Lab on a Chip

PAPER

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/c8lc00108a

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

Multivariate computational analysis of biosensor's data for improved CD64 quantification for sepsis diagnosis†

U. Hassan,*[abc] R. Zhu[d] and R. Bashir *[abce]

Received 29th January 2018,
Accepted 10th March 2018

DOI: 10.1039/c8lc00108a

rsc.li/loc

Sepsis, as a leading cause of death worldwide, relies on systemic inflammatory response syndrome (SIRS) criteria for its diagnosis. SIRS is highly non-specific as it relies on monitoring of patients' vitals for sepsis diagnosis, which are known to change with many confounding factors. Changes in leukocyte counts and CD64 expression levels are known specific biomarkers of pro-inflammatory host response at the onset of sepsis. Recently, we have developed a biosensor chip that can enumerate the leukocyte counts and quantify the neutrophil CD64 expression levels from a drop of blood. We were able to show improved sepsis diagnosis and prognosis in clinical studies by measuring these parameters during different times of the patients' stay in hospital. In this paper, we investigated the rate of cell capture with CD64 expression levels and used this in a multivariate computational model using artificial neural networks (ANNs) and showed improved accuracy of quantifying CD64 expression levels from the biosensor (n = 106 whole blood experiments). We found a high coefficient of determination and low error between biosensor- and flow cytometry-based neutrophil CD64 expression levels using multiple ANN training methods in comparison to those of univariate regression commonly employed. This approach can find many applications in biosensor data analytics by utilizing multiple features of the biosensor's data for output determination.

Introduction

Sepsis is a leading cause of mortality worldwide and is a complicated immunological condition in response to pathogenic infections.[1,2] In the United States, every year more than one million people are diagnosed with severe sepsis and one third or more of them die.[2–4] The main contributing factor to these statistics is the unavailability of an accurate and sensitive diagnostic test for sepsis.[4] Systemic inflammatory response syndrome (SIRS) criteria are currently used as a diagnostic standard in hospitals and rely on monitoring of patients' vitals (temperature, respiratory rate, heart rate) and white blood cell counts.[4] These parameters are highly non-specific and may change because of many confounding factors. Increased CD64 expression on neutrophils (nCD64) is known to be a specific biomarker for detecting an early inflammatory response in patients and has shown improved sensitivity and specificity to sepsis diagnosis.[5–10] Early diagnosis of sepsis is extremely critical as every hour the appropriate antimicrobial treatment is delayed, the patients' survival rate decreases by 7.7%.[11]

In clinical laboratories, blood cell counting is commonly conducted using a hematology analyzer and specific leukocyte enumeration is performed using flow cytometers. Furthermore, quantification of cell surface receptor expression requires additional expertise in data analysis and is also commonly carried out using flow cytometers. Operating these instruments requires manual sample handling and significant training of laboratory personnel to achieve reproducible protocols. Point-of-care (POC) microfluidic sensors provide attractive alternatives and significant work has been performed to develop POC cell enumerating sensors. Microfluidic optical detection used waveguides and photodetectors to enumerate cells by looking at their scatter plots based on their size and granularity, while specific cells were counted by their fluorescence labelling and subsequent optical detection and imaging.[12–15] On the other hand, electrical cell counting requires miniaturized Coulter counters aligned with microfluidic channels.[16–19] Cells have been investigated at multiple frequencies to differentiate leukocytes based on their size and membrane capacitance.[16,19] Furthermore, specific

[a] Department of Bioengineering, University of Illinois at Urbana-Champaign, 1270 Digital Computer Laboratory, 1304 W. Springfield Ave, Urbana, IL 61801, USA. E-mail: uhassan2@illinois.edu, rbashir@illinois.edu
[b] Micro and Nanotechnology Lab, University of Illinois at Urbana-Champaign, 208 N. Wright St., Urbana, IL 61801, USA
[c] Stevens Family Biomedical Research Center, Carle Foundation Hospital, Urbana, IL 61801, USA
[d] Department of Statistics, University of Illinois at Urbana Champaign, Illini Hall, 725S Wright St. 101, 61820, Champaign, IL, USA
[e] Carle Illinois College of Medicine, Urbana, IL 61801, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8lc00108a

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

leukocytes can also be enumerated by their specific tagging using labeled microparticles to increase the corresponding impedance signature of the desired cells.[18] Furthermore, antigen expression-based cell capture in microfluidic devices has been reported by Murthy *et al.* by capturing high-expressing CD31 cells[20] and Pappas *et al.* by capturing Ramos B and HuT 78 T lymphocytes based on CD71 expression levels for their respective cell lines.[21] However, these devices haven't quantified expression levels from whole blood without manual processing or sample spiking. Mathematical modeling of expression-based cell capture has also been previously reported.[22–24]

Our group has developed POC microfluidic biosensors based on "differential immunoaffinity capture technology"

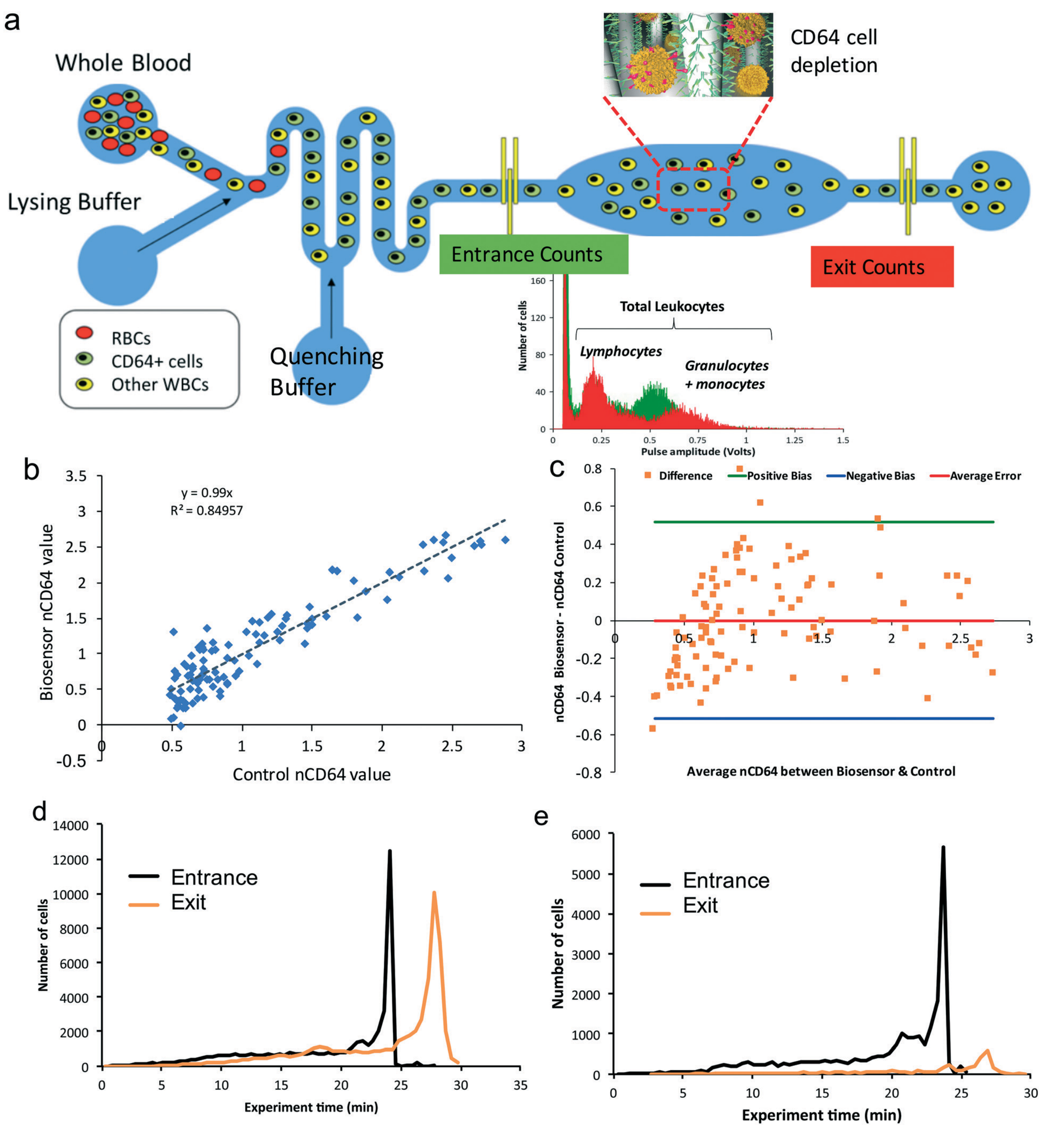

Fig. 1 (a) Sepsis biosensor schematic for enumerating leukocytes and quantifying CD64 expression levels from 10 μL of whole blood (adapted from ref. 30). Cells get captured based on the nCD64 expression level.[30] Pulse amplitude histogram obtained between entrance and exit counters showing granulocyte and monocyte expressing CD64 cell capture,[30] (b) comparison between nCD64 values obtained from the biochip and flow cytometry control (univariate regression with R^2 = 0.85),[30] (c) Bland–Altman analysis of data in b, (d) numbers of leukocytes counted at entrance and exit electrical counters are shown over time using a BSA blocked chamber. (e) Numbers of leukocytes counted at entrance and exit electrical counters are shown over time using an anti-CD64 antibody adsorbed chamber.

 This journal is © The Royal Society of Chemistry 2018

View Article Online

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

and enumerated complete blood cell counts (total leukocyte counts and their differentials, erythrocyte counts and platelets), specific leukocyte counts (CD4+ and CD8+ lymphocytes for HIV/AIDS diagnostics) and antigen expression quantification (CD64 expression levels for sepsis diagnosis) from a drop of whole blood in around 30 min.[25–30] Leukocyte enumeration is based on Coulter counting and impedance analysis of cells.[25] For specific leukocyte enumeration, we differentially count cells before and after the desired cells are captured in an immunoaffinity capture chamber.[28,29] Recently, we developed a 'sepsis biosensor' capable of enumerating leukocytes and nCD64 expression levels from patients' blood samples.[30] CD64 expressing cells were captured in a chamber based on their expression levels and the differential counts were directly correlated with the nCD64 expression levels.[30] The CD64 value from the biosensor is based on univariate (differential capture of CD64 expressing cells) regression.[30]

In this article, we investigated the rate of cell capture in comparison to CD64 expression values and used this as an additional parameter to determine biosensors' nCD64 expression values by employing multivariate regression using artificial neural networks. We used different training methods including Levenberg Marquardt, Bayesian regularization, and scaled conjugate gradient. Multivariate regression using artificial neural networks has shown improved accuracy, correlation coefficients and mean squared error in comparison to univariate and standard multivariate linear regression. We conducted this computational analysis on 106 whole blood experiments performed on blood samples collected from SIRS positive or infection suspected patients at Carle Foundation Hospital, Urbana. Experimental data are used from our recent study.[30]

Experimental setup

Our biosensor is based on differential immunocapture technology by performing differential cell counts (electrical) and affinity-based cell capture. Fig. 1a shows the overall schematic of our sepsis biosensor.[30] Whole blood (10 μL) is infused into the sensor along with lysis buffer to effectively lyse erythrocytes for a total lysis time of 6 s.[30] A quenching buffer is then infused into the biosensor to preserve the remaining leukocytes and maintain the osmolarity of the solution. Leukocytes pass through the microfluidic channel to which the microfabricated metal electrodes are aligned (fabrication procedures are given in our earlier papers).[28–30] As a single leukocyte passes through the electrical counter, it generates an electrical voltage signal (a bipolar pulse). The peak amplitude of the pulse is directly proportional to the cell size. This results in lymphocytes (6–9 μm in diameter) producing smaller pulses as compared to granulocytes and monocytes (10–15 μm in diameter). Leukocytes then flow through the immunoaffinity capture chamber to which anti-CD64 (clone 10.1) monoclonal antibody is adsorbed. CD64 leukocytes become captured based on their expression level. The remaining leukocytes are counted again and the difference of counts is directly proportional to nCD64 expression levels.[30]

Results

CD64 expression-based cell capture

The pulse peak amplitude histogram of both entrance (green) and exit counters (red) is shown in Fig. 1a.[30] Lymphocytes (6–9 μm) can be easily distinguished from granulocytes and monocytes in amplitude histograms owing to their smaller size. Furthermore, lymphocytes weren't captured in the chamber as they didn't express the CD64 antigen on their surface. However, granulocytes and monocytes were captured in the chamber based on their CD64 antigen expression levels.[30] Earlier, we have performed experiments on whole blood samples from our biosensor and correlated the cell capture percentage and CD64 expression levels.[30] Fig. S1† shows a linear correlation between the percentage of cell capture and CD64 expression levels (n = 106) obtained from a flow cytometer.[30] Subsequently, we performed univariate regression and 3-fold cross-validation with 1000 times random sample selection on the granulocyte + monocyte capture percentage data *vs.* CD64 expression values.[30] Fig. 1b shows the corresponding linear correlation plot between biosensor's nCD64 expression values and control nCD64 values obtained from the flow cytometer.[30] It shows a correlation coefficient of 0.85. We also performed Bland–Altman analysis to compare the nCD64 expression values obtained from the biosensor and control; it shows an average difference of zero between the two methods but a high value of 0.52 as limits of agreement (shown in Fig. 1c).[30]

Rate of cell capture over experiment duration

Here, we investigated the rate at which cells become captured in the capture chamber and its correlation with the CD64 expression values. To investigate this, we monitored the number of cells at the entrance and exit counters over the entire experiment duration. The number of leukocytes counted at the entrance and exit counters over time is shown in Fig. 1d. The cell count spike in the plots represents the release of cells bound to the walls of the microfluidic biosensor during the course of experiment. First, we investigated the capture using a BSA blocked chamber (Fig. 1d), which shows a minimal cell capture of less than 5%. The corresponding plots from individual granulocytes + monocytes and lymphocyte populations are shown in Fig. S2a and b,† respectively. We used the BSA blocked chamber for the initial optimization of the BSA concentrations in the chamber to prevent any non-specific cell adsorption. Similarly, using an anti-CD64 antibody adsorbed chamber, the plot of leukocytes counted at the entrance and exit electrical counters over time is shown in Fig. 1e. It shows much higher capture owing to the specific CD64 expression level-based capture. The corresponding plots from individual granulocyte + monocyte and lymphocyte populations are shown in Fig. S3a and b,† respectively.

Machine learning (hierarchical clustering) on biosensor's data (heat maps)

We used hierarchical clustering[31] machine learning techniques to investigate the effect of the rate of CD64 expression

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

levels on the rate of cell capture in the chamber. The normalized ratio of exit to entrance granulocyte + monocyte counts *versus* experiment time is shown in Fig. 2a for all 106 whole blood experiments. Each experiment shows a unique progression and the slope of the normalized ratio at different points in time. We selected a time window of 15–20 min and divided that interval into 30 s sub-intervals. Fig. S4† shows the correlation plots between slope values *vs.* cell counts (R^2 = 0.034) and slope values *vs.* cell capture percentage (R^2 = 0.4). To observe the underlying pattern between all the biosensor data features, the slope values, nCD64 and total granulocyte + monocyte counts are used as input data to the unsupervised hierarchical clustering methods. The detailed programs are described in the Materials and methods section. The corresponding heat map is shown in Fig. 2b. The heat map shows the underlying pattern between features, and it is interesting to find that the slope values across different time points share consistent patterns, while they are jointly inversely related to nCD64 expression values and cell counts. Similarly, when the pairwise distance is calculated with a 'correlation' method for hierarchical clustering of the input data, the resulting heat map is shown in Fig. S5.† This further validates the pattern between features, in particular, the inverse relationship of the slope values to the CD64 expression values. Therefore, the slope values can be used as promising predictors in more accurate determination of nCD64 expression levels from the biosensor's data.

Clustering on biosensor's data using artificial neural networks (self-organizing maps)

In order to further validate the inverse relationship of slope values to nCD64 expression values, we used the self-organizing map (SOM) clustering technique which is an unsupervised version of the artificial neural network (ANN).[32] A single layer model with 100 neurons is used as shown in Fig. S6.† The input features to the model are slope values, nCD64 values and granulocyte + monocyte counts. The network was trained using a batch SOM algorithm. Fig. S7† shows the SOM clustering analysis of the input feature set from the 106 whole blood nCD64 cell capture experiments. The plot of SOM neighbor weight distances shows neurons as gray-blue patches. The neighboring hexagons are colored black–yellow, which represents the closeness of each neuron's weight vector to its corresponding neighbors. Fig. S7† also shows the SOM layer showing sample hits. Each neuron represented by a hexagon shows numbers which represent the number of inputs it classifies. The colored patch size represents the number of vectors for each neuron. SOM input weight planes are shown in Fig. S8.† Each plot represents the specific input to the neuron layers. Zero connections are represented as black, and strong positive connections are represented as red. Input 1 is nCD64 values with low black connections showing their dependency on the slope values.

Multivariate regression on biosensor's data

We performed standard multivariate regression using Matlab on the biosensor's input feature data to predict nCD64 expression values for each blood sample. We used nCD64 values as a response variable and cell counts, percentage of captured cells and slope values as predictor variables in the function. The correlation plot between the biosensor and control nCD64 values is shown in Fig. 3a, which shows a correlation coefficient of R^2 = 0.77 and a recovery of 95%. The corresponding Bland–Altman analysis of the data is shown in Fig. 3b, which shows an average difference of −0.0003 and limits of agreement to be (−0.5403, 0.5395). The regular multivariate linear regression to predict nCD64 values from biosensor's input features didn't show any improvement and resulted in lower correlation coefficients. However, the multivariate linear regression may suffer from several drawbacks, such as highly correlated variables and a nonlinear relationship between the predictor and the response. Hence, we used

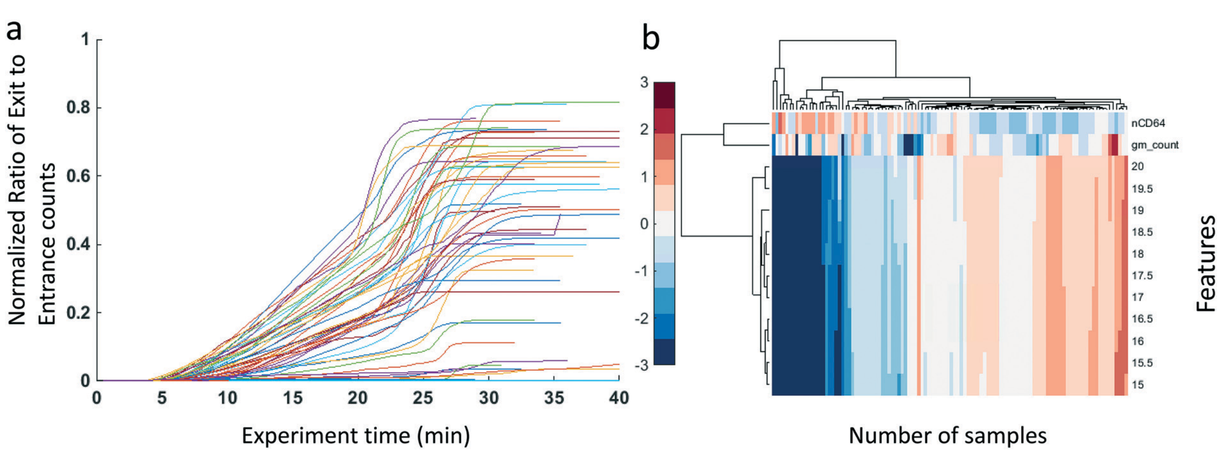

Fig. 2 (a) Normalized ratio of exit to entrance granulocyte + monocyte counts *vs.* experiment time showing different slope values. (b) Heat maps showing the underlying pattern between features, in particular the inverse relationship of the slope values across different time points to nCD64 expression levels and granulocyte + monocyte counts. The heat map is obtained through hierarchical clustering with Euclidean distance using features from 106 blood sample experiments. The *x*-axis represents each individual experiment corresponding to a unique blood sample.

This journal is © The Royal Society of Chemistry 2018

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

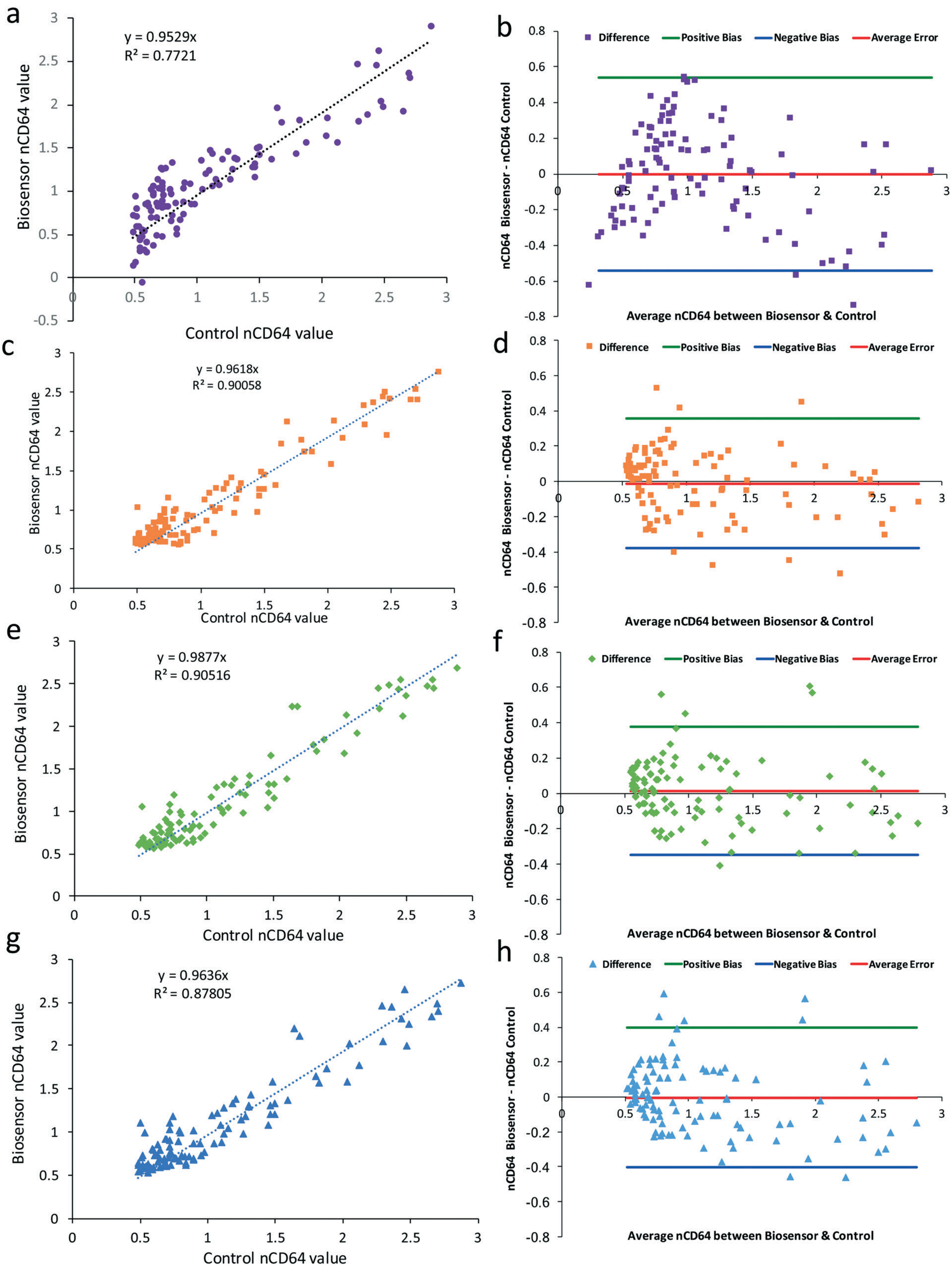

Fig. 3 Correlation between nCD64 values obtained from the biochip and flow cytometry control using multivariate regression analysis. (a) Standard multiple linear regression showing R^2 = 0.90, (b) Bland–Altman analysis of data in a, (c) multivariate regression using the Levenberg Marquardt ANN training method with R^2 = 0.90, (d) Bland–Altman analysis of data in c, (e) multivariate regression using the Bayesian regularization ANN training method with R^2 = 0.91, (f) Bland–Altman analysis of data in e, (g) multivariate regression using the scaled conjugate ANN training method with R^2 = 0.88, (h) Bland–Altman analysis of data in g. The control nCD64 values (from the flow cytometer) shown in Fig. 3 are obtained from ref. 30.

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

instead an artificial neural network to achieve higher predictability.

Multivariate regression on biosensor's data using artificial neural networks

Next, we used artificial neural networks,[33] performed multivariate regression on the biosensor's input feature data (slope values, percentage of captured granulocytes + monocytes and corresponding cell counts) and determined the biosensor's nCD64 expression values for each blood sample. We used a two-layer feed-forward artificial neural network with 10 neurons for hidden layers and 1 neuron for output layers, respectively, as shown in Fig. S9.† We trained the model with different methods in Matlab including Levenberg Marquardt, Bayesian regularization, and scaled conjugate. The resulting comparison plots between flow cytometry (control) nCD64 values and biosensor's nCD64 values for all the training methods are shown in Fig. 3. First, when the model is trained with the Levenberg Marquardt method, the corresponding Hinton diagram representing the weight and bias values of all neurons in hidden and output layers is shown in Fig. S10.† Furthermore, the output regression plots of training, validation and testing data sets are shown in Fig. S11.† The error histogram and mean squared error (0.0445) plots are shown in Fig. S12a and b,† respectively. The correlation plot between the biosensor and control nCD64 values using the Levenberg Marquardt method is shown in Fig. 3c, which shows a correlation coefficient of $R^2 = 0.90$ and a recovery of 96%. The corresponding Bland–Altman analysis of the data is shown in Fig. 3d, which shows an average difference of −0.0116 and limits of agreement to be (−0.3793, 0.3561). Second, when the model is trained with the Bayesian regularization method, the corresponding Hinton diagram representing weight and bias values of all neurons in the network is shown in Fig. S13.† The output regression plots of data sets are shown in Fig. S14† and the corresponding error histogram and mean squared error (0.0369) plots are shown in Fig. S15a and b,† respectively. Similarly, the correlation plot between the biosensor and control nCD64 values using the Bayesian regularization training method is shown in Fig. 3e, which shows a correlation coefficient of $R^2 = 0.91$ and a recovery of 99%. The corresponding Bland–Altman analysis of the data (Fig. 3e) is shown in Fig. 3f, which gives an average difference of −0.0127 and limits of agreement to be (−0.3499, 0.3753). Third, when the model is trained with the scaled conjugate method, the corresponding Hinton diagram is shown in Fig. S16.† The output regression plots of data sets are shown in Fig. S17† and the corresponding error histogram and mean squared error (0.0669) plots are shown in Fig. S18a and b,† respectively. Finally, the correlation plot between the biosensor and control nCD64 values using the scaled conjugate training method is shown in Fig. 3g, which shows a correlation coefficient of $R^2 = 0.88$ and a recovery of 96%. The corresponding Bland–Altman analysis of the data (Fig. 3g) is shown in Fig. 3h, which gives an average difference of −0.0021 and limits of agreement to be (−0.4039, 0.3976). Furthermore, Tables S1–S3† list all the weight and bias values of all the neurons in the input and output layers of the network when trained with Levenberg Marquardt, Bayesian regularization, and scaled conjugate training methods, respectively.

Comparison between computational methods

We also calculated the accuracy to determine nCD64 expression values from the biosensor as compared to control values for all the computational methods used. Fig. 4 shows the accuracy plot to predict nCD64 bins (equally divided over the entire range) using the univariate regression, multivariate linear regression and ANN training methods. Univariate regression accuracy results are obtained from earlier analysis. In order to predict the nCD64 bins with accuracy greater than 85%, the minimum size of an nCD64 bin is found to be 1.2 for univariate regression. For multivariate linear regression, the minimum size of the nCD64 bin is found to be 1.33, which is higher compared to that for univariate regression, thus showing no improvement. However artificial neural network training methods worked better and showed much better performance. The minimum bin size with prediction accuracy greater than 85% is found to be 0.632 for both Levenberg Marquardt and Bayesian regularization methods and 0.706 for the scaled conjugate method. This shows that ANN methods perform better than univariate regression based on the calculated nCD64 values.

Furthermore, Table 1 shows a more detailed comparative analysis between univariate regression, multivariate regression, and different artificial neural network training methods including Levenberg Marquardt, Bayesian regularization and scaled conjugate. In terms of highest accuracy (89.5%), coefficient of determination ($R^2 = 0.91$), lowest mean squared error (0.0369), and Bland–Altman's lower limits of agreement (−0.3499, 0.3753), Bayesian regularization not only performed better than univariate and standard multivariate linear

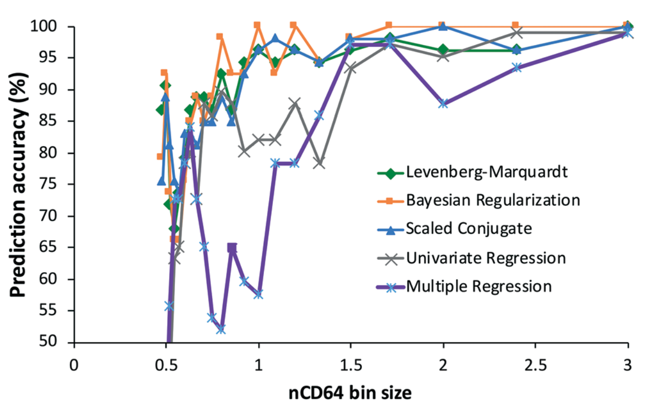

Fig. 4 Accuracy to predict nCD64 bins using univariate regression, multivariate linear regression and multivariate ANN different training methods. ANN methods perform better than univariate and standard multivariate linear regression methods based on the calculated nCD64 values. Univariate regression accuracy results are obtained from earlier analysis.[30]

This journal is © The Royal Society of Chemistry 2018

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

Table 1 Comparison between different computational methods including univariate regression, standard multivariate regression and multivariate ANN training methods (Levenberg Marquardt, Bayesian regularization and scaled conjugate methods). Bayesian regularization method performs best with the lowest mean squared error, the highest R^2 value, nCD64 recovery, and prediction accuracy

		Artificial neural network results			Linear regression and Bland–Altman analysis						
	Computational method	*R* value	Mean squared error	Epochs	R^2 value	nCD64 recovery	Bland–Altman (mean error)	Bland–Altman (pos. bias)	Bland–Altman (neg. bias)	nCD64 bin size (accuracy ≥ 85%)	Overall accuracy
Multivariate ANN	Levenberg Marquardt	0.95	0.0445	4	0.90	96%	−0.0116	0.3561	−0.3793	0.632	89.2%
Multivariate ANN	Bayesian regularization	0.95	**0.0369**	68	**0.91**	**99%**	0.0127	0.3753	−0.3499	0.632	**89.5%**
Multivariate ANN	Scaled conjugate	0.94	0.0669	29	0.88	96%	−0.0031	0.3976	−0.4039	0.706	88.8%
	Multivariate regression				0.77	95%	−0.0003	0.5395	−0.5403	1.333	71.7%
	Univariate regression				0.85	99%	0.000	0.5171	−0.5171	1.200	78.9%

- Overall accuracy: average accuracy of all nCD64 bin sizes

regression but also outperformed other ANN multivariate Levenberg Marquardt, and scaled conjugate training methods.

Discussion and conclusions

Here, we show an increase in the R^2 value (0.85 to 0.91) and accuracy (78.9% to 89.5%) to quantify CD64 values from the biosensor using ANNs for multi-feature regression as compared to single feature regression analysis. Similarly, the use of ANN methods for multivariate regression (R^2 = 0.91) has shown much better performance than the standard multivariate linear regression (R^2 = 0.77). Furthermore, for univariate and multivariate linear regression, the CD64 bin prediction accuracy starts to drop as the nCD64 bin size decreases from 1.5; however, the accuracy remains greater than 90% for the ANN methods used for CD64 calculation. In prospective clinical adjudication based on patients' CD64 levels, CD64 bins can be assigned to different patient categories depending on their state of sepsis progression.[34] Thus, the ability to accurately predict more bins of smaller size will help in developing more accurate diagnostic and prognostic disease models.

The selection of the ANN multivariate regression training method depends on time and space complexity *i.e.* the processing time and computing memory the data training requires. Considering these two parameters, the Levenberg Marquardt method takes the least time and the scaled conjugate gradient the least memory to train the input data for regression. Meanwhile, Bayesian regularization requires more time and memory than the other two methods. However, biosensors relatively produce smaller data sets as compared to other genomics, proteomics and other high bandwidth data sets *e.g.* climate and astronomical data sets. Therefore, multivariate regression using training algorithms requiring more memory and time, like Bayesian regularization, can easily be implemented in a reasonable time on the biosensor's data owing to its smaller datasets. Bayesian regularization also performed best in terms of highest accuracy (89.5%) and coefficient of determination (R^2 = 0.91), lowest mean squared error (0.0369), and Bland–Altman's lower limits of agreement (−0.3499, 0.3753) as shown in Table 1.

The accuracy and coefficient of determination can be further improved by reducing the non-specific capture of cells in the chamber which is currently 2% of the total leukocytes and 8% of granulocytes + monocytes. Lowering the flow rate can decrease the normalized experimental slopes (shown in Fig. 2) as the experiment takes more time to complete. Changing the flow rate should be accompanied by the change of the geometry of the erythrocyte lysing module to keep the lysing and quenching times the same. Furthermore, flow rate changes can result in different shear stresses in the capture chamber for the immunoaffinity capture, and the chamber geometry also needs to be changed to keep similar shear conditions for cellular capture in the chamber.

The underlying inverse pattern of slope values to the CD64 expression levels in Fig. 2b shows some 'singularities' in the heat map when the linear inverse relationship does not apply. There are two possibilities for those 'singularities' or outliers. First, it could be a random error in our features or their measurements using the biosensor, which could potentially come from experimental variability. Second, we had a very generalized patient inclusion criterion, *i.e.* SIRS positive or/and suspicion of an infection.[30] Our patient population had many confounding factors including acute and chronic co-morbidities and other systemic inflammatory diseases that may influence the level of nCD64 expression and other features including cell counts.[30]

Although our biosensor is designed for determining nCD64 expression levels, it can easily be adapted to quantify any other cell surface receptor expression levels specific to sepsis. For example, reduced HLA-DR expression on monocytes and increased CD11b expression on neutrophils are also known to be specific biomarkers of pro-inflammatory and anti-inflammatory responses of the immune system in septic patients.[35–38] The corresponding changes in the biosensor

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

are required to immobilize their respective antibodies in the capture chamber. Similarly, ANN-based multivariate regression can also be applied to such biomarker determination considering the same parameters of normalized rates of capture and percentage of captured cells.

Furthermore, our biosensor can also be adapted to capture CYP1A1 and CYP1B1 in endothelial cells and our computational analysis can be used to quantify their expression levels to explore their pro/anti-atherogenic roles.[39] CD71 expression changes in circulating tumor cells (CTCs) based on growth regulation and proliferation can also be quantified by their expression-based capture;[40] however, owing to the very low concentration of CTCs in whole blood, the cellular capture and differential counts also depend on non-specific cellular adhesion to the chamber which should be incorporated into the model.

In our whole blood experiments, we have many quality control processes, to keep the variability minimum to collect the best datasets. The capture chambers and lysing modules were discarded after each experiment. Capture chambers are also observed under a microscope to check for any pillars collapsing and damaged areas before use in an experiment. Each electrical counter is also tested before use in the experiments. The resistance of the metal lines on PCB and electrodes is measured and the counter is only used if the resistance is less than 5 ohms.

In conclusion, we have shown the improved accuracy for CD64 expression level quantification by employing multivariate regression using artificial neural networks as compared to commonly used univariate regression. We developed an ANN model and trained the biosensor's input data to quantify CD64 expression by using multivariate training methods including Levenberg Marquardt, Bayesian regularization, and scaled conjugate. ANN multivariate regression has shown improved accuracy, correlation coefficients and mean squared error in determining CD64 values in comparison to univariate regression. We conducted this computational analysis on 106 whole blood experiments performed on blood samples collected from SIRS positive or infection suspected patients at Carle Foundation Hospital, Urbana. This approach finds many applications in biosensor data analytics by utilizing multiple features of the biosensor's data for output determination.

Materials and methods

Statistical analysis

Univariate regression: univariate regression is conducted considering the percentage of captured cells (as a single variable) to determine nCD64 expression values. Data on the scatter plot between variables are fitted for regression using Microsoft Excel. Rotation estimation (cross validation): the cross validation technique is used to determine biosensors' nCD64 values from the percentage of captured cells. We employed 3-fold validation by dividing the entire data into 3 sets and performing regression on 2 sets and validating on the third set (blinded). We used 1000 rotations to reduce error, randomly selected the samples and performed cross-validation. The results are averaged for all the rotations. Bland–Altman (BA) analysis: BA analysis compares the two methods (biosensor and flow cytometer) to quantify nCD64 values. It shows the difference (biosensor nCD64–control nCD64) of values *vs.* average of both methods. The bias value is the average of the difference (represents error) and the limits of agreement are calculated using ±1.96 × SD of the difference. Box plot: we used the *boxplot*(x,g) Matlab command to plot the box plot, where g is the time window groupings and x is the respective nCD64 values. Outliers are represented as red colored crosses. Standard multiple regression: standard multi-feature regression is performed using the Matlab function *regress*(y, X). This function outputs a vector of coefficient estimates for multivariate regression of the responses (nCD64) in y on the predictors in X (slope values, cell counts, and percentage of cell capture).

Hierarchical clustering (clustergram)

We performed the unsupervised hierarchical clustering analysis on the biosensor data (cell counts, nCD64 expression values and slope values at different experiment times) using Matlab's *clustergram*(*Data*) function. The function displays the heat map and the dendrogram. Furthermore, input data are normalized by log-mean normalization. The pairwise distance between pairs of objects is calculated by two different methods, 'Euclidean' and 'correlation'. *RowPdist* and *ColumnPdist* input variables to the clustergram function are used to input these pairwise distance methods.

Artificial neural networks (self-organizing maps)

We also clustered the data using self-organizing maps (SOMs). SOMs classify the input variables according to their grouping in the input space. Neighboring neurons in the SOM recognize neighboring sections of the input space and learn the distribution and topology for subsequent clustering of input variables. We used artificial neural networks (ANNs) in Matlab to cluster the biosensor's input data using the SOM. The corresponding network diagram of a single layer artificial neural network representing a self-organizing map with 100 neurons is shown in Fig. S6.† The network was trained using the batch SOM algorithm. The SOM neighbor weight distances show neurons as gray-blue patches and the neighbor relationship is shown with red lines. The neighboring hexagons are colored black–yellow, which represents the closeness of each neuron's weight vector to its corresponding neighbors. SOM sample hits: each neuron is represented by a hexagon and shows numbers which represent the number of inputs it classifies. SOM weight positions represent the input space that is classified by dots representing neuron weight vectors and the corresponding neighboring neurons are connected with red lines. SOM input weight planes represent the interconnectivity of different neurons. Zero connections

 This journal is © The Royal Society of Chemistry 2018

View Article Online

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

are represented as black, and strong positive connections are represented as red.

Artificial neural networks (multivariate regression)

Multivariate regression on the biosensor's data is performed using artificial neural networks in Matlab. The network diagram of a double layer forward feed artificial neural network with softmax output and sigmoid hidden neurons is shown in Fig. S9.† It consists of 10 neurons (hidden layer) and 1 neuron (output layer). The input space to the ANN model was normalized by the log-mean normalization method. Input variables including the slope values, cell counts and percentage of captured cells and the output target values for prediction are the corresponding nCD64 expression values for each sample. The network was trained using three different training methods.[41] (1) Levenberg Marquardt: this training method takes more memory but less time to train. Training stops when there is no improvement in generalization, which is also indicated by the increase in the mean squared error of samples.[41] (2) Bayesian regularization: this training method normally takes more time (as also indicated by epochs in Table 1) but results in better prediction accuracy and is used for more noisy datasets. This method stops training according to a regularization protocol *i.e.* adaptive weight minimization. (3) Scaled conjugate gradient: this method takes the least memory compared to other methods. Training stops when there is no improvement in generalization, which is also indicated by the increase in the mean squared error of samples.

Blood sample acquisition

Blood samples were obtained from patients who are SIRS+ and/or whose blood culture test is ordered by a doctor. The left-over samples were de-identified by hospital staff and sent to us. The blood sample collection process is approved through authorized University of Illinois Urbana-Champaign and Carle Foundation Hospital Institutional Review Board (IRB) applications (UIUC IRB number 15500 and Carle IRB number 15014).[30]

Conflicts of interest

R. B. and U. H. have financial interests in Prenosis, Inc. R. Z. declares no competing financial interests.

Acknowledgements

The authors thank the clinical collaborators including K. White, T. Jensen, B. Davis and J. Kumar at Carle Foundation Hospital, Urbana. The authors also thank M. Patel, N. Hung, M. Rappleye, Z. Haidry, T. Ghonge, C. Hwu, J. Berger, V. F. Bartumeus, E. Flaugher, G. Tufte, A. Hasnain, M. Saadah, S. Liu, and M. Chheda for their help in the device fabrication. The authors acknowledge R. Healey, N. Mansury, Z. Price, and A. Tanna for their help in performing whole blood experiments. The authors acknowledge the funding support from the University of Illinois at Urbana-Champaign and grant support from the Consortia for Improving Medicine with Innovation & Technology (CIMIT)'s Point-of-Care Technology Research Center in Primary Care.

References

1 T. Lagu, M. B. Rothberg, M. Shieh, P. S. Pekow, J. S. Steingrub and P. K. Lindenauer, Hospitalizations, costs, and outcomes of severe sepsis in the United States 2003 to 2007, *Crit. Care Med.*, 2012, **40**, 3.

2 D. F. Gaieski, M. Edwards, K. J. Kallan and B. G. Carr, Benchmarking the Incidence and Mortality of Severe Sepsis in the United States, *Crit. Care Med.*, 2013, **41**, 1167.

3 N. K. Adhikari, R. A. Fowler, S. Bhagwanjee and G. D. Rubenfeld, Critical care and the global burden of critical illness in adults, *Lancet*, 2010, **375**, 1339.

4 R. Daniels, Surviving the first hours in sepsis: getting the basics right (an intensivist's perspective), *J. Antimicrob. Chemother.*, 2011, **66**, 2.

5 J. Du, L. Li, Y. Dou, P. Li, R. Chen and H. Liu, Diagnostic utility of neutrophil CD64 as a marker for early-onset sepsis in preterm neonates, *PLoS One*, 2014, **9**, 7.

6 A. Dimoula, O. Pradier, Z. Kassengra, D. Dalcomune, H. Turkan and J. Vincent, Serial determination of neutrophil CD64 expression for the diagnosis and monitoring of sepsis in critically ill pateints, *Clin. Infect. Dis.*, 2014, **58**(6), 820–829.

7 J. Cid, R. Aguinaco, R. Sánchez, G. García-Pardo and A. Llorente, Neutrophil CD64 expression as marker of bacterial infection: A systematic review and meta-analysis, *J. Infect.*, 2010, **60**(5), 313–319.

8 X. Wang, Z. Y. Li. L. Zeng, A. Q. Zhang, W. Pan. W. Gu and X. J. Jiang, Neutrophil CD64 expression as a diagnostic marker for sepsis in adult patients: a meta-analysis, *Crit. Care*, 2015, **10**(19), 245.

9 S. Li, X. Huang, Z. Chen, H. Zhong, Q. Peng, Y. Deng, X. Qin and J. Zhao, Neutrophil CD64 expression as a biomarker in the early diagnosis of bacterial infection: a meta-analysis, *Int. J. Infect. Dis.*, 2013, **17**, e12–e23.

10 R. Repp, T. Valerius and A. Sendler, Neutrophils express the high affinity receptor for IgG (Fc gamma RI, CD64) after in vivo application of recombinant human granulocyte colony-stimulating factor, *Blood*, 1991, **78**, 885–889.

11 A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar and M. Cheang, Duration of hypotension before initiation of effective antimicrobial is the critical determinant of survival in human septic shock, *Crit. Care Med.*, 2006, **34**(6), 1589–1596.

12 X. Li, A. Ymeti, B. Lunter, C. Breukers, A. G. J. Tibbe, L. W. M. M. Terstappen and J. Greve, CD4+ T lymphocytes enumeration by an easy-to-use single platform image cytometer for HIV monitoring in resource-constrained settings, *Cytometry, Part B*, 2007, **72**(5), 397–407.

13 W. Rodriguez, N. Christodoulides, P. N. Floriano, S. Graham, S. Mohanty, M. Dixon, M. Hsiang, T. Peter, S. Zavahir, I.

Published on 16 March 2018. Downloaded by University of Illinois - Urbana on 13/04/2018 18:05:19.

Thior, D. Romanovicz, B. Bernard, A. P. Goodey, B. D. Walker and J. T. McDevitt, A microchip CD4 counting method for HIV monitoring in resource-poor settings, *PLoS Med.*, 2005, **2**, 182.

14 X. Cheng, D. Irimia, M. Dixon, K. Sekine, U. Demirci, L. Zamir, R. G. Tompkins, W. R. Rodriguez and M. Toner, A microfluidic device for practical label-free CD4(+) T cell counting of HIV-infected subjects, *Lab Chip*, 2007, **7**(2), 170–178.

15 Z. Wang, S. Chin, C. Chin and J. Sarik, Microfluidic CD4+ T-cell counting device using chemiluminescence-based detection, *Anal. Chem.*, 2010, **82**(1), 36–40.

16 W. Coulter, Means for counting particles suspended in fluid, *US Pat.*, 26565081953, 1953.

17 A. Adams, P. I. Okagbare, J. Feng, M. L. Hupert, D. Patterson, J. Göttert, R. L. McCarley, D. Nikitopoulos, M. C. Murphy and S. Soper, S. Highly efficient circulating tumor cell isolation from whole blood and label-free enumeration using polymer-based microfluidics with an integrated conductivity sensor, *J. Am. Chem. Soc.*, 2008, **130**(27), 8633–8641.

18 D. Holmes, D. Pettigrew, C. H. Reccius, J. D. Gwyer, C. Van Berkel, J. Holloway, D. E. Davies, H. Morgan and D. Morganti, Leukocyte analysis and differentiation using high speed microfluidic single cell impedance cytometry, *Lab Chip*, 2009, **9**(20), 2881–2889.

19 D. Holmes and H. Morgan, Single cell impedance cytometry for identification and counting of CD4 T-cells in human blood using impedance labels, *Anal. Chem.*, 2010, **82**(4), 1455–1461.

20 D. A. L. Vickers, E. J. Chory and S. K. Murthy, Seperation of two phenotypically similar cells via a single common marker in microfluidic channels, *Lab Chip*, 2012, **12**, 3399–3407.

21 Y. Zhang and D. Pappas, Microfluidic cell surface antigen expression using a single antibody type, *Analyst*, 2016, **141**, 1440–1447.

22 D. A. McQuarrie, Kinetics of small systems. I, *J. Chem. Phys.*, 1963, **38**, 433–435.

23 J. W. Piper, R. A. Swerlick and C. Zhu, Determining force dependence of two-dimensional receptor–ligand binding affinity by centrifugation, *Biophys. J.*, 1998, **74**, 492–513.

24 P. Decuzzi and M. Ferrari, The adhesive strength of non-spherical particles mediated by specific interactions, *Biomaterials*, 2006, **27**, 5307–5314.

25 U. Hassan, B. Reddy, G. Damhorst, O. Sonoiki, T. Ghonge, C. Yang and R. Bashir, A Microfluidic Biochip for Complete Blood Cell Counts at the Point-of-Care, *Technology*, 2015, **3**(4), 201–213.

26 U. Hassan, N. Watkins, C. Edwards and R. Bashir, Flow Metering Characterization within an Electrical Cell Counting Microfluidic Device, *Lab Chip*, 2014, **14**, 1469.

27 U. Hassan and R. Bashir, Coincidence detection of heterogeneous cell populations from whole blood with coplanar electrodes in a Microfluidic impedance cytometer, *Lab Chip*, 2014, **14**(22), 4370–4381.

28 N. Watkins, U. Hassan, G. Damhorst, H. Ni, A. Vaid, W. Rodriguez and R. Bashir, Microfluidic CD4+ and CD8+ T Lymphocyte Counters for Point-of-Care HIV Diagnostics Using Whole Blood, *Sci. Transl. Med.*, 2013, **5**(214), 214ra170.

29 U. Hassan, N. Watkins, B. Reddy, G. Damhorst and R. Bashir, Microfluidic Biochips for Cell Counting, *Nat. Protoc.*, 2016, **11**(4), 714–726.

30 U. Hassan, T. Ghonge, B. Reddy Jr., M. Patel, M. Rappleye, I. Taneja, A. Tanna, R. Healey, N. Mansury, Z. Price, T. Jensen, J. Berger, A. Hasnain, E. Flaugher, S. Liu, B. Davis, J. Kumar, K. White and R. Bashir, A point-of-care microfluidic biochip for quantification of CD64 expression from whole blood for sepsis stratification, *Nat. Commun.*, 2017, **8**, 15949, DOI: 10.1038/ncomms15949.

31 J. Friedman, T. Hastie and R. Tibshirani, *The elements of statistical learning*, Springer Series in Statistics, New York, 2001, vol. 1.

32 T. Kohonen, The self-organizing map, *Neurocomputing*, 1998, **21**(1), 1–6.

33 *MATLAB and Neural Network Toolbox Release 2017b*, The MathWorks, Inc., Natick, Massachusetts, United States.

34 I. Taneja, B. Reddy Jr., G. Damhorst, D. Zhao, U. Hassan, Z. Price, T. Jensen, T. Ghonge, M. Patel, S. Waschpress, J. Winter, M. Rapport, G. Smith, R. Healey, M. Ajmal, M. Khan, J. Patel, H. Rawal, R. Sarwar, S. Soni, B. Davis, J. Kumar, K. White, R. Bashir and R. Zhu, Predictive Power of Including Novel Biomarker Measurements with Electronic Medical Record Data in Identification of Sepsis, *Sci. Rep.*, 2017, 7, 10800, DOI: 10.1038/s41598-017-09766-1.

35 F. Venet, A. Lepape and G. Monneret, Clinical review: flow cytometry perspectives in the ICU – from diagnosis of infection to monitoring of injury-induced immune dysfunctions, *Crit. Care*, 2011, **15**(5), 231.

36 K. Wang, V. Bhandari and S. Chepustanova, *et al.* Which biomarkers reveal neonatal sepsis?, *PLoS One*, 2013, **8**(12), e82700.

37 K. Fairchild, Predictive monitoring for early detection of sepsis in neonatal ICU patients, *Curr. Opin. Pediatr.*, 2013, **25**(2), 172–179.

38 M. A. Owens, H. G. Vall and A. A. Hurley, *et al.*, Validation and quality control of immunophenotyping in clinical flow cytometry, *J. Immunol. Methods*, 2000, **243**, 33–50.

39 A. Vranic, Antigen expression on recurrent meningioma cells, *Radiol. Oncol.*, 2010, **44**, 107–112.

40 T. R. Daniels, T. Delgado, J. A. Rodriguez, G. Helquera and M. L. Penichet, The transferrin receptor part I: Biology and targeting with cytotoxic antibodies for the treatment of cancer, *Clin. Immunol.*, 2006, **121**, 144–158.

41 M. T. Hagan and M. Menhaj, Training feed-forward networks with the Marquardt algorithm, *IEEE Trans. Neural Netw.*, 1994, **5**(6), 989–993.

 This journal is © The Royal Society of Chemistry 2018