

# Machine-learning algorithms define pathogen-specific local immune fingerprints in peritoneal dialysis patients with bacterial infections



see commentary on page 16

OPEN

Jingjing Zhang<sup>1</sup>, Ida M. Friberg<sup>1,9</sup>, Ann Kift-Morgan<sup>1</sup>, Gita Parekh<sup>2</sup>, Matt P. Morgan<sup>1,3</sup>, Anna Rita Liuzzi<sup>1</sup>, Chan-Yu Lin<sup>1,4</sup>, Kieron L. Donovan<sup>5,6</sup>, Chantal S. Colmont<sup>5</sup>, Peter H. Morgan<sup>7</sup>, Paul Davis<sup>2</sup>, Ian Weeks<sup>8</sup>, Donald J. Fraser<sup>1,5,6,8</sup>, Nicholas Topley<sup>5,8</sup> and Matthias Eberl<sup>1,8</sup>

<sup>1</sup>Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, UK; <sup>2</sup>Mologic Ltd., Bedford Technology Park, Thurleigh, Bedford, UK; <sup>3</sup>Directorate of Critical Care, Cardiff and Vale University Health Board, University Hospital of Wales, Heath Park, Cardiff, UK; <sup>4</sup>Kidney Research Center, Chang Gung Memorial Hospital, Chang Gung University, College of Medicine, Taoyuan City, Taiwan; <sup>5</sup>Wales Kidney Research Unit, Heath Park Campus, Cardiff, UK; <sup>6</sup>Directorate of Nephrology and Transplantation, Cardiff and Vale University Health Board, University Hospital of Wales, Heath Park, Cardiff, UK; <sup>7</sup>Cardiff Business School, Cardiff University, Cardiff, UK; and <sup>8</sup>Systems Immunity Research Institute, Cardiff University, Cardiff, UK

The immune system has evolved to sense invading pathogens, control infection, and restore tissue integrity. Despite symptomatic variability in patients, unequivocal evidence that an individual's immune system distinguishes between different organisms and mounts an appropriate response is lacking. We here used a systematic approach to characterize responses to microbiologically well-defined infection in a total of 83 peritoneal dialysis patients on the day of presentation with acute peritonitis. A broad range of cellular and soluble parameters was determined in peritoneal effluents, covering the majority of local immune cells, inflammatory and regulatory cytokines and chemokines as well as tissue damage-related factors. Our analyses, utilizing machine-learning algorithms, demonstrate that different groups of bacteria induce qualitatively distinct local immune fingerprints, with specific biomarker signatures associated with Gram-negative and Gram-positive organisms, and with culture-negative episodes of unclear etiology. Even more, within the Gram-positive group, unique immune biomarker combinations identified streptococcal and non-streptococcal species including coagulase-negative *Staphylococcus* spp. These findings have diagnostic and prognostic implications by informing patient management and treatment choice at the point of care. Thus, our data establish the power of non-linear mathematical models to analyze complex biomedical datasets and highlight key pathways involved in pathogen-specific immune responses.

*Kidney International* (2017) **92**, 179–191; <http://dx.doi.org/10.1016/j.kint.2017.01.017>

**Correspondence:** Matthias Eberl, Division of Infection and Immunity, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK. E-mail: [eberlm@cf.ac.uk](mailto:eberlm@cf.ac.uk)

<sup>9</sup>Present address: School of Environment and Life Sciences, University of Salford, Salford M5 4WT, UK.

Received 1 November 2016; revised 4 January 2017; accepted 12 January 2017; published online 17 March 2017

**KEYWORDS:** biomarkers; inflammation; machine learning methods; microbial infection; peritoneal dialysis

Copyright © 2017, International Society of Nephrology. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

The immune system is an intricate network of specialized cell types and molecular structures evolved to sense and target invading pathogens, control and clear the infection, and repair and restore the integrity of affected tissues and organs. The human body is constantly exposed to a plethora of pathogenic, opportunistic, commensal, and environmental microorganisms and has developed mechanisms to discriminate between harmful and harmless colonization through receptors and pathways that specifically recognize pathogen and danger-associated molecular patterns and unique antigenic epitopes.<sup>1–4</sup> However, unequivocal evidence that the human immune system distinguishes between different types of organisms in a physiologic context and mounts appropriate responses that are distinct enough to be exploited as rapid diagnostic indicators driving appropriate therapy is lacking.<sup>5–10</sup>

Individuals with end-stage kidney disease receiving peritoneal dialysis (PD) serve as well-defined exemplar of a clinical infection requiring immediate medical intervention. Peritonitis is a common complication of PD and remains a major cause of early dropout, technical failure, and mortality.<sup>11,12</sup> In addition to its clinical relevance for individuals with end-stage kidney failure who depend on dialysis as life-saving renal replacement therapy, PD offers unparalleled insights into complex local cell interactions and molecular mechanisms that underpin the clinical severity of infectious episodes and that are readily translatable to improve patient management and outcomes.<sup>13–15</sup> Importantly, peritoneal effluent can be sampled repeatedly and noninvasively, thus providing early and continuous access to the site of infection, even before antibiotic treatment is

initiated. Moreover, PD-related peritonitis is caused by a wide spectrum of bacterial species, thereby allowing the study of acute responses to defined groups of organisms under closely related conditions.<sup>6,15</sup> However, although highly elevated white cell counts with a proportion of >50% granulocytes in the peritoneal effluent are used as indicators of peritonitis, only little progress has been made with regard to reliable discrimination between infection and noninfectious inflammation. Culture-based diagnosis of infection is slow and unsatisfactory, and rapid identification of disease-causing organisms using molecular techniques with sufficient sensitivity and specificity remains a challenge.<sup>11,12,16</sup> Treatment of peritonitis therefore continues to be largely empirical, and early but untargeted treatment with broad-spectrum antibiotics and antifungals is recommended.<sup>12,17</sup>

As alternative to organism-based diagnostics, we aimed at exploiting the human host response and used a systematic approach based on machine learning algorithms to identify diagnostically relevant, pathogen-specific local immune fingerprints in PD patients who presented with acute peritonitis. The introduction of “big data” technologies in biomedical sciences to address the complexity of the molecular and cellular mechanisms underlying disease has brought about an increasing need for advanced statistical models, machine learning, and pattern recognition techniques. In particular, wrapped feature selection methods have proved highly efficient for finding the best feature combination compared with time-consuming exhaustive searches.<sup>18</sup> Support Vector Machines (SVMs) are data-driven methods that try to find a separating hyperplane with the maximal “margin” for classification problems and that can also be used for regression or density estimation.<sup>19–21</sup> Artificial neural networks (ANNs) are inspired by biological neural networks with data processing from the input through a network of multiple nodes that are connected with each other in different layers.<sup>22–24</sup> Random Forests (RFs) are ensemble methods constructed on multiple decision trees for classification and regression.<sup>25–27</sup> By combining biomarker measurements during acute peritonitis and feature selection approaches based on SVMs, ANNs, and RFs, our findings demonstrate the power of advanced mathematical models to analyze complex biomedical datasets and highlight key pathways involved in pathogen-specific inflammatory responses at the site of infection. The observation that different infecting bacteria induce consistent and unique local immune responses has immediate diagnostic implications at the point of care by directing appropriate antibiotic treatment before conventional microbiological culture results become available.

## RESULTS

### Local immune biomarkers form distinct hierarchical clusters

In order to define combinations of local biomarkers that would constitute relevant disease-specific immune fingerprints, we measured a broad range of cellular and soluble biomarkers in 83 PD patients presenting with microbiologically well-defined episodes of acute peritonitis (Table 1).

To cover the breadth and the complexity of local inflammatory and regulatory immune responses during early infection, these biomarkers included frequencies and total numbers of infiltrating leukocytes as well as levels of common cytokines, chemokines, and tissue damage-associated molecules, the majority of which were elevated during acute peritonitis compared with baseline parameters in stable individuals (Supplementary Table S1). Perhaps not surprisingly, due to the redundant roles of many inflammatory mediators within the human immune system, some of the 49 biomarkers correlated with each other and formed 5 distinct hierarchical clusters during acute peritonitis (Figure 1). These data suggested that a signature comprising as few as 5 parameters might already suffice to define a reliable immune fingerprint.

### Feature selection methods define local fingerprints associated with Gram-negative infections

We next divided the patients into groups according to the type of infecting organism. We initially attempted to define immune fingerprints that would reliably discriminate patients presenting with Gram-negative infections against all other cases of peritonitis (Supplementary Table S2A), based on our earlier observation of certain differences between Gram-negative and Gram-positive infections using logistic regression analyses.<sup>6</sup> To this end, recursive feature elimination was used by evaluating the model performance according to the area under the receiver operating characteristic curve achieved and eliminating the least important features in each step. To reduce variability, 5 rounds of resampling methods were applied in the outer layer of the iteration, and cross-validation was used to avoid overfitting. These steps clearly demonstrated that Gram-negative infections were associated with unique different immune fingerprints. Figure 2a shows the number of features changing during feature elimination and the corresponding performance based on 3 different models, using SVMs, ANNs, and RFs. Whilst all 3 models successfully discriminated between Gram-negative infections and all other causes of peritonitis, RF-based feature elimination showed the best average performance, with the optimum biomarker combination comprising 8 features (area under receiver operating characteristic curve [AUC] = 0.993; sensitivity = 98.5% and specificity = 92.6%). In comparison, SVMs and ANNs were far less powerful for the recursive elimination of pathogen-related biomarkers, reaching overall lower degrees of sensitivity and specificity and requiring combinations comprising 10 and 30 features, respectively (Figure 2a).

The top 5 and 10 individual biomarkers selected by the 3 different models and the corresponding average performance of the models based on combinations of these biomarkers are listed in Supplementary Table S2B. Of note, although the 3 models yielded different sets of biomarkers, the frequencies of  $V\gamma 9^+$  and  $V\delta 2^+$  T cells within peritoneal T cells featured prominently in each. These findings appear to concur with our previous data suggesting a key role for  $V\gamma 9/V\delta 2$  T cells in Gram-negative infections and emphasize

**Table 1 | Characteristics of the patient cohort investigated in this study**

	Stable		No growth		Gram-positive		Gram-negative	
	no.	%	no.	%	no.	%	no.	%
Age, yr, mean $\pm$ SEM	70.1 $\pm$ 2.2		63.8 $\pm$ 2.9		65.9 $\pm$ 2.1		68.6 $\pm$ 3.3	
18–40	1	4.2	1	5.3	4	8.5	1	5.9
40–50	1	4.2	1	5.3	3	6.4	0	0
50–60	0	0	1	5.3	6	12.8	3	17.7
60–70	8	33.3	11	57.9	14	29.8	3	17.7
70–80	10	41.7	3	15.8	12	25.5	8	47.1
$\geq 80$	4	16.7	2	10.5	8	17.0	2	11.8
Sex								
Male	18	75	11	57.9	32	68.1	9	52.9
Female	6	25	8	42.1	15	31.9	8	47.1
Days on PD, mean $\pm$ SEM	1165.1 $\pm$ 186.5		638.8 $\pm$ 163.5		1022.9 $\pm$ 144.2		768.5 $\pm$ 245.8	
0–30	0	0	3	15.8	1	2.1	0	0
30–360	1	4.2	4	21.1	15	31.9	9	52.9
360–720	9	37.5	6	31.6	8	17.0	3	17.7
720–1800	8	33.3	4	21.1	16	34.0	3	17.7
1800–3600	5	20.8	2	10.5	6	12.8	1	5.9
$\geq 3600$	1	4.2	0	0	1	2.1	1	5.9
Hypertension								
Yes	12	50	2	10.5	14	29.8	7	41.2
No	12	50	17	89.5	33	70.2	10	58.8
Diabetes								
Yes	7	29.2	6	31.6	13	27.7	7	41.2
No	17	70.8	13	68.4	34	72.3	10	58.8
Technique failure								
Days 0–14	N/A		1	5.3	2	4.3	2	11.8
Days 14–30	N/A		2	10.5	3	6.5	2	11.8
Days 30–90	N/A		1	5.3	6	12.9	4	23.5
Mortality								
Days 0–14	N/A		1	5.3	0	0	1	5.9
Days 14–90	N/A		1	5.3	0	0	1	5.9

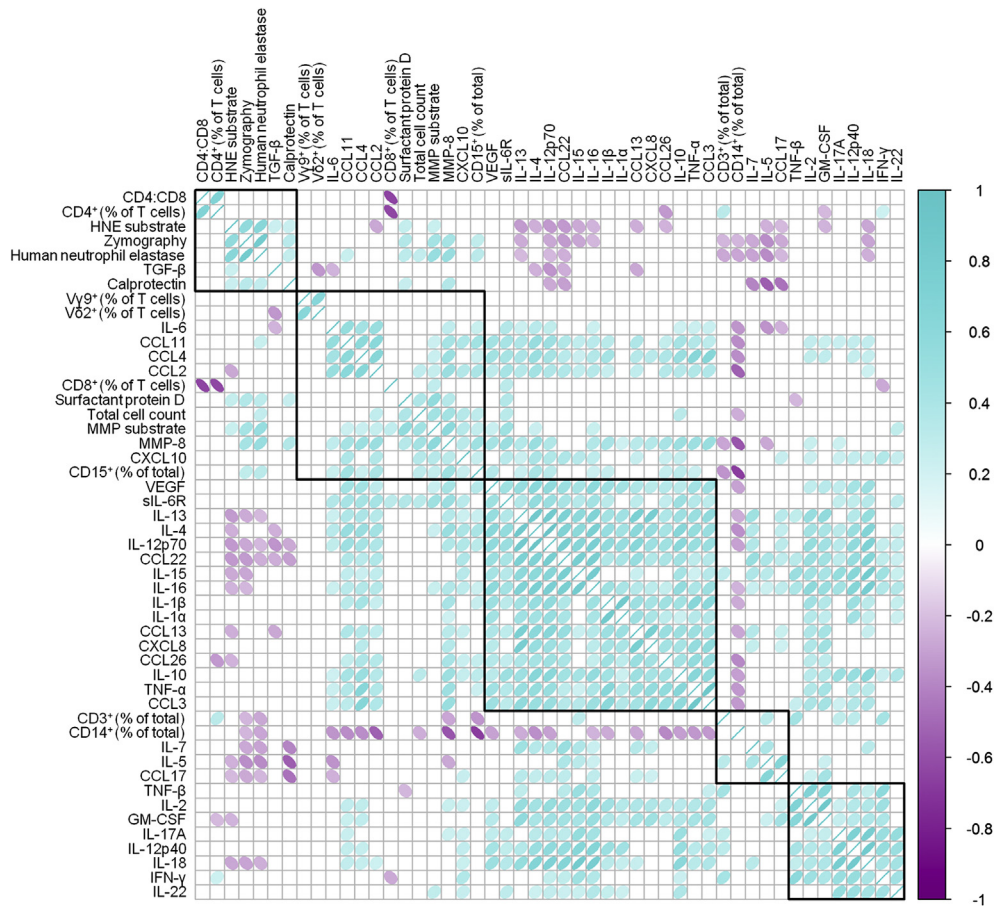
N/A, not available; PD, peritoneal dialysis.

the diagnostic potential of those cells at the point of care.<sup>15,28</sup> Soluble biomarkers of particular interest for the prediction of Gram-negative infections using the RF model included local levels of tumor necrosis factor (TNF)- $\alpha$ , interleukin (IL)-12p40, and vascular endothelial growth factor (VEGF). Taken together, our findings demonstrate that RF models showed striking performance at combinations of 5 and fewer biomarkers, thereby making it the algorithm of choice for a clinically viable prediction of the causative pathogen in individuals presenting with PD-related peritonitis.

#### Internal validation of individual biomarkers constituting Gram-negative immune fingerprints

We next sought to validate the findings from the feature elimination process by assessing the distribution and performance of the top 5 biomarkers that had been selected by the RF model. Using skewness as a measure of symmetry and kurtosis as a measure of peakedness, all 5 biomarkers showed very low positive skewness and only limited or even negative kurtosis in patients presenting with Gram-negative infections (Figure 2b). In contrast, these 5 biomarkers had generally higher skewness and considerable positive kurtosis in patients with other episodes of peritonitis, especially in the case of V $\delta$ 2<sup>+</sup> T-cell frequencies and levels of IL-12p40 and VEGF. Among the top 5 biomarkers, the local frequencies of V $\gamma$ 9<sup>+</sup>

and V $\delta$ 2<sup>+</sup> T cells and the levels of TNF- $\alpha$  on their own already showed relatively good sensitivities and specificities to identify Gram-negative infections with AUCs  $\geq 0.75$  for each biomarker, much more so than levels of IL-12p40 and VEGF (Figure 2c), but far lower than the full signature that reached an AUC of 0.99 (Supplementary Table S2B). This conclusion was supported by classic analyses using the Mann-Whitney *U* test showing that levels of V $\gamma$ 9<sup>+</sup> and V $\delta$ 2<sup>+</sup> T cells and TNF- $\alpha$ , but not of IL-12p40 and VEGF, were markedly different between patients with Gram-negative infections and patients with other episodes of peritonitis (Figure 2d). These results confirmed the importance of the biomarkers selected by recursive feature elimination, especially of local levels of V $\gamma$ 9<sup>+</sup> and V $\delta$ 2<sup>+</sup> T cells and of TNF- $\alpha$ , in identifying Gram-negative organisms. However, our analyses also identified shortcomings of conventional statistical methods that were especially apparent when visualizing the individual readings across all patients in the form of a heat map, where the overall differences between Gram-negative and other episodes of peritonitis were not very pronounced (Figure 2e). Overall, the performances of the individual biomarkers lagged behind their combined performance in an RF model, demonstrating the importance of defining complex signatures comprising distinct biomarkers and of assessing their relationships in nonlinear models.



**Figure 1 | Correlation analysis of local biomarkers in a total of 83 peritoneal dialysis patients on the day of presentation with acute peritonitis.** Ellipses depict the correlation coefficients for each pair of biomarkers in the corresponding cell of the matrix, with the direction of the dip and the color of the shading representing positive and negative correlations, respectively. Only pairs with significant correlations ( $P < 0.05$ ) are shown. Analyses were performed using the corplot R and Hmisc R packages. GM-CSF, granulocyte macrophage colony-stimulating factor; HNE, human neutrophil elastase; IFN- $\gamma$ , interferon- $\gamma$ ; IL, interleukin; MMP, matrix metalloproteinase; sIL-6R, soluble IL-6 receptor; TGF- $\beta$ , transforming growth factor- $\beta$ ; TNF- $\alpha$ , tumor necrosis factor- $\alpha$ ; VEGF, vascular endothelial growth factor.

### Patients with culture-negative episodes of peritonitis show distinct local immune fingerprints associated with milder inflammation

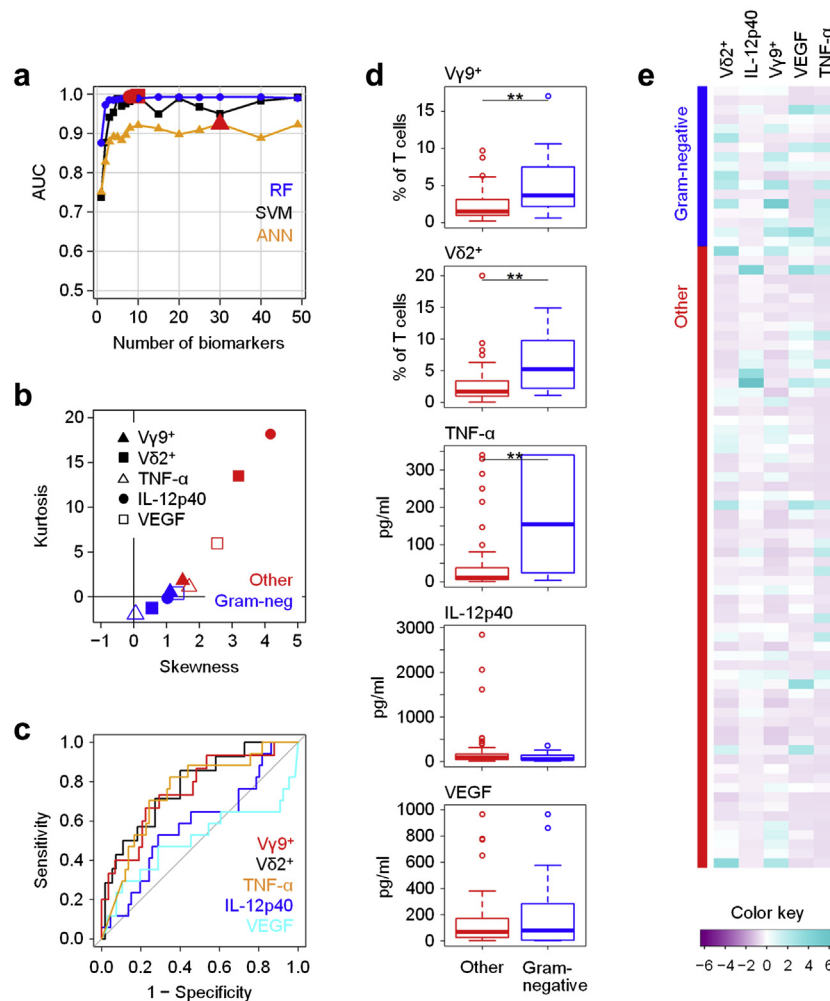
Although microbiological culture remains the method of choice for diagnosis of infection, a considerable proportion of samples does not yield any culture results, thereby not allowing a reliable designation of the underlying cause of the inflammatory episode, which may or may not be infectious.<sup>29</sup> Patients with culture-negative episodes often show less severe inflammation and have better clinical outcomes, suggesting that local biomarkers might aid in the diagnosis, treatment, and prognosis of such episodes. Here, RF models selected biomarker signatures that reliably distinguished samples with no growth from cases with confirmed bacterial infection (Figure 3a, Supplementary Tables S3A and S3B). The top 5 biomarkers all showed great potential in identifying culture-negative episodes that were characterized by relatively low total cell counts, an increased proportion of CD14<sup>+</sup> monocytes/macrophages in the cellular infiltrate, and lower levels of IL-1 $\beta$ , matrix metalloproteinase (MMP)-8 and the chemokine CCL4 compared with confirmed infections

(Figure 3b–e, Supplementary Table S3B). Despite the heterogeneity of this patient group, in which the failure to grow organisms might be due to inappropriate sampling, poor handling and culture techniques, low organism numbers, ongoing treatment with antibiotics for unrelated infections, or nonmicrobial disease such as sterile inflammation or viral infection, our findings indicate that culture-negative episodes are immunologically distinct from confirmed cases of bacterial peritonitis and are characterized by a less severe inflammatory response.

### Different types of Gram-positive bacteria induce distinct immune responses that allow discrimination between organism subgroups

We next sought to define immune fingerprints in PD patients with confirmed infections caused by Gram-positive bacteria (Supplementary Table S4A). Here, feature elimination models were able to discriminate between Gram-positive infections and other episodes, yet required combinations of  $\geq 30$  biomarkers for optimal performance and failed to reach satisfactory AUCs (Supplementary Figure S1). The best





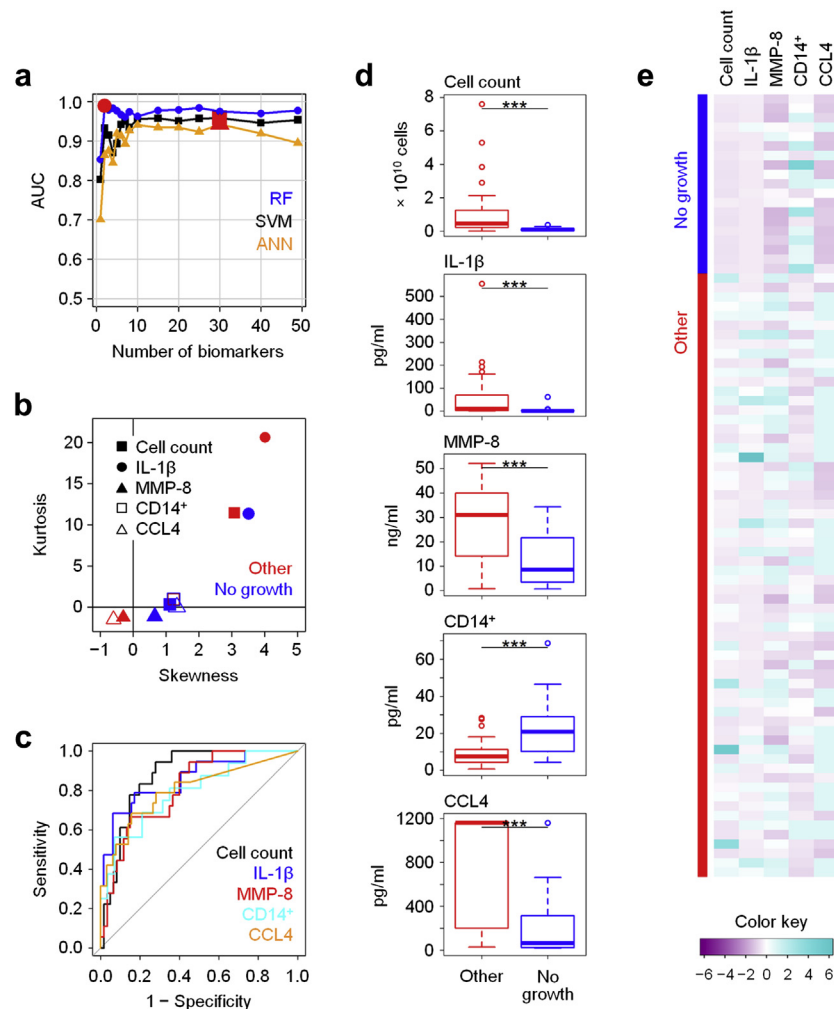
**Figure 2 | Identification of local immune fingerprints associated with peritonitis caused by Gram-negative bacteria.** (a) Performance of recursive feature elimination models based on Random Forest (RF), Support Vector Machines (SVM), and artificial neural networks (ANN) for the prediction of Gram-negative infections ( $N = 17$ ) against all other episodes of peritonitis ( $N = 66$ ), shown as area under the receiver operating characteristic curve (AUC) depending on the number of biomarkers. Red symbols depict the maximum AUC achieved for each model. (b) Kurtosis and skewness of the top 5 biomarkers selected by RF-based feature elimination. (c) Receiver operating characteristic analysis showing specificity and sensitivity of the top 5 biomarkers. (d) Tukey plots of the top 5 biomarkers in patients with confirmed Gram-negative infections and with all other episodes of peritonitis, as assessed by Mann-Whitney tests ( $**P < 0.01$ ). (e) Heat map showing the top 5 biomarkers across all patients presenting with acute peritonitis. IL, interleukin;  $TNF-\alpha$ , tumor necrosis factor- $\alpha$ ; VEGF, vascular endothelial growth factor.

5 biomarkers together only reached an AUC of 0.711 in the RF model, and none of the individual markers on their own—IL-17A, IL-12p40, interferon- $\gamma$ , IL-1 $\beta$ , and total cell count—exceeded an AUC of 0.72 or stayed well below that value (Supplementary Table S4B).

Of note, there is considerable heterogeneity in the Gram-positive organisms causing PD-related peritonitis, comprising streptococci, staphylococci, coryneforms, and other bacteria that cause clinically distinct diseases with different outcomes and require different antibiotics.<sup>12</sup> We therefore attempted to define the pathogen-specific immune responses to subtypes of Gram-positive organisms. These analyses demonstrated that in the Gram-positive group, streptococcal infections caused by *Streptococcus* and *Enterococcus* species were associated with markedly distinct immune responses compared with all other cases of

peritonitis (Figure 4a, Supplementary Table S5A). The most promising biomarker combination consisted of the local levels of IL-1 $\beta$ ,  $TNF-\beta$ , and IL-15 together with the enzymatic activity of MMPs in the PD effluent, as measured by specific cleavage of a fluorogenic MMP substrate and by gelatin zymography on gel electrophoretic separation (Figure 4b). Individually, IL-1 $\beta$ ,  $TNF-\beta$ , and zymography showed differences between streptococcal infections and all other patients, yet only an RF-based model revealed their full diagnostic potential with an AUC of 0.969 (Figure 4c–e, Supplementary Table S5B).

Similar to the definition of *Streptococcus/Enterococcus*-specific immune fingerprints, nonstreptococcal Gram-positive infections caused by *Staphylococcus aureus*, coagulase-negative *Staphylococcus* spp. (CNS) and *Corynebacterium* spp. were associated with biomarker signatures that distinguished them



**Figure 3 | Local immune fingerprints in culture-negative episodes of peritonitis.** (a) Performance of Random Forest (RF), Support Vector Machine (SVM), and artificial neural network (ANN)-based feature elimination models for the prediction of culture-negative episodes (no growth,  $N = 19$ ) against microbiologically confirmed infections (other,  $N = 64$ ), shown as area under the curve (AUC) depending on the number of biomarkers. Red symbols depict the maximum AUC for each model. (b) Kurtosis and skewness of the top 5 biomarkers selected by RF-based feature elimination. (c) Receiver operating characteristic analysis showing specificity and sensitivity of the top 5 biomarkers. (d) Tukey plots of the top 5 biomarkers in patients with culture-negative peritonitis and with infectious (other) episodes of peritonitis, as assessed by Mann-Whitney tests ( $***P < 0.001$ ). (e) Heat map showing the top 5 biomarkers across all patients presenting with acute peritonitis. IL, interleukin; MMP, matrix metalloproteinase.

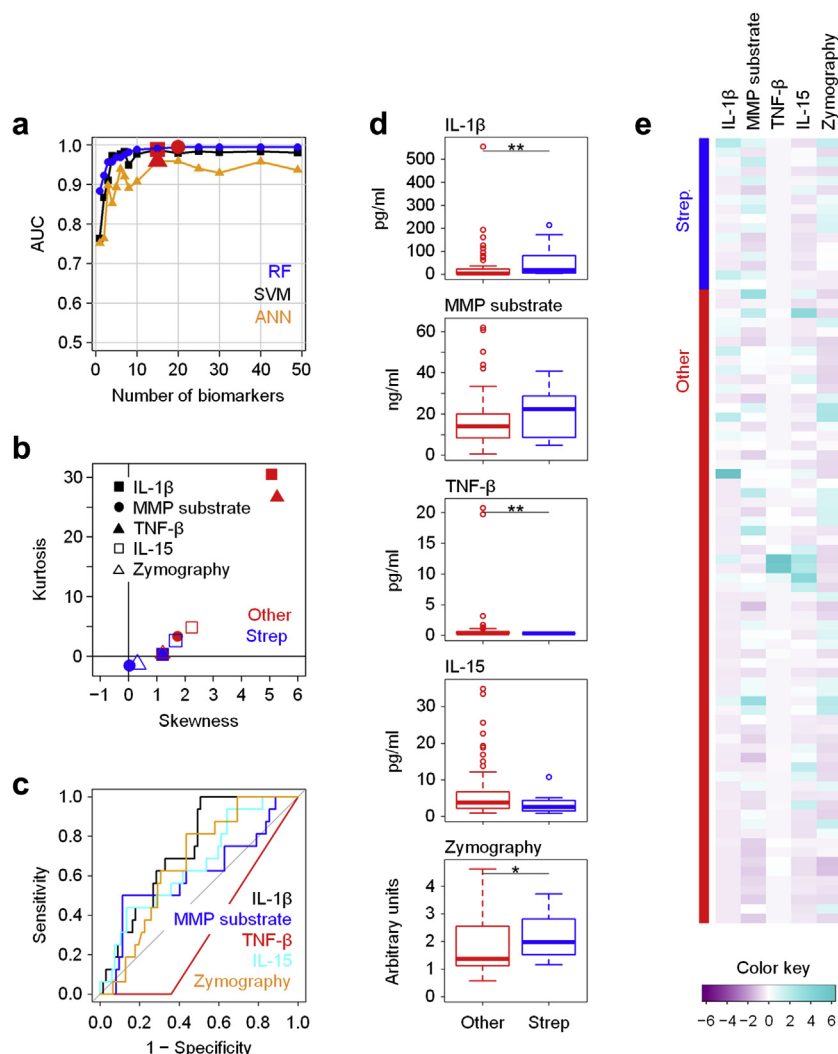
from all other episodes of peritonitis (Supplementary Figure S2, Supplementary Table S6A). Yet, despite a statistically significant difference between such nonstreptococcal Gram-positive infections and other episodes, especially for local levels of IL-17A, interferon- $\gamma$ , and IL-15, RF-based algorithms were not as powerful in this case as for the above predictions of Gram-negative or streptococcal infections, most likely due to the remaining heterogeneity of the organisms in that patient group (Supplementary Table S6B).

Given that CNS species such as *Staphylococcus epidermidis* are the major cause of peritonitis in PD patients and are also clinically associated with a relatively benign outcome,<sup>30</sup> we finally determined immune fingerprints that would specifically define CNS infections. Despite having only marginal or no statistical significance as individual biomarkers on conventional tests, the combination of IL-15, IL-16, and soluble

IL-6 receptor (sIL-6R) levels, total cell count, and MMP substrate turnover showed excellent performance in the RF model, with an AUC of 0.961 (Figure 5, Supplementary Tables S7A+B), demonstrating that CNS infections are sufficiently immunologically distinct to allow a pathogen-specific diagnosis in PD patients.

#### Local biomarkers on the day of presentation correlate with subsequent clinical outcomes over the following 90 days

The nature of the causative pathogen and the underlying inflammatory response profoundly affect clinical outcomes of peritonitis in PD patients,<sup>6,15</sup> indicating a need for prognostic biomarkers in early disease. When applying recursive feature elimination methods to our dataset, certain models accurately predicted the risk of downstream complications, defined as technique failure over the following 90 days, including

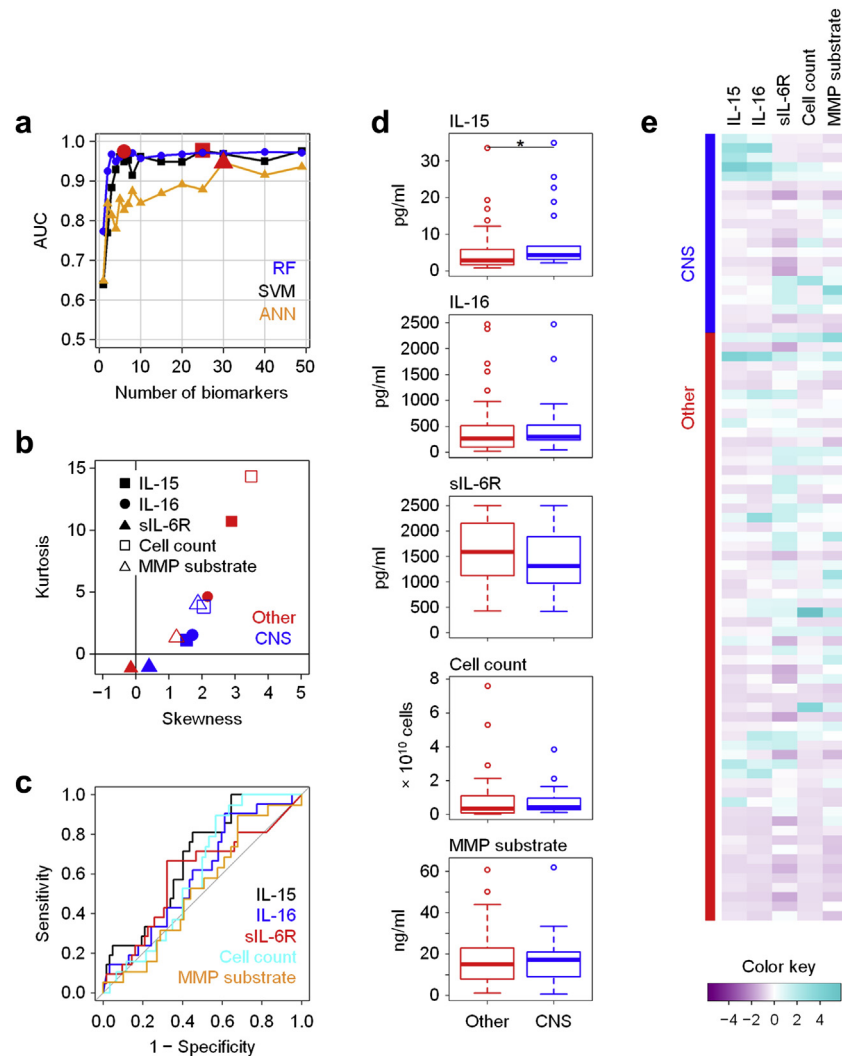


**Figure 4 | Local immune fingerprints in streptococcal (Strep) infections.** (a) Performance of Random Forest (RF), Support Vector Machine (SVM), and artificial neural network (ANN)-based feature elimination models for the prediction of infections caused by streptococcal species (*Streptococcus* spp. and *Enterococcus* spp.,  $N = 16$ ) against all other episodes of peritonitis ( $N = 67$ ), shown as area under the curve (AUC) depending on the number of biomarkers. One episode of peritonitis classified as streptococcal infection was a coinfection caused by *Enterococcus* sp. with light growth of coagulase-negative *Staphylococcus* spp. Red symbols depict the maximum AUC for each model. (b) Kurtosis and skewness of the top 5 biomarkers selected by RF-based feature elimination. (c) Receiver operating characteristic analysis showing specificity and sensitivity of the top 5 biomarkers. (d) Tukey plots of the top 5 biomarkers in patients with confirmed streptococcal infections and with all other episodes of peritonitis, as assessed by Mann-Whitney tests ( $*P < 0.05$ ;  $**P < 0.01$ ). (e) Heat map showing the top 5 biomarkers across all patients presenting with acute peritonitis. IL, interleukin; MMP, matrix metalloproteinase; TNF- $\beta$ , tumor necrosis factor- $\beta$ .

catheter removal, transfer to hemodialysis, and death (Figure 6). Patients experiencing technique failure after an episode of acute peritonitis had marginally higher levels of calprotectin, MMP-8, sIL-6R, and transforming growth factor- $\beta$  in their dialysis effluent as well as lower CD4 $^{+}$ :CD8 $^{+}$  T-cell ratios compared with uncomplicated cases. Although not being significantly different on their own, the combination of these 5 parameters in RF models yielded an AUC of 0.911 with excellent sensitivity (Supplementary Tables S8A and S8B), implying that the development of prognostic tests is feasible.

Taken together, our findings show that combinations of local biomarkers readily identify clinically meaningful

subgroups of peritonitis patients, depending on the culture results and subsequent clinical outcomes (Figure 7). With microbiologically distinct infections also displaying immunologically distinct immune responses, most individual parameters constituting meaningful fingerprints were only associated with single patient groups. However, certain biomarkers featured more prominently in these mathematical models than others, suggesting that they are particularly important factors in acute peritonitis, such as IL-1 $\beta$  and IL-15, each of which contributed to 3 different biomarker signatures, and the total cell count that was included in 4 different signatures and is thus likely to be of the utmost relevance in the diagnosis of PD patients.



**Figure 5 | Local immune fingerprints in coagulase-negative *Staphylococcus* (CNS) infections.** (a) Performance of Random Forest (RF), Support Vector Machine (SVM), and artificial neural network (ANN)-based feature elimination models for the prediction of infections caused by CNS (*Staphylococcus epidermidis* and related species;  $N = 21$ ) against all other episodes of peritonitis ( $N = 62$ ), shown as area under the curve (AUC) depending on the number of biomarkers. Red symbols depict the maximum AUC for each model. (b) Kurtosis and skewness of the top 5 biomarkers selected by RF-based feature elimination. (c) Receiver operating characteristic analysis showing specificity and sensitivity of the top 5 biomarkers. (d) Tukey plots of the top 5 biomarkers in patients with confirmed CNS infections and with all other episodes of peritonitis, as assessed by Mann-Whitney tests (\* $P < 0.05$ ). (e) Heat map showing the top 5 biomarkers across all patients presenting with acute peritonitis. IL, interleukin; MMP, matrix metalloproteinase; sIL-6R, soluble IL-6 receptor.

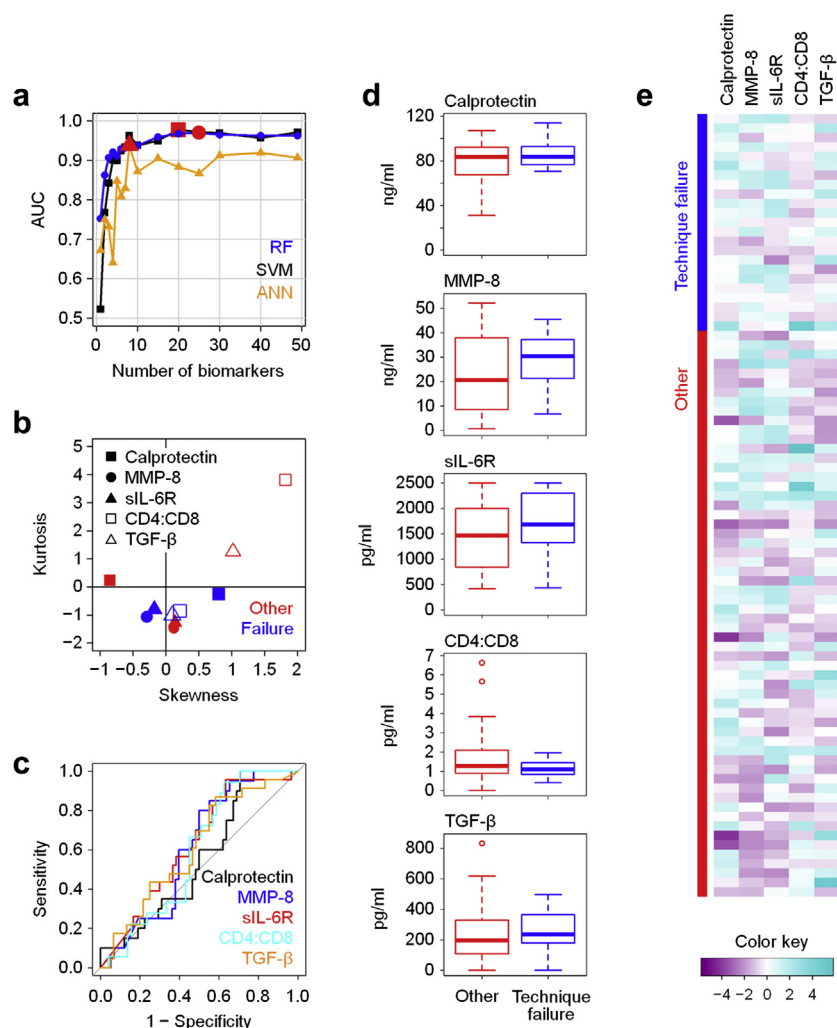
## DISCUSSION

This study demonstrates that different groups of bacteria induce qualitatively distinct local immune responses in infected patients. Specific biomarker signatures were associated with acute infections by Gram-negative and Gram-positive organisms, respectively, and with culture-negative episodes of peritonitis of unclear etiology. Taking advantage of the unique access to local inflammatory responses that is possible in PD patients via the peritoneal catheter and drained dialysate, we were also able for the first time to characterize pathogen-specific local immune responses to defined genera of organisms such as streptococcal species and coagulase-negative staphylococci. These findings demonstrate the feasibility of developing rapid pathogen-specific diagnostics

that exploit the exquisite responsiveness and specificity of the human immune system for different types of organisms. Such rapid methods might have greater utility than microbiological and molecular methods that aim at directly detecting such organisms but that are too slow, subject to confounding contaminants and often lack sufficient sensitivity to inform early patient management and target antibiotic therapy on the day of presentation with acute disease.<sup>31</sup>

Our study also demonstrates the power of using nonlinear approaches for mining complex biomedical datasets where conventional statistical methods fail to yield satisfactory results and where individual biomarkers on their own are unlikely to reach sufficient sensitivity and specificity to change clinical practice. Notably, the nature of the signatures



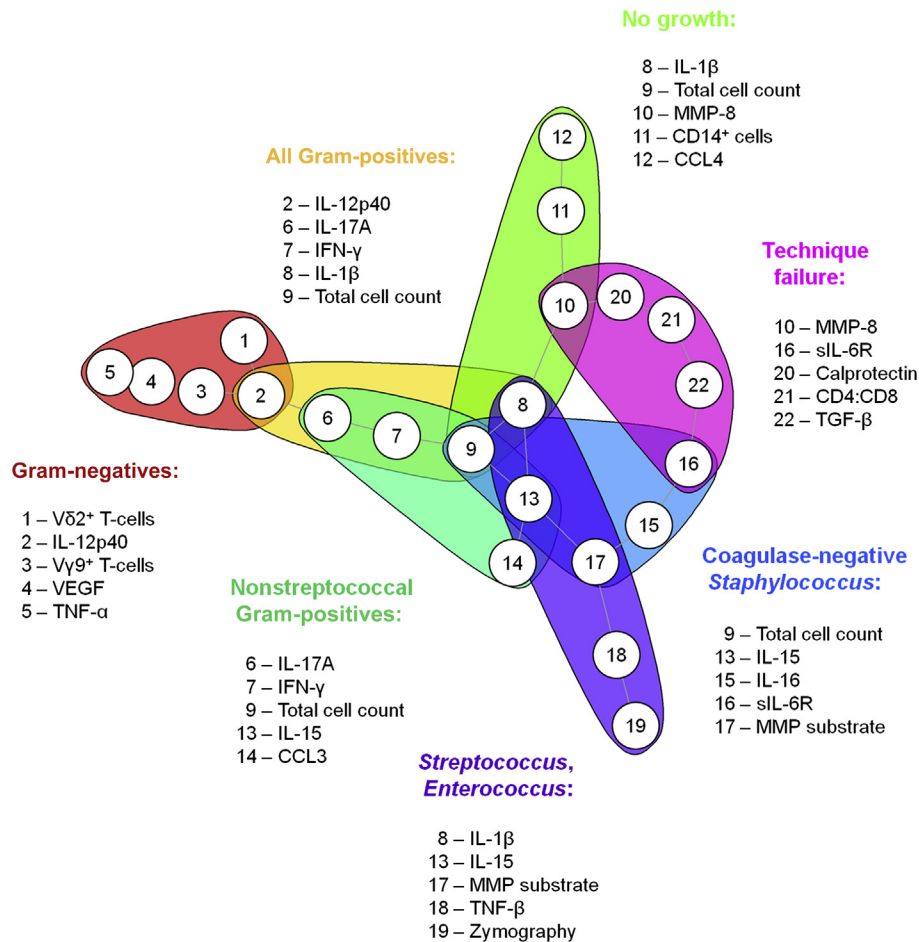


**Figure 6 | Local immune fingerprints associated with poor clinical outcomes.** (a) Performance of Random Forest (RF), Support Vector Machine (SVM), and artificial neural network (ANN)-based feature elimination models for the prediction of technique failure over the next 90 days (catheter removal, transfer to hemodialysis, or peritonitis-related death;  $N = 23$ ) against all other episodes of peritonitis ( $N = 60$ ), shown as area under the curve (AUC) depending on the number of biomarkers. Red symbols depict the maximum AUC for each model. (b) Kurtosis and skewness of the top 5 biomarkers selected by RF-based feature elimination. (c) Receiver operating characteristic analysis showing specificity and sensitivity of the top 5 biomarkers, as assessed by Mann-Whitney tests. (d) Heat map showing the top 5 biomarkers across all patients presenting with acute peritonitis. MMP, matrix metalloproteinase; sIL-6R, soluble IL-6 receptor; TGF- $\beta$ , transforming growth factor- $\beta$ .

identified in this study varied according to the mathematical model applied. By directly comparing 3 different approaches and assessing their performance when predicting microbiological and clinical endpoints in PD patients, we identified RFs as the most suitable models in this study and patient cohort, yielding superior performances and at fewer biomarkers than SVMs and ANNs, in accordance with investigations in other fields of research.<sup>32–34</sup>

Gram-negative bacteria, streptococci, and CNS are the major types of bacteria causing peritonitis in individuals receiving PD.<sup>11</sup> Strikingly, clinical outcomes of infections by those organisms differ, implying differences in their pathogenicity, their susceptibility to antibiotics, and/or the pathophysiology of the host responses they trigger.<sup>12,15</sup> Our findings demonstrate that in addition to their microbiological

differences, Gram-negative bacteria, streptococci, and CNS elicit fundamentally distinct immune responses, which is not only of relevance for the development of novel diagnostics but potentially also highlights key factors and cell types involved in sensing and fighting such infections. In this context, V $\gamma$ 9/V $\delta$ 2 T cells and TNF- $\alpha$  appear to be particularly relevant in Gram-negative infections, IL-1 $\beta$  and MMPs in streptococcal infections, and IL-16 and sIL-6R in CNS infections. IL-15 featured in prediction models for streptococcal, non-streptococcal, and CNS infection, indicating that this cytokine may play a role in all Gram-positive infections. In addition, the fact that IL-17A and interferon- $\gamma$  featured prominently in nonstreptococcal Gram-positive infections (which included CNS), but not in CNS infections themselves, may argue for a particular role of both cytokines in response to Gram-positive



**Figure 7 | Summary of disease-specific immune fingerprints in patients presenting with acute peritonitis.** Shown are the top 5 biomarkers associated with the type of causative organism as indicated or with the risk of technique failure over the next 90 days. IFN-γ, interferon-γ; IL, interleukin; MMP, matrix metalloproteinase; sIL-6R, soluble IL-6 receptor; TGF-β, transforming growth factor-β; TNF-α, tumor necrosis factor-α; VEGF, vascular endothelial growth factor.

bacteria other than streptococci and CNS (i.e., in infections by *Staphylococcus aureus* and/or coryneform bacteria). The organism-specific contributions of these soluble mediators during acute disease, the cell types that produce them, and the targets that respond to them can now be addressed in appropriate cellular assays, suitable animal models that mimic the situation in patients as closely as possible, and well-defined cohorts with bacterial peritonitis and other infections with access to the site of inflammation. This will ultimately further our understanding of antimicrobial immune responses and how to exploit such knowledge diagnostically and therapeutically. The roles of calprotectin, MMP-8, sIL-6R, and transforming growth factor-β as well as the balance between CD4<sup>+</sup> and CD8<sup>+</sup> T cells may deserve special attention with regard to their involvement in regulating pathologic processes in the peritoneal cavity, and their contribution to predicting clinical outcomes.

Taken together, we successfully applied different machine learning models to complex biomedical datasets and identified key pathways involved in pathogen-specific immune

responses at the site of infection. It is apparent that the nature of the signatures identified depends on both biological and analytical parameters. However, our current findings demonstrate that such methodologies have immediate diagnostic and prognostic implications at the point of care, by informing patient management and the choice of treatment before traditional culture results become available. Being based on a relatively small population in a single hospital, the biomarkers identified in this study and the corresponding algorithms now await external validation in larger patient cohorts at multiple sites<sup>35</sup> in order to demonstrate the applicability of the chosen approach to other centers where the spectrum of the infecting organisms and the previous infection history as well as patient demographics and health care settings may vary. Validated biomarker combinations can then be incorporated into appropriate diagnostic tests to be used in central laboratories or at the point of care and into new patient management and treatment guidelines based on such test results. The final choice of biomarkers to be taken forward will depend on the desired performance

requirements, with soluble proteins being equally suitable for automated immunodiagnostic analyzers and bedside or home tests, whilst assessments of immune cell subsets such as V $\gamma$ 9/V $\delta$ 2 T cells would require standardized flow cytometric protocols. In the meantime, our study reaffirms the importance of correctly interpreting simple parameters such as the total cell count (contributing to 4 different immune fingerprints) and the differential leukocyte count (reflected in the proportion of CD14<sup>+</sup> cells among total cells), which already convey vital information about the nature of the causative pathogen.

## MATERIALS AND METHODS

### Patient samples

This study was approved by the South East Wales Local Ethics Committee (04WSE04/27) and registered on the UK Clinical Research Network Study Portfolio under reference number #11838 “Patient Immune Responses to Infection in PD.” All individuals provided written informed consent. The local cohort comprised 83 adults PD patients admitted between 2008 and 2016 to the University Hospital of Wales, Cardiff, on day 1 of acute peritonitis. Twenty-four age- and sex-matched stable PD patients with no infection in the previous 3 months were included as controls. Subjects positive for HIV or hepatitis C virus were excluded.

Clinical diagnosis of acute peritonitis was based on the presence of abdominal pain and cloudy peritoneal effluent with >100 white blood cells per cubic millimeter. According to the microbiological analysis of the effluent from preinoculated blood culture bottles by the routine Microbiology Laboratory, Public Health Wales, peritonitis episodes were defined as culture-negative ( $N = 19$ , after incubation of up to 5 days) or as confirmed bacterial infections by Gram-positive ( $N = 47$ ) and Gram-negative organisms ( $N = 17$ ) (Table 1). Cases of fungal infection and mixed or unclear culture results were excluded from the study. Clinical outcomes were recorded by following patients for 90 days after presenting with peritonitis. Technique failure was defined as catheter removal, transfer to hemodialysis or death within 90 days, and occurred in 21.1% of culture-negative episodes, 23.4% of Gram-positive infections, and 47.1% of Gram-negative infections (Table 1). Samples from  $\geq 8$ -hour dwells with volumes of 1 to 2.5 L were collected for biomarker measurements and processed as previously described.<sup>6,13–15</sup>

### Cellular biomarkers

Cells from cloudy peritoneal effluents were acquired on an 8-color FACSCanto II (BD Biosciences, San Diego, CA) and analyzed with FlowJo 10.1 (TreeStar, Ashland, OR), using monoclonal antibodies against CD3 (SK7), CD4 (RPA-T4), CD8 (RPA-T8), CD15 (HI98 or HIM1), and TCR-V $\delta$ 2 from BD Biosciences; anti-TCR-V $\gamma$ 9 (Immu360) from Beckman Coulter (Brea, CA); and anti-CD14 (61D3) from eBioscience (San Diego, CA); together with appropriate isotype controls. Leukocyte populations were gated based on their appearance in side scatter and forward scatter area/height and exclusion of live/dead staining (Fixable Aqua; Invitrogen, Carlsbad, CA). Biomarkers determined were the total cell counts; the percentages of CD3<sup>+</sup> T cells, CD14<sup>+</sup> monocytes/macrophages, and CD15<sup>+</sup> neutrophils among total cells; the frequencies of CD4<sup>+</sup>, CD8<sup>+</sup>, V $\gamma$ 9<sup>+</sup>, and V $\delta$ 2<sup>+</sup> cells within the CD3<sup>+</sup> T cell gate; and the ratios of CD4<sup>+</sup> to CD8<sup>+</sup> T cells.

### Soluble biomarkers

Cell-free peritoneal effluents were analyzed on a SECTOR Imager 6000 (Meso Scale Discovery, Rockville, MD) using the V-PLEX Human Cytokine 30-Plex Kit to measure levels of IL-1 $\alpha$ , IL-1 $\beta$ , IL-2, IL-4, IL-5, IL-6, IL-7, IL-10, IL-12p40, IL-12p70, IL-13, IL-15, IL-16, IL-17A, interferon- $\gamma$ , TNF- $\alpha$ , TNF- $\beta$ , granulocyte macrophage colony-stimulating factor, and VEGF as well as the chemokines CCL2, CCL3, CCL4, CCL11, CCL13, CCL17, CCL22, CCL26, CXCL8, and CXCL10; ultrasensitive single-plex assays for sIL-6R and IL-18 (Meso Scale Discovery); and a customer-made single-plex assay for IL-22 using capture (MAB7822) and biotinylated detection antibodies (BAM7821) and recombinant human IL-22 from R&D Systems. Conventional enzyme-linked immunosorbent assay kits were used to measure transforming growth factor- $\beta$ , total MMP-8, total MMP-9, and surfactant protein D (R&D Systems); calprotectin (Hycult Biotech, Inc., Plymouth Meeting, PA); and CCL2 (BD Biosciences). Human neutrophil elastase was measured using a B.I.T.S. enzyme-linked immunosorbent assay kit (Mologic, Bedford, UK). Active human neutrophil elastase was measured using the fluorogenic substrate MeOSuc-Ala-Ala-Pro-Val-AMC (Bachem, Bubendorf, Switzerland); active MMP was measured using the fluorogenic substrate Mca-Lys-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH<sub>2</sub> (Enzo Life Sciences, Farmingdale, NY) and by zymography using precast Novex gelatin zymogram gels (Invitrogen) scanned on a Bio-Rad GS800 densitometer (Bio-Rad Laboratories, Berkeley, CA) and analyzed using ImageJ software. Total protein was measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA). Measurements that were below or above the detection limit were replaced by the lowest and highest detectable values for each biomarker, respectively.

### Data preprocessing

All analyses were performed using R software version 3.2.5 (R Foundation, Vienna, Austria). Before applying machine learning models, missing data imputation was applied to fit gaps due to missing or failed measurements by adopting Multivariate Imputation by Chained Equations,<sup>36</sup> which impute an incomplete feature by generating synthetic values taking into account their relationship with other biomarkers, using RF models (Supplementary Table S9). Data were then standardized to a mean of 0 and a variance of 1 to reduce the effect of large feature range variation. After preprocessing, the samples in the minority groups were unsampled so that minority and majority groups had equal frequencies.

### Feature elimination

The caret package in R<sup>37</sup> was adopted for the implementation of recursive feature elimination methods using 3 different machine learning models: SVMs with radial basis function kernel in the kernlab R package,<sup>38</sup> RFs with Breiman's algorithm in the randomForest R package,<sup>25</sup> and single-hidden-layer ANNs in the nnet R package.<sup>39</sup> To reduce variability, resampling methods were applied in the outer layer of the iteration, and cross-validation was used in the model fitting and parameter tuning to avoid overfitting. During parameter tuning, the search regions for hyperparameters in different classification models were set as follows: the penalty factor  $C$  that controls the tradeoff between learning errors and the complexity term and the radial basis function kernel parameter  $\sigma$  both ranged from  $2^{-5}$  to  $2^5$  with steps of 2; the number of trees in RFs were selected from [100, 300, 500, 1000, 3000], and the size of ANNs ranged from 2 to  $2^{10}$  hidden nodes with steps of 2 and a

weight decay from  $10^{-4}$  to  $10^{-10}$  with steps of 10. The number of repeats for resampling was set as 5 and for cross-validation and model selection as 10.

### Basic statistical analyses

Correlations between all 49 biomarkers were determined using the *corrplot* R package,<sup>40</sup> based on correlation calculations using the *Hmisc* R package.<sup>41</sup> Means, SEs, skewness, and kurtosis were determined with *plotrix* R<sup>42</sup> and *e1071* R packages.<sup>43</sup> Mann-Whitney tests (2-sample Wilcoxon tests) were applied to assess the relationship between two patient groups. Heat maps were visualized using the *gplots* R package.<sup>44</sup>

### DISCLOSURE

C-YL, NT, and ME are inventors on patent applications filed by University College Cardiff Consultants Ltd. (Cardiff University) in Europe and the United States on the identification of bacterial infections in peritoneal dialysis patients. PD is the cofounder and chief scientific officer of Mologic Ltd.; PD and GP are shareholders/option holders in Mologic. All the other authors declared no competing interests.

### ACKNOWLEDGMENTS

We are grateful to all patients and volunteers for participating in this study and to the clinicians and nurses for their cooperation. We especially thank Delyth Davies, Sally Jones, Billy Keogh, Chia-Te Liao, and Sharron Tatchell for help with patient recruitment and sampling and Eryl Francis for her support throughout this study. This research received support from the Wales Kidney Research Unit (WKRU), UK Clinical Research Network (UKCRN) Study Portfolio, NISCHR/MRC Health Research Partnership Award HA09-009, Kidney Research UK grant RP6/2014, MRC grant MR/N023145/1, NIHR i4i Product Development Award IL-LA-0712-20006, NISCHR/Wellcome Trust Institutional Strategic Support Fund (ISSF), MRC Confidence in Concept scheme, SARTRE/SEWAHSP Health Technology Challenge scheme, and EU-FP7 Initial Training Network 287813 "European Training & Research in Peritoneal Dialysis" (EuTriPD).

### SUPPLEMENTARY MATERIAL

**Figure S1.** Local immune fingerprints in Gram-positive infections. (A) Performance of Random Forest (RF), Support Vector Machine (SVM), and artificial neural network (ANN)-based feature elimination models for the prediction of Gram-positive infections ( $N = 47$ ) against all other episodes of peritonitis ( $N = 36$ ), shown as area under the curve (AUC) depending on the number of biomarkers. Red symbols depict the maximum AUC for each model. (B) Kurtosis and skewness of the top 5 biomarkers selected by RF-based feature elimination. (C) Receiver operating characteristic analysis showing specificity and sensitivity of the top 5 biomarkers. (D) Tukey plots of the top 5 biomarkers in patients with confirmed Gram-positive infections and with all other episodes of peritonitis, as assessed by Mann-Whitney tests ( $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ ). (E) Heat map showing the top 5 biomarkers across all patients presenting with acute peritonitis. IFN- $\gamma$ , interferon- $\gamma$ ; IL, interleukin.

**Figure S2.** Local immune fingerprints in nonstreptococcal Gram-positive infections. (A) Performance of Random Forest (RF), Support Vector Machine (SVM), and artificial neural network (ANN)-based feature elimination models, for the prediction of infections caused by nonstreptococcal Gram-positive species (*Staphylococcus aureus*, coagulase-negative *Staphylococcus* spp., *Corynebacterium* spp.;  $N = 31$ ) against all other episodes of peritonitis ( $N = 62$ ), shown as area under the curve (AUC) depending on the number of biomarkers.

One episode of peritonitis classified as nonstreptococcal (Non-strep) infection was a coinfection caused by *Corynebacterium* sp. and coagulase-negative *Staphylococcus* sp. Red symbols depict the maximum AUC for each model. (B) Kurtosis and skewness of the top 5 biomarkers selected by RF-based feature elimination. (C) Receiver operating characteristic analysis showing specificity and sensitivity of the top 5 biomarkers. (D) Tukey plots of the top 5 biomarkers in patients with confirmed nonstreptococcal Gram-positive infections and with all other episodes of peritonitis, as assessed by Mann-Whitney tests ( $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ ). (E) Heat map showing the top 5 biomarkers across all patients presenting with acute peritonitis.

**Table S1.** Local biomarkers in stable PD patients and in patients presenting with acute peritonitis.

**Table S2A.** Local biomarkers in patients presenting with acute peritonitis caused by Gram-negative organisms or with other episodes.

**Table S2B.** Performance of local biomarkers in predicting Gram-negative infections in PD patients against all other episodes of peritonitis.

**Table S3A.** Local biomarkers in patients presenting with culture-negative peritonitis or with other episodes.

**Table S3B.** Performance of local biomarkers in predicting culture-negative episodes in PD patients against all other microbiologically confirmed infections.

**Table S4A.** Local biomarkers in patients presenting with acute peritonitis caused by Gram-positive organisms or with other episodes.

**Table S4B.** Performance of local biomarkers in predicting Gram-positive infections in PD patients against all other episodes of peritonitis.

**Table S5A.** Local biomarkers in patients presenting with acute peritonitis caused by streptococcal species (*Streptococcus* spp. and *Enterococcus* spp.) or with other episodes.

**Table S5B.** Performance of local biomarkers in predicting infections caused by streptococcal species (*Streptococcus* spp. and *Enterococcus* spp.) against all other episodes of peritonitis.

**Table S6A.** Local biomarkers in patients presenting with acute peritonitis caused by nonstreptococcal Gram-positive species (*Staphylococcus aureus*, coagulase-negative *Staphylococcus* spp., *Corynebacterium* spp.) or with other episodes.

**Table S6B.** Performance of local biomarkers in predicting infections caused by nonstreptococcal Gram-positive species (*Staphylococcus aureus*, coagulase-negative *Staphylococcus* spp., *Corynebacterium* spp.) against all other episodes of peritonitis.

**Table S7A.** Local biomarkers in patients presenting with acute peritonitis caused by coagulase-negative staphylococci or with other episodes.

**Table S7B.** Performance of local biomarkers in predicting infections caused by coagulase-negative staphylococci against all other episodes of peritonitis.

**Table S8A.** Local biomarkers in patients technique failure over the next 90 days or with other episodes.

**Table S8B.** Performance of local biomarkers in predicting technique failure over the next 90 days against all other episodes of peritonitis.

**Table S9.** Proportion of missing values.

Supplementary material is linked to the online version of the paper at [www.kidney-international.org](http://www.kidney-international.org).

### REFERENCES

1. Janeway CA Jr. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb Symp Quant Biol.* 1989;54:1-13.
2. Matzinger P. The danger model: a renewed sense of self. *Science.* 2002;296:301-305.



3. Blander JM, Sander LE. Beyond pattern recognition: five immune checkpoints for scaling the microbial threat. *Nat Rev Immunol.* 2012;12: 215–225.
4. Liuzzi AR, McLaren JE, Price DA, Eberl M. Early innate responses to pathogens: pattern recognition by unconventional human T-cells. *Curr Opin Immunol.* 2015;36:31–37.
5. Ramilo O, Allman W, Chung W, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood.* 2007;109: 2066–2077.
6. Lin CY, Roberts GW, Kift-Morgan A, et al. Pathogen-specific local immune fingerprints diagnose bacterial infection in peritoneal dialysis patients. *J Am Soc Nephrol.* 2013;24:2002–2009.
7. Smith CL, Dickinson P, Forster T, et al. Identification of a human neonatal immune-metabolic network associated with bacterial infection. *Nat Commun.* 2014;5:4649.
8. Bajpai A, Prasad KN, Mishra P, et al. Distinct cytokine pattern in response to different bacterial pathogens in human brain abscess. *J Neuroimmunol.* 2014;273:96–102.
9. Mejias A, Suarez NM, Ramilo O. Detecting specific infections in children through host responses: a paradigm shift. *Curr Opin Infect Dis.* 2014;27: 228–235.
10. Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci Transl Med.* 2016;8:346ra91.
11. Cho Y, Johnson DW. Peritoneal dialysis-related peritonitis: towards improving evidence, practices, and outcomes. *Am J Kidney Dis.* 2014;64: 278–289.
12. Li PK, Szeto CC, Piraino B, et al. ISPD Peritonitis Recommendations: 2016 Update on Prevention and Treatment. *Perit Dial Int.* 2016;36: 481–508.
13. Eberl M, Roberts GW, Meuter S, et al. A rapid crosstalk of human  $\gamma\delta$  T cells and monocytes drives the acute inflammation in bacterial infections. *PLOS Pathog.* 2009;5:e1000308.
14. Davey MS, Lin CY, Roberts GW, et al. Human neutrophil clearance of bacterial pathogens triggers anti-microbial  $\gamma\delta$  T cell responses in early infection. *PLOS Pathog.* 2011;7:e1002040.
15. Liuzzi AR, Kift-Morgan A, Lopez-Anton M, et al. Unconventional human T cells accumulate at the site of infection in response to microbial ligands and induce local tissue remodeling. *J Immunol.* 2016;197: 2195–2207.
16. Szeto CC, Lai KB, Chow KM, et al. Dialysate bacterial endotoxin as a prognostic indicator of peritoneal dialysis related peritonitis. *Nephrology.* 2016;21:1069–1072.
17. Campbell DJ, Johnson DW, Mudge DW, et al. Prevention of peritoneal dialysis-related infections. *Nephrol Dial Transplant.* 2015;30: 1461–1472.
18. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res.* 2003;3:1157–1182.
19. Cortes C, Vapnik V. Support-vector networks. *Machine Learning.* 1995;20: 273–297.
20. Furey TS, Cristianini N, Duffy N, et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics.* 2000;16:906–914.
21. Orrù G, Pettersson-Yeo W, Marquand AF, et al. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev.* 2012;36: 1140–1152.
22. Bishop CM. *Neural Networks for Pattern Recognition.* New York, NY: Oxford University Press; 1995.
23. Khan J, Wei JS, Ringnér M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med.* 2001;7:673–679.
24. Amato F, López A, Peña-Méndez EM, Vañhara P, et al. Artificial neural networks in medical diagnosis. *J Appl Biomed.* 2013;11:47–58.
25. Liaw A, Wiener M. Classification and regression by randomForest. *R News.* 2002;2:18–22.
26. Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics.* 2006; 7:3.
27. Touw WG, Bayjanov JR, Overmars L, et al. Data mining in the life sciences with Random Forest: a walk in the park or lost in the jungle? *Brief Bioinform.* 2013;14:315–326.
28. Eberl M, Friberg IM, Liuzzi AR, et al. Pathogen-specific immune fingerprints during acute infection: the diagnostic potential of human  $\gamma\delta$  T-cells. *Front Immunol.* 2014;5:572.
29. Fahim M, Hawley CM, McDonald SP, et al. Culture-negative peritonitis in peritoneal dialysis patients in Australia: predictors, treatment, and outcomes in 435 cases. *Am J Kidney Dis.* 2010;55:690–697.
30. Fahim M, Hawley CM, McDonald SP, et al. Coagulase-negative staphylococcal peritonitis in Australian peritoneal dialysis patients: predictors, treatment and outcomes in 936 cases. *Nephrol Dial Transplant.* 2010;25:3386–3392.
31. Bieber SD, Anderson AE, Mehrotra R. Diagnostic testing for peritonitis in patients undergoing peritoneal dialysis. *Semin Dial.* 2014;27: 602–606.
32. Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics.* 2008;9:319.
33. Hsieh CH, Lu RH, Lee NH, et al. Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. *Surgery.* 2011;149: 87–93.
34. Liu M, Wang M, Wang J, Li D. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. *Sensors Actuators B: Chem.* 2013;177: 970–980.
35. Taylor JM, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. *Clin Cancer Res.* 2008;14:5977–5983.
36. Van Buuren S, Groothuis-Oudshoorn K. MICE: multivariate imputation by chained equations in R. *J Stat Softw.* 2011;45:1–67.
37. Kuhn M. Contributions from Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z and the R Core Team. 2014. caret: classification and regression training. R package version 6.0.24. Available at: <https://cran.r-project.org/package=caret>. Accessed February 16, 2014.
38. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab - An S4 package for kernel methods in R. *J Stat Softw.* 2004;11:1–20.
39. Venables WN, Ripley BD. *Modern Applied Statistics with S.* 4th ed. New York, NY: Springer; 2002.
40. Wei T, Simko V. corrplot: Visualization of a correlation matrix. R package version 0.77, 2016. Available at: <https://cran.r-project.org/package=corrplot>. Accessed April 21, 2016.
41. Harrell FE Jr, with contributions from Dupont C and many others. 2016. Hmisc: Harrell Miscellaneous. R package version 3.17-3. Available at: <https://cran.r-project.org/package=Hmisc>. Accessed April 3, 2016.
42. Lemon J. Plotrix: a package in the red light district of R. *R News.* 2006;6:8–12.
43. Meyer D, Dimitriadou E, Hornik K, et al. e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-7, 2015. Available at: <https://cran.r-project.org/package=e1071>. Accessed August 5, 2015.
44. Warnes GR, Bolker B, Bonebakker L, et al. gplots: various R programming tools for plotting data. R package version 3.0.1, 2016. Available at: <https://cran.r-project.org/package=gplots>. Accessed March 30, 2016.