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

see commentary on page 16

OPEN

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

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

[9]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

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.[1–4] 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.[5–10]

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.[11,12] 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.[13–15] 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.[6,15] 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.[11,12,16] Treatment of peritonitis therefore continues to be largely empirical, and early but untargeted treatment with broad-spectrum antibiotics and antifungals is recommended.[12,17]

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.[18] 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.[19–21] 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.[22–24] Random Forests (RFs) are ensemble methods constructed on multiple decision trees for classification and regression.[25–27] 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.[6] 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 $V\gamma9^+$ and $V\delta2^+$ T cells within peritoneal T cells featured prominently in each. These findings appear to concur with our previous data suggesting a key role for $V\gamma9/V\delta2$ T cells in Gram-negative infections and emphasize