E (%)	PPV (%)	NPV (%)
5	BDA	90.6	95.7	80	90.9	89.8
10	BDA	85.6	92	72.3	87.4	81.1
15	BDA	87.6	93.9	74.5	88.5	85.4
20	BDA	90.1	94.3	81.4	91.4	87.3
5	RF	88.2	95.7	72.7	88	88
10	RF	91.2	91.3	90.9	95.5	83.3
15	RF	88.2	91.3	81.8	91.3	81.8
20	RF	94.1	95.7	90.9	95.7	90.9

**Table 1.** Performance of the classification models with the top 5, 10, 15, and 20 peaks. A, accuracy; S, sensitivity; E, specificity; PPV, positive predictive value; NPV, negative predictive value.

Our results showed that the performance of the test with whole plasma and standard acquisition settings used in routine microbiological biotyping were excellent in terms of discriminatory power. First, an ensemble between supervised and unsupervised algorithms allowed the overall discrimination of plasma samples corresponding to the three experimental groups. These results showed that each state comprises a distinctive MALDI-TOF-MS pattern. A recent study reported a top-down plasma analysis showing that differences between samples from healthy donors and cancer patients could be appreciated without the depletion of large abundant proteins by setting a cut-off at 30KDa<sup>36</sup>, supporting our results.

Furthermore, the performance of our ML algorithms challenged with the test dataset, was able to identify basal, pro-inflammatory and anti-inflammatory/immunosuppressive groups in a very satisfactory way, exceeding 90% of accuracy in some of the tested conditions. The cross validation strategy used herein allows a fine-tuning of the models according to particular needs. Thus, depending on the clinical and epidemiological settings, algorithms with higher sensitivity and lower computational demand like our 5-peaks BDA model could be preferred over those with a good overall accuracy.

The peaks found in the 2-20KDa range mostly correspond to small peptides. Plasma peptidome has been proposed as an interesting source of information for diagnostic purposes<sup>36,37</sup>. These peptides are presumably fragments of larger proteins derived from digestion by proteases or non-enzymatic degradation<sup>38</sup> but the lack of comprehensive digestome/degradome databases is still a limitation to their study<sup>39</sup>. Nevertheless, recent reports demonstrate the importance of plasma peptidome as a useful tool for clinical monitoring in septic shock<sup>40,41</sup>. Although the identification of individual peaks is beyond the objectives of this study, it would be interesting to evaluate them in the future with specific methods with higher resolving power.

A major limitation of this study is related to the animal models chosen. Although LPS-induced inflammation/immunosuppression models are widely used, they do not represent the dynamic physiological changes that occur in sepsis<sup>42</sup>. Endotoxin challenge promotes a faster and transient release of pro/anti-inflammatory mediators compared to the septic processes. Notwithstanding the limitations, these models allowed us to define two discrete immunological states that actually occur concurrently during a septic process. For this reason, we considered

this was the most appropriate approach to test the potential of MALDI-TOF-MS analysis in the proof-of concept phase. We are currently validating our results in a poly-microbial model of sepsis as cecal ligation and puncture model (CLP), with very encouraging results. Considering the complexity of human sepsis, in terms of organ systems involved, etiological agent/s, presence of co-morbidities, etc., it would be interesting to challenge our approach in several preclinical models in addition to CLP, following the recommendations of the "Minimum Quality Threshold in Preclinical Sepsis Studies" (MQTiPSS), : An International Expert Consensus Initiative for Improvement of Animal Modelling in Sepsis" guidelines<sup>43–45</sup> to improve the chances of translation to the clinic.

Finally, our approach has numerous advantages. MALDI-TOF-MS is a simple and speedy method, a critical feature considering the dynamism of sepsis, which is spreading in the clinical setting even in resource-constrained regions. In addition, we used an accessible sample as plasma, without complex pre-processing steps. Finally, we performed our ML analysis in the open source software R that also guarantees the transferability of the bioinformatics tool. In conclusion, our results indicate that plasma peptidome analysis by MALDI-TOF-MS has the potential to be a powerful tool for sepsis patient stratification and therapeutic decision making based on their immunological state.

## Materials and methods

**Animals.** Female and male BALB/c mice (8–12 weeks old) were provided by the IMEX-CONICET-Academia Nacional de Medicina, Buenos Aires, Argentina. Animals were maintained under a 12 h light–dark cycle at  $22 \pm 2$  °C and were fed with standard diet and water ad libitum. Animals were bred and housed in accordance with the NIH Guide and Use of Laboratory Animals<sup>46</sup>. Experimental designs were conducted according to the 3 R's guidelines and approved by the Institutional Animal Care and Use Committee (IACUC) of Institute of Experimental Medicine (IMEX)-CONICET- Academia Nacional de Medicina (Protocol code 71/2019, date of approval 12/27/2019). The authors complied with the ARRIVE guidelines.

**Murine models.** All the inoculation and sample collection schemes are depicted in Supplementary Figure S1.

Briefly, for the inflammatory state model (LPS group), BALB/c mice were inoculated with a lethal dose 50 (LD50) of LPS of *Escherichia coli* O111: B4 (100 µg/ mouse; Sigma-Aldrich, St Louis, MO, USA) intraperitoneally (i. p.)<sup>25</sup>. Plasma samples were obtained at 1.5 h and 6 h after the injection. In the anti-inflammatory/immunosuppression state model (IS group), BALB/c mice were inoculated daily with LPS for 10 consecutive days. The inoculation scheme consisted of increasing doses starting from 5 µg/mouse i. p. for the first 4 days, followed by 50 µg/mouse i. p. for 3 days, and 100 µg/mouse i. p. for the last 3 days<sup>30</sup>. Plasma samples were obtained 24 h after the last LPS dose. A third group of mice inoculated with vehicle (saline solution) served as control (CTL group). The plasma in this group was collected at the same time points indicated in the experimental groups. Peripheral blood was collected by submandibular bleeding method in order to maximize the quality of the sample. Part of the heparinized samples were used for blood cell count and the remaining volume was centrifuged twice (400×g, 10 min at 4°C) and plasma were stored at  $-20$  °C until analysis.

**Immunological, hematological and biochemical parameters.** Peripheral blood leukocytes were counted in a Coulter hematology analyzer (Diatron Abacus Junior Vet, Budapest, Hungary) at 1.5 h and 24 h post LPS for the LPS and the IS groups respectively. Proinflammatory (TNF-α; IL-12p70; IFN-γ; IL-6) and anti-inflammatory cytokines (IL-10, TGF-β) were determined in plasma by ELISA at the indicated time points (OptEIA set; BD Biosciences, San Diego, CA, USA) according to the manufacturer's instructions. The tissue damage markers creatinine (Cre), alanine aminotransferase (ALT) and aspartate transaminase (AST) were determined in plasma, at 6 and 24 h in the LPS and IS groups respectively, using a kit from BioTecnica (Varginha, Minas Gerais, Brazil) and the MINDRAY BS-200E auto-analyzer according to the manufacturer's instructions. For the leukocytes and plasma cytokines determinations, the number of animals (n) used per group for each analysis was of n = 6 to 7 mice; for the biochemical parameters, an n = 5 to 7 mice per group were used.

For flow cytometry analysis, 24 h after the last LPS dose, a splenocyte suspension was obtained and the cells were immunolabeled as described previously<sup>30,31</sup>. The following cell types were evaluated: CD4 (FITC-anti-CD4, clone RM4-5) and CD8 (PECy5-anti-CD8, clone 53-6.7) T-lymphocytes, B-lymphocytes (FITC-anti-CD45R-B220, clone RA3-6B2), polymorphonuclear neutrophils (FITC-anti-CD11b, clone M1/70 and PE-anti-Ly6G, clone 1A8). The expression levels of PDL-1 (PE-anti-PDL-1, clone MIH5) on the CD11b gate of myeloid lineage cells were also evaluated. Labeled monoclonal antibodies were obtained from Invitrogen and BD Pharmingen™. Cells were acquired in a Becton Dickinson FACScan flow cytometer using Cell Quest software (Becton Dickinson, San Jose, CA, USA). An n = 6 to 8 mice per group was used for each analysis.

Primary antibody response to sheep red blood cells (SRBC) was evaluated by immunizing animals 24 h after the last dose of LPS ( $5 \times 10^8$  SRBC /mouse, 0.1 ml i. p.). The anti-SRBC antibody titer was evaluated in serum through a hemagglutination assay 7 days after immunization as described previously<sup>30,31</sup>. An n = 8 mice per group was used in these assays.

**Acquisition of MALDI-TOF-MS spectra.** Plasma samples were analyzed with the MALDI Biotyper System (Bruker Daltonik GmbH, Bremen, Germany). Total number of plasma samples analyzed per group were n = 29 for the LPS group, n = 22 for the IS group and n = 25 for the CTL group. Plasma samples were analyzed with the standard method for microbial biotyping. Briefly, 1 µl of plasma was loaded onto each spot in duplicate and 1 µl of the matrix (alpha-cyano-4-hydroxycinnamic acid matrix in 50% acetonitrile and 2.5% trifluoroacetic acid) was added to each dried spot.

Continuous mass spectra were obtained with a Microflex LT/SH MALDI-TOF mass spectrometer using the flexControl software version 3.4.135.0. Acquisition conditions were ionization mode: LD +, acquisition method: MBT\_FC.par, acquisition mode: qsim, tof Mode: linear, acquisition Operator Mode: linear, and digitizerType: Bruker BD0G5, within a mass range of 2000–20,000 Da. Spectra were obtained in the manual mode, using 60% of laser intensity with 40 laser repetition in each shot and reaching between 400–500 spectra by acquisition. Internal calibration was performed every day following manufacturer's instructions (bacterial test standard; Bruker Daltonik GmbH).

**MALDI-TOF-MS data pre-processing.** We generated a dataset containing 152 mass spectra from plasma samples corresponding to the three mice groups in duplicate. Mass spectra were read as fid/aqus files with MALDIquantForeign (v0.10)<sup>47</sup> and they were processed using MALDIquant (v1.16.2)<sup>48</sup> R package. Briefly, spectra were square root-transformed, smoothed using the Savitzky-Golay algorithm, and were baseline-corrected applying the SNIP process across 100 iterations.

The peaks were detected by a function that estimates the noise of mass spectrometry data by calculating the median absolute deviation (MAD). The signal-to-noise-ratio (SNR) was set up in 4, with a half Window Size of 40 and a tolerance of 0.2 for peak binning. Duplicates were averaged except from one replicate corresponding to the CTL group that was removed due to low quality and 76 averaged spectra were subject to further analysis. Peaks that occurred in less than 33% of the spectra were removed. MALDI-TOF-MS data were transformed by a categorization of the peak intensity, performed with the Binda R package<sup>49</sup> which compares the intensity value with the peak group average, returning 1 when it was equal or higher and 0 when it was lower. The analysis workflow is depicted in the Fig. 1.

**Statistical analysis.** *Immunological, hematological and biochemical parameters.* Graph Pad Prism 6 software (GraphPad Software, San Diego, CA) was used for statistical analysis and plotting. The number of mice or biological replicates (n) analyzed in each experiment were the replicates considered for statistical analysis. The n and the repetitions of each experiment are indicated the figure legends. Values are expressed as the mean  $\pm$  standard error of the mean (SEM) of n samples. No outliers were removed. The assumption test to determine the Gaussian distribution was performed by the Kolmogorov and Smirnov method. For parameters with a Gaussian distribution, the differences between two experimental groups were assessed by unpaired Student's t test, and for multiple group comparisons, as the differences in peripheral blood leukocytes, were analyzed using a one-way ANOVA followed by a Tukey's multiple comparison test. For parameters with a non-Gaussian distribution, comparisons between two experimental groups were performed using Mann-Whitney U test. All statistical tests were interpreted in a two-tailed fashion and a  $P < 0.05$  was considered statistically significant.

**MALDI-TOF-MS data.** Statistical analysis and plotting were performed using Rstudio. Programmed peaks selection was performed in the entire dataset seeking for biomarkers of each class representing the different experimental groups by the Binary Discriminant Analysis (BDA) algorithm, which outputs the t.score (Class means vs. Pooled mean) of each peak. The sign of the t.score indicates the presence (positive t.score) or absence (negative t.score) of that peak in each group. A significance level of 95% was achieved if the modulus of the t.score was equal or higher than 2.5. The best-extracted features were then used to perform a hierarchical k-means clustering-principal component analysis (PCA) with the factoextra R package<sup>50</sup>. The binary distance was used to measure dissimilarity between observations by the ward.D2 agglomeration method with a k of 4. In addition, spectra were analyzed with the random forest (RF) classifier to test another classification model based on a different underlying criterion.

Afterwards, two ML methods were applied in the dataset to train both BDA and RF classifiers in R<sup>51</sup>. Initially, the dataset was randomly partitioned into a training set (60% of plasma samples) and a test set (40% of plasma samples). Programmed feature (peaks) selection was performed with the respective algorithms in the training set seeking for discriminant peaks corresponding to each experimental group. The extracted features were then used to train several ML models. Accuracy, sensitivity, and specificity were used to evaluate the performances of all resulting models in the test set by a cross-validation strategy.

## Data availability

The MS datasets generated for this study as well as the analysis pipelines can be found in GitHub [<https://github.com/MarManLed/SepsisData>].

Received: 25 November 2021; Accepted: 29 June 2022

Published online: 06 July 2022

## References

- Novosad, S. A. *et al.* Vital signs: Epidemiology of sepsis: Prevalence of health care factors and opportunities for prevention. *MMWR Morb. Mortal Wkly. Rep.* **65**, 864–869. <https://doi.org/10.15585/mmwr.mm6533e1> (2016).
- Vincent, J. L. *et al.* International study of the prevalence and outcomes of infection in intensive care units. *JAMA* **302**, 2323–2329. <https://doi.org/10.1001/jama.2009.1754> (2009).
- Fleischmann, C. *et al.* Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *Am. J. Respir. Crit. Care Med.* **193**, 259–272. <https://doi.org/10.1164/rccm.201504-0781OC> (2016).
- Estenssoro, E. *et al.* Predictive validity of sepsis-3 definitions and sepsis outcomes in critically ill patients: A cohort study in 49 ICUs in Argentina. *Crit. Care Med.* **46**, 1276–1283. <https://doi.org/10.1097/CCM.0000000000003208> (2018).
- Estenssoro, E. *et al.* Health inequities in the diagnosis and outcome of sepsis in Argentina: A prospective cohort study. *Crit. Care* **23**, 250. <https://doi.org/10.1186/s13054-019-2522-6> (2019).

6. Azevedo, L. C. P., Cavalcanti, A. B., Lisboa, T., Pizzol, F. D. & Machado, F. R. Sepsis is an important healthcare burden in Latin America: A call to action!. *Rev. Bras. Ter Intensiva* **30**, 402–404. <https://doi.org/10.5935/0103-507X.20180061> (2018).
7. Machado, F. R. *et al.* The epidemiology of sepsis in Brazilian intensive care units (the Sepsis PREvalence Assessment Database, SPREAD): An observational study. *Lancet Infect. Dis.* **17**, 1180–1189. [https://doi.org/10.1016/S1473-3099\(17\)30322-5](https://doi.org/10.1016/S1473-3099(17)30322-5) (2017).
8. Singer, M. *et al.* The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* **315**, 801–810. <https://doi.org/10.1001/jama.2016.0287> (2016).
9. van der Poll, T., van de Veerdonk, F. L., Scicluna, B. P. & Netea, M. G. The immunopathology of sepsis and potential therapeutic targets. *Nat. Rev. Immunol.* **17**, 407–420. <https://doi.org/10.1038/nri.2017.36> (2017).
10. Venet, F., Rimmelé, T. & Monneret, G. Management of sepsis-induced immunosuppression. *Crit. Care Clin.* **34**, 97–106. <https://doi.org/10.1016/j.ccc.2017.08.007> (2018).
11. Rubio, I. *et al.* Current gaps in sepsis immunology: New opportunities for translational research. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(19\)30567-5](https://doi.org/10.1016/S1473-3099(19)30567-5) (2019).
12. van der Poll, T., Shankar-Hari, M. & Wiersinga, W. J. The immunology of sepsis. *Immunity* **54**, 2450–2464. <https://doi.org/10.1016/j.immuni.2021.10.012> (2021).
13. Parlato, M. & Cavaillon, J. M. Host response biomarkers in the diagnosis of sepsis: A general overview. *Methods Mol. Biol.* **1237**, 149–211. [https://doi.org/10.1007/978-1-4939-1776-1\\_15](https://doi.org/10.1007/978-1-4939-1776-1_15) (2015).
14. Al Jalbout, N., Troncoso, R. Jr., Evans, J. D., Rothman, R. E. & Hinson, J. S. Biomarkers and molecular diagnostics for early detection and targeted management of sepsis and septic shock in the emergency department. *J. Appl. Lab. Med.* **3**, 724–729. <https://doi.org/10.1373/jalm.2018.027425> (2019).
15. Gunsolus, I. L., Sweeney, T. E., Liesenfeld, O. & Ledebner, N. A. Diagnosing and managing sepsis by probing the host response to infection: Advances, opportunities, and challenges. *J. Clin. Microbiol.* <https://doi.org/10.1128/JCM.00425-19> (2019).
16. Parlato, M. *et al.* Circulating biomarkers may be unable to detect infection at the early phase of sepsis in ICU patients: The CAP-TAIN prospective multicenter cohort study. *Intensive Care Med.* **44**, 1061–1070. <https://doi.org/10.1007/s00134-018-5228-3> (2018).
17. Schenz, J., Weigand, M. A. & Uhle, F. Molecular and biomarker-based diagnostics in early sepsis: Current challenges and future perspectives. *Expert Rev. Mol. Diagn.* **19**, 1069–1078. <https://doi.org/10.1080/14737159.2020.1680285> (2019).
18. Pierrakos, C. & Vincent, J. L. Sepsis biomarkers: A review. *Crit. Care* **14**, R15. <https://doi.org/10.1186/cc8872> (2010).
19. van Engelen, T. S. R., Wiersinga, W. J., Scicluna, B. P. & van der Poll, T. Biomarkers in sepsis. *Crit. Care Clin.* **34**, 139–152. <https://doi.org/10.1016/j.ccc.2017.08.010> (2018).
20. López Fernández, H., Reboiro-Jato, M., Pérez Rodríguez, J. A., Fdez-Riverola, F. & Glez-Peña, D. Implementing effective machine learning-based workflows for the analysis of mass spectrometry data. *J. Integr. OMICS.* <https://doi.org/10.5584/jiomics.v6i1.196> (2016).
21. Hou, T. Y., Chiang-Ni, C. & Teng, S. H. Current status of MALDI-TOF mass spectrometry in clinical microbiology. *J. Food Drug Anal.* **27**, 404–414. <https://doi.org/10.1016/j.jfda.2019.01.001> (2019).
22. Ludwig, K. R. & Hummon, A. B. Mass spectrometry for the discovery of biomarkers of sepsis. *Mol. Biosyst.* **13**, 648–664. <https://doi.org/10.1039/c6mb00656f> (2017).
23. Dolin, H. H., Papadimos, T. J., Chen, X. & Pan, Z. K. Characterization of pathogenic sepsis etiologies and patient profiles: A novel approach to triage and treatment. *Microbiol. Insights* **12**, 1178636118825081. <https://doi.org/10.1177/1178636118825081> (2019).
24. van Engelen, T. S. R., Joost Wiersinga, W. & van der Poll, T. *Handbook of Sepsis* Chapter 3, 31–43 (2018).
25. Córdoba-Moreno, M. O. *et al.* Consequences of the lack of IL-10 in different endotoxin effects and its relationship with glucocorticoids. *Shock* **52**, 264–273. <https://doi.org/10.1097/SHK.0000000000001233> (2019).
26. Landoni, V. I. *et al.* Tolerance to lipopolysaccharide promotes an enhanced neutrophil extracellular traps formation leading to a more efficient bacterial clearance in mice. *Clin. Exp. Immunol.* **168**, 153–163. <https://doi.org/10.1111/j.1365-2249.2012.04560.x> (2012).
27. Martire-Greco, D. *et al.* all-trans-retinoic acid improves immunocompetence in a murine model of lipopolysaccharide-induced immunosuppression. *Clin. Sci. (Lond.)* **126**, 355–365. <https://doi.org/10.1042/CS20130236> (2014).
28. Montagna, D. R. *et al.* Meta-tyrosine modulates the immune response induced by bacterial endotoxins. *Immunobiology* **225**, 151856. <https://doi.org/10.1016/j.imbio.2019.10.005> (2020).
29. Rearte, B., Landoni, V., Laborde, E., Fernandez, G. & Isturiz, M. Differential effects of glucocorticoids in the establishment and maintenance of endotoxin tolerance. *Clin. Exp. Immunol.* **159**, 208–216. <https://doi.org/10.1111/j.1365-2249.2009.04052.x> (2010).
30. Rearte, B. *et al.* Mifepristone (RU486) restores humoral and T cell-mediated immune response in endotoxin immunosuppressed mice. *Clin. Exp. Immunol.* **162**, 568–577. <https://doi.org/10.1111/j.1365-2249.2010.04262.x> (2010).
31. Rearte, B. *et al.* Dehydroepiandrosterone and metyrapone partially restore the adaptive humoral and cellular immune response in endotoxin immunosuppressed mice. *Innate Immun.* **20**, 585–597. <https://doi.org/10.1177/1753425913502243> (2014).
32. Genga, K. R., Shimada, T., Boyd, J. H., Walley, K. R. & Russell, J. A. The understanding and management of organism toxicity in septic shock. *J. Innate Immun.* **10**, 502–514. <https://doi.org/10.1159/000487818> (2018).
33. Opal, S. M. *et al.* Relationship between plasma levels of lipopolysaccharide (LPS) and LPS-binding protein in patients with severe sepsis and septic shock. *J. Infect. Dis.* **180**, 1584–1589. <https://doi.org/10.1086/315093> (1999).
34. Hayashi, N. *et al.* Multiple biomarkers of sepsis identified by novel time-lapse proteomics of patient serum. *PLoS ONE* **14**, e0222403. <https://doi.org/10.1371/journal.pone.0222403> (2019).
35. Harberts, E. *et al.* Toll-like receptor 4-independent effects of lipopolysaccharide identified using longitudinal serum proteomics. *J. Proteome Res.* **19**, 1258–1266. <https://doi.org/10.1021/acs.jproteome.9b00765> (2020).
36. Cheon, D. H. *et al.* Comprehensive analysis of low-molecular-weight human plasma proteome using top-down mass spectrometry. *J. Proteome Res.* **15**, 229–244. <https://doi.org/10.1021/acs.jproteome.5b00773> (2016).
37. Dufresne, J. *et al.* The plasma peptidome. *Clin. Proteomics* **15**, 39. <https://doi.org/10.1186/s12014-018-9211-3> (2018).
38. Greening, D. W. & Simpson, R. J. Characterization of the low-molecular-weight human plasma peptidome. *Methods Mol. Biol.* **1619**, 63–79. [https://doi.org/10.1007/978-1-4939-7057-5\\_6](https://doi.org/10.1007/978-1-4939-7057-5_6) (2017).
39. Shen, Y. *et al.* Strategy for degradomic-peptidomic analysis of human blood plasma. *J. Proteome Res.* **9**, 2339–2346. <https://doi.org/10.1021/pr901083m> (2010).
40. Aletti, F. *et al.* Peptidomic analysis of rat plasma: Proteolysis in hemorrhagic shock. *Shock* **45**, 540–554. <https://doi.org/10.1097/SHK.0000000000000532> (2016).
41. Bauza-Martinez, J. *et al.* Proteolysis in septic shock patients: Plasma peptidomic patterns are associated with mortality. *Br. J. Anaesth.* **121**, 1065–1074. <https://doi.org/10.1016/j.bja.2018.05.072> (2018).
42. Lewis, A. J., Seymour, C. W. & Rosengart, M. R. Current murine models of sepsis. *Surg. Infect. (Larchmt)* **17**, 385–393. <https://doi.org/10.1089/sur.2016.021> (2016).
43. Hellman, J. *et al.* Part III: Minimum Quality Threshold in Preclinical Sepsis Studies (MQTiPSS) for fluid resuscitation and antimicrobial therapy endpoints. *Shock* **51**, 33–43. <https://doi.org/10.1097/SHK.0000000000001209> (2019).
44. Libert, C. *et al.* Part II: Minimum Quality Threshold in Preclinical Sepsis Studies (MQTiPSS) for types of infections and organ dysfunction endpoints. *Shock* **51**, 23–32. <https://doi.org/10.1097/SHK.0000000000001242> (2019).
45. Zingarelli, B. *et al.* Part I: Minimum Quality Threshold in Preclinical Sepsis Studies (MQTiPSS) for study design and humane modeling endpoints. *Shock* **51**, 10–22. <https://doi.org/10.1097/SHK.0000000000001243> (2019).
46. National Research Council (U.S.). *Guide for the care and use of laboratory animals* 8th edn. (National Academy Press, 2011).

47. MALDIquantForeign: Import/Export Routines for MALDIquant. v. 0.12 (2019).
48. Gibb, S. & Strimmer, K. MALDIquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics* **28**, 2270–2271. <https://doi.org/10.1093/bioinformatics/bts447> (2012).
49. Gibb, S. & Strimmer, K. binda: Multi-class discriminant analysis using binary predictors. R package version 1.0.3. (2015).
50. factoextra: Extract and visualize the results of multivariate data analyses v. R package version 1.0.6 (2019).
51. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).

## Acknowledgements

We thank Dr. Martín Isturiz for his eternal accompaniment and support for the development of this study. His outstanding contribution stems from his singular thoughts. Forever in our hearts, Dr. Isturiz. We also thank Dr. Viviana Ritacco for her thoughtful comments on the manuscript.

## Author contributions

Conceptualization, B.R. and N.Y.; methodology, B.R., M.F.T., L.M.; software, M.L.; validation, M.G., B.L.; formal analysis, B.R., M.F.T., M.L.; investigation, B.R.; resources, M.P., C.V., M.G.; data curation, M.L., N.Y.; writing—original draft preparation, B.R., N.Y.; writing—review and editing, M.L., B.L.; visualization, M.P., C.V.; supervision, B.R.; project administration, B.R.; funding acquisition, B.R. All authors have read and agreed to the published version of the manuscript.

## Funding

This research was funded by the Agencia Nacional de Promoción Científica y Tecnológica, grant number PICT 2015-0412.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15792-5>.

**Correspondence** and requests for materials should be addressed to B.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022