analytical chemistry

pubs.acs.org/ac Article

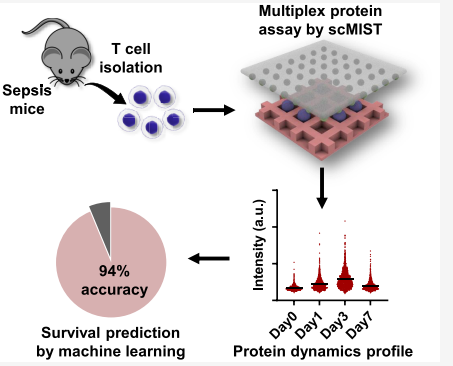
Rapid, High-Throughput Single-Cell Multiplex In Situ Tagging (MIST) Analysis of Immunological Disease with Machine Learning

Liwei Yang, Pratik Dutta, Ramana V. Davuluri, and Jun Wang*

Cite This: https://doi.org/10.1021/acs.analchem.3c01157 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The cascade of immune responses involves activation of diverse immune cells and release of a large amount of cytokines, which leads to either normal, balanced inflammation or hyperinflammatory responses and even organ damage by sepsis. Conventional diagnosis of immunological disorders based on multiple cytokines in the blood serum has varied accuracy, and it is difficult to distinguish normal inflammation from sepsis. Herein, we present an approach to detect immunological disorders through rapid, ultrahigh-multiplex analysis of T cells using single-cell multiplex in situ tagging (scMIST) technology. scMIST permits simultaneous detection of 46 markers and cytokines from single cells without the assistance of special instruments. A cecal ligation and puncture sepsis model was built to supply T cells from two groups of mice that survived the surgery or died after 1 day. The scMIST assays have captured the T cell features and the dynamics over the course of recovery. Compared with cytokines in the peripheral blood, T cell markers show different dynamics and cytokine levels. We have applied a random forest machine learning model to single T cells from two groups of mice. Through training, the model has been able to predict the group of mice through T cell classification and majority rule with 94% accuracy. Our approach pioneers the direction of single-cell omics and could be widely applicable to human diseases.



INTRODUCTION

Advanced single-cell analysis has profoundly impacted many biological fields and is widely regarded as an important propellant of precision medicine.[1−3] However, there is still a deep gap between the current single-cell analysis and early diagnosis of immunological diseases. The low-content single-cell studies do not possess much prediction value, while the high-multiplex single-cell tools are not feasible to routinely analyze samples from individuals. The prevailing single-cell sequencing[4] and mass cytometry[5,6] technologies require expensive instruments that are not accessible for many researchers and clinicians and are not portable either. Since cytokines play the central role in inflammation, single-cell cytokine technologies have been developed to correlate secretome with immune responses.[7−9] Most of them only focus on cytokines and do not include intracellular protein markers. The lack of a practically useful, information-rich, inexpensive, and portable single-cell technology severely hampers the effective treatment of an array of diseases including sepsis.

Precision diagnosis and prognosis of immunological diseases are still challenging, as there is no effective way to predict the consequence of inflammation, control the hyperinflammatory response, and prevent organ failure. Given the spatiotemporal complexity of the immune system, targeting a single inflammatory factor such as tumor necrosis factor alpha (TNFα) often fails the expectations in clinical therapies.[10−12] In the case of sepsis, currently there are no specific therapeutic interventions for clinical treatment of this life-threatening disease that are responsible for >250,000 deaths every year in the US.[13] Because there are numerous cellular processes involved in sepsis, finding reliable diagnostic biomarkers specific for sepsis has been almost impossible.[14] Currently, the most reliable way to predict sepsis is through multiplexed assays of a few cytokine biomarkers in the blood,[15−17] although the accuracy is varied. In addition, the mechanism of sepsis is debatable since the host immunity mechanism is not clear. In the traditional view, the host response to sepsis starts from a hyperinflammatory state, followed with chronic immunosuppression with impaired innate immune functions of phagocytes to remove infections.[18,19] But some reports asserted that the unabated inflammation persists for a long time and therefore causes organ injury.[20,21] The dysregulated inflammation is mainly mediated by cytokines, which are largely secreted by immune cells. However, almost no diagnosis and prognosis have been directly based on immune cell functions.

In recent years, machine learning and artificial intelligence have been widely adopted in biomedical fields and show their

Received: March 15, 2023
Accepted: April 25, 2023

superior performance over conventional models.[22,23] For example, single-cell RNA-sequencing data have been used to predict patient phenotypes, determine disease status, and stratify immune responses.[24,25] These methods are limited in practical use because of high cost and relatively low number of cells.[26] High-multiplex cytometry is relatively affordable to characterize dozens of samples or more through machine learning algorithms DGCyTOF and DeepCyTOF.[27,28] But the studies on predicting disease status and progression are still very limited due to the complexity and expensive instruments. Machine learning models normally need at least dozens of samples in training and prediction, whereas the most advanced single-cell technologies can hardly analyze more than 10 samples for each study. Thus, machine learning on rich single-cell data has been mostly focused on cell classification and cell subpopulation identification. Besides, most machine learning algorithms, particularly those based on deep learning, need extensive computer skills and require large data and computational resources to train the data.

In this report, we have combined single-cell multiplex in situ tagging (scMIST) technology with machine learning to precisely predict sepsis outcomes of a mouse model. We have analyzed 46 marker proteins from single T cells using the scMIST technology that can conveniently quantify both cytokines and surface markers from thousands of cells across dozens of samples. The 46 proteins panel was selected from the conventional phenotype biomarkers of T cell subsets, surface markers of T cell migration, proliferation and differentiation, and typical cytokines during inflammatory and anti-inflammatory responses (Supporting Information Table S1). The scMIST technology is simple, easy to follow, portable, and inexpensive for point of detection purposes. Meanwhile, it possesses the ability to measure dozens of marker proteins rapidly without the assistance of special instruments. We generated a sepsis mouse model and employed scMIST to assay T cell features from 20 mice. A novel approach based on a random forest algorithm has been created to bypass the limitation of sample size in machine learning training and predict sepsis outcomes. We found T cell function can be new biomarkers to predict sepsis onset and distinguish sepsis from normal inflammation. The trained model predicted sepsis outcomes with 94% accuracy, which is higher than most of the conventional, sophisticated diagnosis methods. Thus, our technology and the associated method can potentially be a new approach in early diagnosis and better prediction or treatment for septic shock patients.

EXPERIMENTAL SECTION

Microwell Chip Fabrication. The fabrication of poly(dimethylsiloxane) (PDMS) microwell chip was performed according to the standard soft lithography and microfabrication methods using SU-8 2025 photoresist (Microchem) and SYLGARD 184 Silicone Elastomer Kit (Dow). Briefly, a 10:1 mixture of PDMS prepolymer and curing agent was poured on an SU-8 mold and cured in an oven at 70 °C for 2 h after vacuum degassing. The cured PDMS slab was separated and shaped into the appropriate size for the subsequent cell-loading step. The size of each microwell of the chip is 45 μm (length) × 45 μm (width) × 40 μm (depth).

MIST Microbead Array Preparation. 46 oligo DNA pairs (15−20 bases for each) were custom-designed and purchased from Integrated DNA Technologies (Supporting Information Table S2). The cross-reactivity between oligos has been fully validated in our previous study.[29] Before oligo modification, microbeads (2 μm; Life Technologies) were first functionalized with poly-L-lysine (Ted Pella) through 10 mM bis(sulfosuccinimidyl)suberate (Thermo Scientific) cross-linking. The microbeads were further mixed with 300 μM amine-ended oligo and 2 mM bis(sulfosuccinimidyl)suberate for 3 h. Finally, the microbeads were thoroughly washed by Milli-Q water and resuspended to the original volume. To make MIST arrays, 46 oligo-modified microbeads with equal portion were mixed and then added into blank microbeads with a volume ratio of 4:1. Afterward, the mixed microbeads were deposited on an adhesive tape to form a uniform and monolayered MIST microbead array. To determine the detection sensitivity of the microbead array, various concentrations of complementary oligo tagged with Cy5 fluorophore were added on the array and incubated for 1 h before imaging on a Nikon Ti2 inverted fluorescence microscope. The fluorescence intensities of each oligo-modified microbead were quantified and identified by an in-house MATLAB program code. The limit of detection for each type of microbead was calculated as the background intensity plus 3 times standard deviation (SD) based on the fitting curves.

Complementary Oligo−Antibody Conjugation and Purification. The preparation of biotinylated complementary oligo−antibody conjugate was conducted by click chemistry coupling as reported.[29] 50 μg of antibody at 1 mg/mL was reacted with photocleavable azido-NHS Ester (Click Chemistry Tools) at 1:10 molar ratio for 3 h, and meanwhile 6 μL of 1 mM oligo was reacted with photocleavable DBCO-NHS Ester (Click Chemistry Tools) at 1:20 molar ratio for 3 h. The azido-antibody and DBCO-oligo were purified by 7 K Zeba spin desalting column and mixed for overnight reaction, before purification by a Fast Protein Liquid Chromatography (FPLC; ÄKTA) system. The collected product was concentrated to 0.5 mg/mL measured by a NanoDrop Spectrophotometer (Thermo Fisher) and stored at 4 °C for further use.

Mouse Model, Sepsis Induction, and T Cell Isolation. The following animal procedure was conducted in accordance with the US National Institutes of Health Guide for the Care and Use of Laboratory Animals and was approved by the Stony Brook University Institutional Animal Care and Use Committee (IACUC). Male wild-type C57BL/6 mice (age: 8−12 weeks; weight 20−30 g) were purchased from Charles River. Since sex-specific outcomes in rodent models of sepsis by diverse immunomodulatory functions have been reported,[30] we only used male mice in this study to obtain consistent data. Sepsis was induced in mice by a conventional cecal ligation puncture (CLP) procedure as previously described.[31] The mice were divided into two groups: the control group without any CLP surgery (n = 4) and the experimental group varying degrees of ligation procedure (n = 16). For the blood sample collection, the peripheral blood (100−150 μL) was drawn from the superficial temporal vein of the mice at the first, third, and seventh day after the CLP surgery if the mice maintained alive, and then were proceeded to the T cell isolation and scMIST analysis.

The collected blood sample was first incubated with 5 μg/mL Brefeldin A (BioLegend) for 3 h to block protein translocation to the cell surface, so that the expressed proteins can be maximally retained during the immunological activation.[32,33] The blood sample was further processed with RBC Lysis Buffer (Invitrogen) to remove excess red blood cells. The T cell was negatively isolated from the blood by

https://doi.org/10.1021/acs.analchem.3c01157
Anal. Chem. XXXX, XXX, XXX−XXX

using the commercial Mouse Pan T Cell Isolation Kit II (Miltenyi Biotec) according to the manufacturer manual. Cells were stained with CD3, CD4, and CD8 to validate the separation. The T cells were resuspended to a concentration of 100,000 cells/mL with a pH 7.4 phosphate-buffered saline (PBS) buffer containing 0.5% bovine serum albumin (BSA) for the subsequent cell loading procedure.

Cytokines Detection in Blood Plasma. The cytokines in the mouse blood were determined by conventional sandwich-like ELISA with a mouse cytokine antibody array (RayBiotech Inc., Norcross, GA). Briefly, 50 μL of mouse blood was centrifuged at 2500*g* for 15 min to obtain the plasma supernatant. The plasma was applied on a glass slide coated with the antibody array for 1 h. After it was rinsed by a washing buffer five times, the slide was incubated with a biotinylated antibody cocktail solution for 1 h. Then, the slide was washed and treated by Cy3 equiv dye-conjugated streptavidin. The slide was imaged under a fluorescence microscope, and the intensity of the slide was quantitatively analyzed by ImageJ (FIJI Version 1.53).

Multiplexed Single-Cell Protein Assay on MIST Array. Before cell loading, the surface of the PDMS microwell chip was treated with a plasma cleaner (Harrick Plasma) for 2 min and incubated with a Pluronic F-127 (Sigma-Aldrich; 1% in pH 7.4 PBS buffer) solution for 30 min to reduce nonspecific adsorption. The chip was washed with PBS buffer three times, and 100 μL of a mouse T cell suspension with a concentration of 100,000 cells/mL was deposited on the microwells for 12 min with slow shaking to load the cells. After we removed the excess cells, the cells in the microwells were fixed by 4% formaldehyde in PBS buffer for 15 min and permeabilized with 0.1% Triton-X100 in PBS buffer for 7 min. Afterward, the cells were incubated with a blocking buffer containing 10% goat serum (Cell Signaling Technology), 2% BSA (Fisher Scientific), and freshly prepared 1 mg/mL Salmon Sperm DNA (Invitrogen) for 1 h. After washing three times with PBST solution (0.1% Tween 20 in pH 7.4 PBS buffer), the cells were stained with a mixture of oligo-antibody conjugates (5 μg/mL for each in the blocking buffer) for 1 h and then thoroughly washed by the PBST solution. The microwell chip was carefully mated with an MIST array and clamped tightly by magnetic force. The whole setup was exposed to a 365 nm UV light (Thorlabs) for 15 min to release the biotinylated complementary oligo and allow its hybridization on the MIST array. After separation from the microwell chip, the MIST array was washed by PBS buffer and further incubated with 5% goat serum/PBS solution for 30 min. Meanwhile, the cells in the microwells were stained with 5 μg/mL Hoechst 33342 (Pierce) in PBS buffer for 15 min. For signal visualization on the array, 10 μg/mL streptavidin-Alexa Fluor 647 dye (Life Technologies) in 5% goat serum/PBS solution was applied onto the array for 15 min and then imaged under a fluorescence microscope, which was denoted *Protein* signal.

MIST Array Decoding Process. The decoding process is to identify which oligo or its corresponding protein is detected on each microbead of the MIST array. The array was treated with 1 M NaOH solution for 1 min, washed with 2X saline-sodium citrate buffer (SSC; Alfa Aesar), and then incubated with a cocktail of 200 nM complementary oligos tagged with various fluorophores (Decoding Cocktail Cycle 1, Supporting Information) in a hybridization SSC buffer containing 40% formamide (Fisher Scientific) and 10% dextran sulfate (Alfa Aesar) for 1 h. After washing three times with SSC buffer, the array was scanned by a fluorescence microscope, thus obtaining the decoding signal for *Cycle 1*. The second and third decoding cycles were performed as the same procedure of the first decoding cycle, except that the Decoding Cocktail Cycle 2 and Cycle 3 (Supporting Information) solutions were applied onto the array, resulting in the decoding signals for *Cycle 2* and *Cycle 3*, respectively.

Imaging and Registration. All images were autoscanned by a Nikon Ti2 inverted fluorescence microscope equipped with a motorized stage. The PDMS microwells were imaged to identify the number and address of cells in each microwell. A laboratory-developed MATLAB program code was used for image registration and signal quantification. The microbeads detecting the same protein were identified by the program, and their fluorescence intensities in the *Protein* signal images were quantified and averaged for each microwell or array. After combining the information on cell number and location in the microwell chip, a final data set in the table was generated including the cell number (zero-cell or single-cell) and its detected protein species/intensities.

Machine Learning. A Random Forest model was built to classify the cells from 16 mice based on the highest proportion of the cell prediction calls, and then a random forest-based variable selection algorithm[34] was applied to select a small set of nonredundant proteins that could best discriminate cells that belong to different classes of mice. All the cells that belong to a mouse were labeled as same as that of the mouse. By selecting 20 proteins as the most discriminative features between the survival and death groups, a random forest classifier was created for survival prediction.[35,36] The cross-validation analysis of the selected classifier was done by an out-of-bag [OOB] approach. The feature selection and classification modeling were performed by leaving out one mouse at a time, and the classifier was applied to predict the survival of the mouse. Then, the predicted label of the left-out mouse was determined based on the highest proportion of the predicted cells. This procedure was repeated for all the mice, and performance statistics were calculated based on the true and predicted class labels of the mice. In order to mitigate the influence of outliers on the results, we applied equal frequency data discretization to transform each feature into ranked bins as we have done previously.[37,38] Data discretization is a common technique used in machine learning to simplify the data analysis and enhance its interpretability, while also reducing the sensitivity of the results to outliers. Here, we applied data discretization for converting continuous data values into categorical data. By discretizing the data, the effects of extreme values and outliers are minimized, allowing for more robust and reliable statistical analysis.

Data Analysis and Statistics. The protein detection intensities in the empty microwells were taken as the background, and the signal detection limit (confident threshold) was set as the background intensity plus 3 times of its standard deviation. Only the detected protein intensities of each microwell/array above the confident threshold were considered as positive signals. Prism 9 (GraphPad) was used to generate all figure plots and statistically perform the data analysis. A nonparametric Mann−Whitney U test was used to evaluate the statistical significances, and a *P*-value less than 0.05 is considered statistically significant and is denoted with *, while ** and *** represent $P < 0.01$ and $P < 0.001$, respectively. The uniform manifold approximation and

https://doi.org/10.1021/acs.analchem.3c01157
Anal. Chem. XXXX, XXX, XXX−XXX

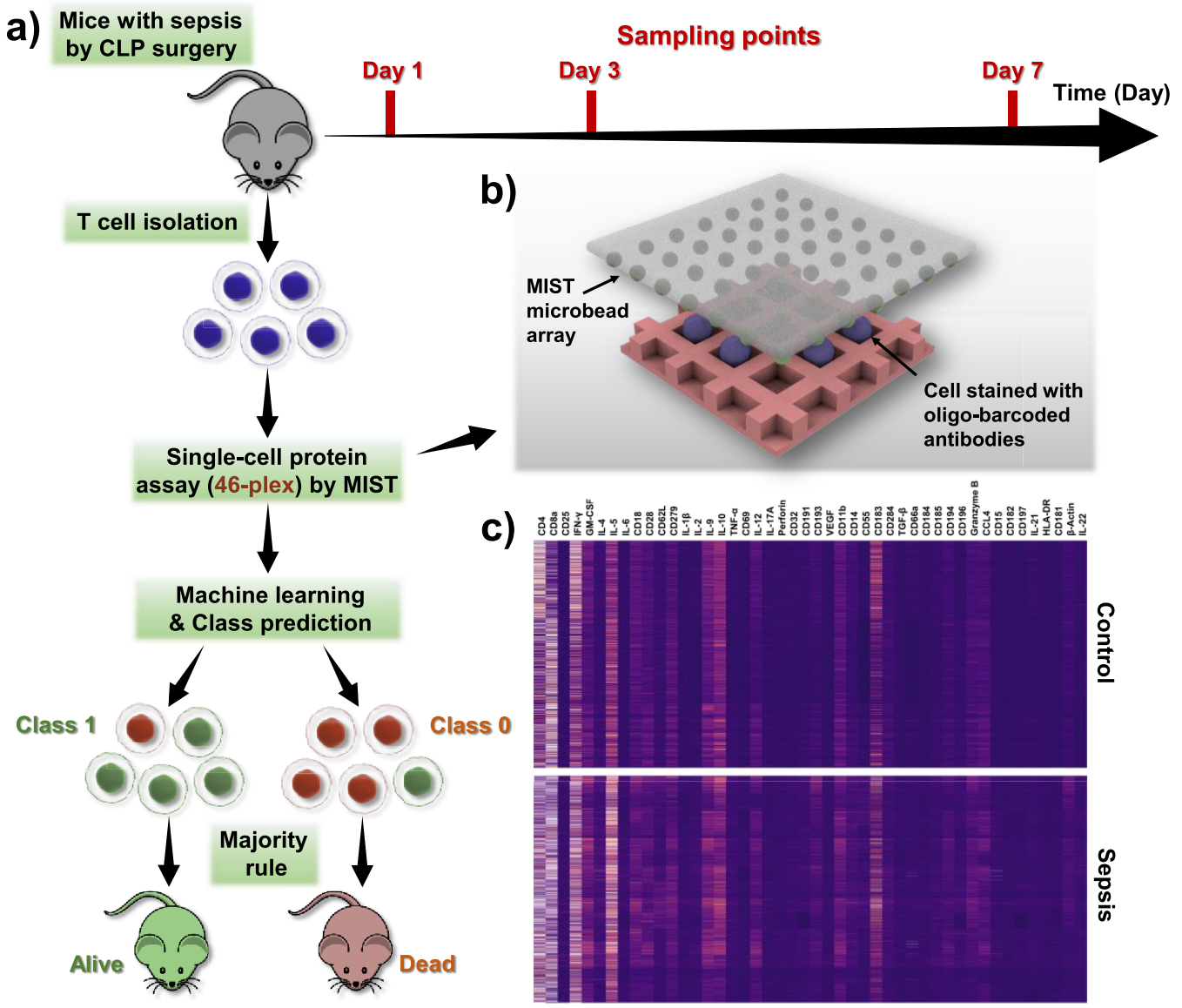


Figure 1. Overview of scMIST and machine learning for prediction of sepsis outcomes. (a) Scheme of workflow. Blood samples from a sepsis mouse model are collected, and single T cells are analyzed by the 46-plex scMIST assay. High-dimensional data are further analyzed by a Random Forest classification model to classify cells and mice into two categories: survival and death. (b) Illustration of scMIST in multiplex detection of isolated cells in PDMS microwells. (c) Representative heat maps (control vs sepsis mice) showing single-cell protein profiles. Each row represents a complete protein profile from individual cell, and each column is the proteins of interest.

projection (UMAP) plot was made by the use of Seurat package in R.

RESULTS AND DISCUSSION

We first generated a sepsis mouse model through CLP surgery and took the purified T cells from mouse blood for scMIST analysis. Mice with no or mild sepsis can survive for at least a week and are visually recovered, while severe sepsis results in death within 1 day. We selected a large panel of 46 marker proteins for the scMIST analysis. A machine learning model random forest has been used to train the pooled single cells from 16 mice, and the trained model is used to classify all the single cells into two categories. Following majority rule, the dominant single cells for each mouse determine the group (survival or death) of that mouse (Figure 1a).

The scMIST technology is a highly multiplexed single-cell protein detection platform composed of a PDMS replica carrying microwells to enclose labeled single cells and a matched MIST array to detect signals from cells (Figure 1b). The scMIST platform is capable of not only assaying cytokines but also detecting surface markers, intracellular proteins, and nuclear proteins, while similar technologies like Isoplexis products primarily focus on cytokines (Figure 1c). The detection involves immunostaining by UV-cleavable complementary oligo-antibody conjugates and multicycle decoding process on an oligo-barcoded microbead array. Specifically, single cells in PDMS microwells are stained by a mixture of the conjugates and are mated with the MIST microbead array. Upon UV exposure, the complementary oligos are released from the conjugates of the cells and are further hybridized and detected on the array. As a result, the signal intensities on the

https://doi.org/10.1021/acs.analchem.3c01157
Anal. Chem. XXXX, XXX, XXX−XXX

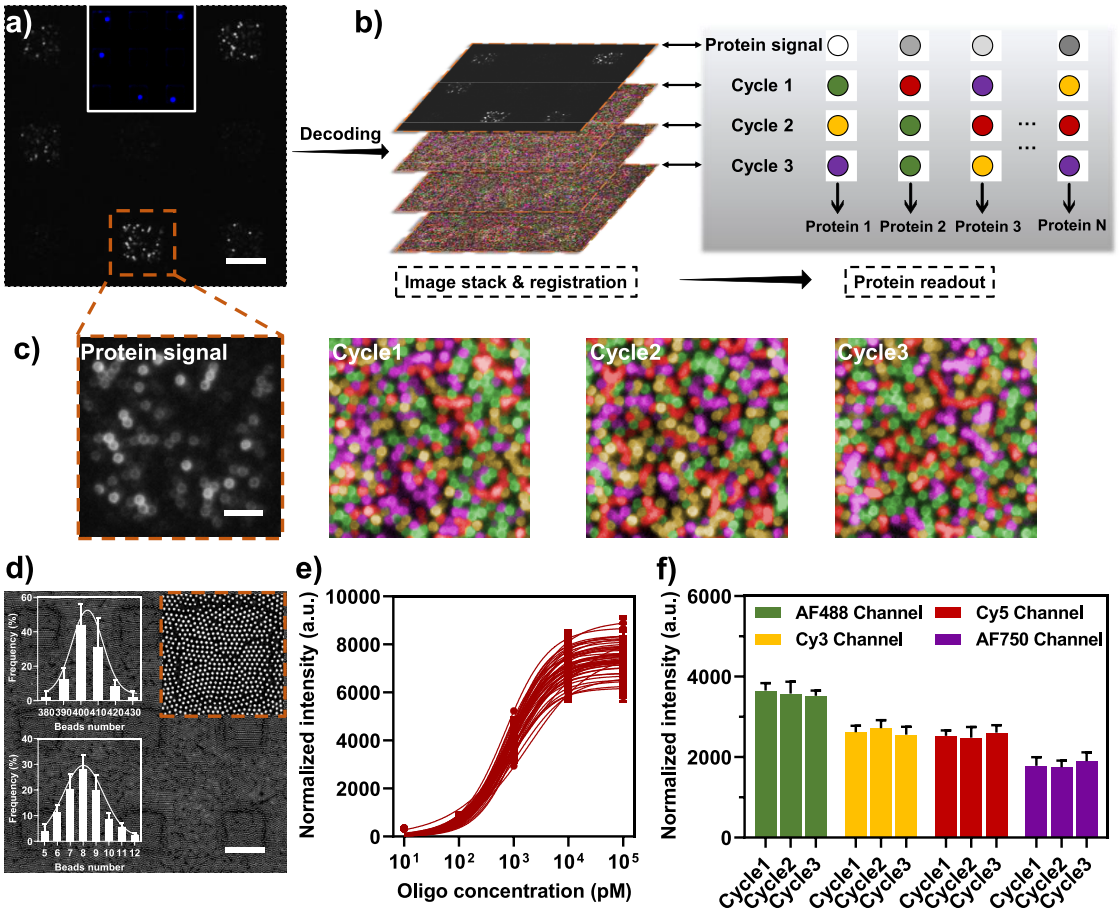


Figure 2. Characterization of scMIST technology. (a) Representative fluorescence image after protein detection on an MIST array and its corresponding cell-loading image (inset). Scale bar = 45 μm. (b) Principle of the decoding process where the quantified fluorescence signals on microbeads are assigned to specific proteins. (c) Representative fluorescence images (Protein signal and the three Decoding Cycles) of the multiplexed protein array from a single cell by MIST technology. Scale bar = 10 μm. (d) Representative bright-field image of an MIST array after clamping with PDMS microwells. Scale bar = 45 μm. Insets are the zoom-in bright-field image of each array and the distribution of the total bead number and the same oligo-modified bead number in each array. (e) Calibration by MIST array by varying the concentration of complementary oligos. The calibration curves fit with the logistic growth function. (f) Reproducibility of the MIST array validated by using the same fluorophore tagged-oligos for three decoding cycles. Data are presented as mean values ± SD of more than three independent experiments.

microbeads, which correspond to the amount of released complementary oligos, are proportional to the amount of single-cell proteins being detected (Figure S1 and Figure 2a).

To identify which oligo is detected on every microbead, a decoding process has been employed through reiterative cycles of dissociation and hybridization with different fluorophore tagged complementary oligos. After the registration of the images of all the decoding cycles, each microbead exhibits its sequential color change as predesigned in the decoding table (Supporting Information Table S3), so that the type of oligo or the species of protein being detected on individual beads can be determined accordingly (Figure 2b). Figure 2c shows the typical images of the array during protein detection and decoding cycles. Protein signals are well-confined within the microwell area, indicating no signal leakage during detection. In the decoding process of this study, 4 different dyes including Alexa Fluor 488 (green), Cy3 (yellow), Cy5 (red), and Alexa Fluor 750 (purple) combined with 3 decoding cycles were utilized to identify each type of microbeads, and theoretically up to 4^3 = 64 proteins can be simultaneously detected for single cells.

The MIST array has been fully characterized to ensure high data quality in the multiplexed single-cell protein detection. Each array at 45 μm × 45 μm contains 380−430 of 2 μm microbeads following Gaussian distribution (Figure 2d). For multiplexity of 46, around 8 copies of the same oligo-coated microbead are found on each array, while such variation of bead number has no significant effect on the assay quantification and reproducibility according to the previous work.[29] The detection sensitivity of the MIST array was quantified through hybridization of Cy5-tagged complementary oligos at various concentrations, the process of which simulates capturing of the released oligos in the scMIST assays. The average detection limit across all 46 oligos was determined as 3.2 pM (Figure 2e). Such a high sensitivity is mainly attributed to the dense oligos coated on the polylysine-grafted microbead surface. Figure 2f validates the reproducibility of the detected signal on the microbeads during successive cycles of dissociation and hybridization. Only <2.1% variation is found across three cycles for each fluorescent color. Our method could have high sensitivity, reproducibility, and multiplexity simultaneously, whereas the multiplexity and sensitivity are

https://doi.org/10.1021/acs.analchem.3c01157
Anal. Chem. XXXX, XXX, XXX−XXX

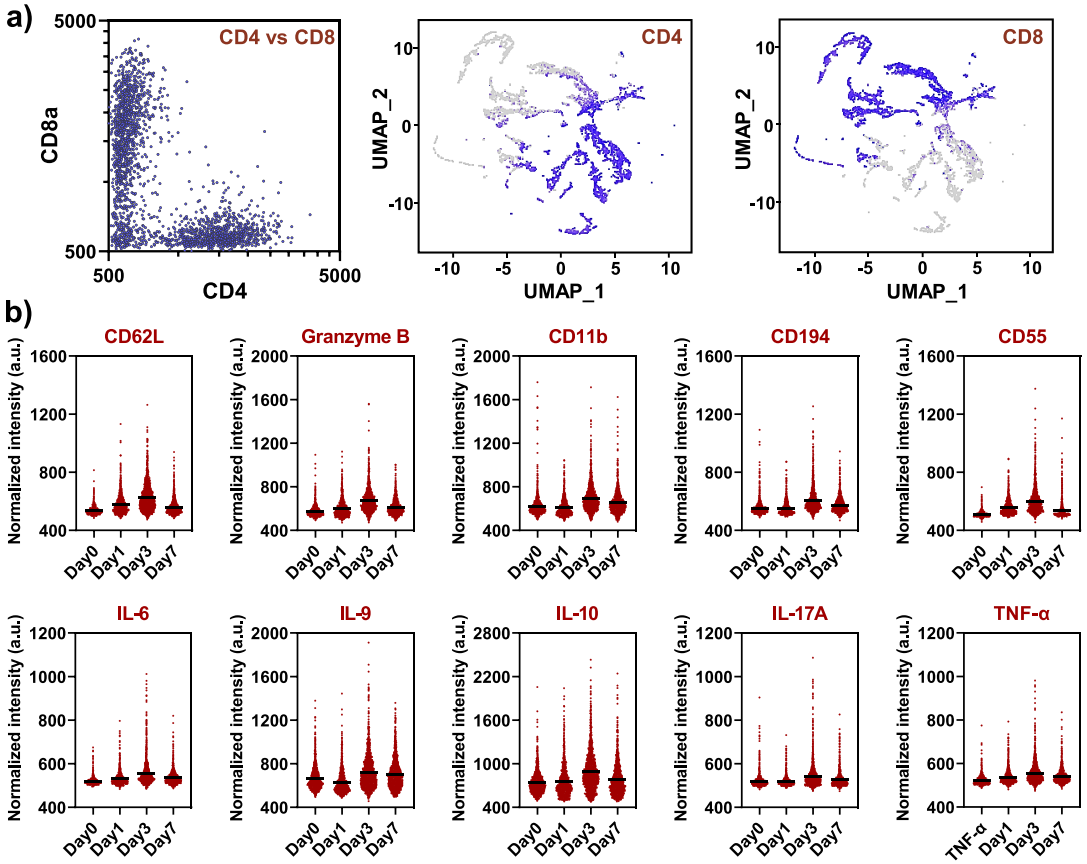


Figure 3. scMIST assay result of T cells from sepsis mice. (a) Scatter plot of T cells expressing CD4 and CD8 and their clusters on a UMAP plot. (b) Expression dynamics of selected proteins in single T cells from scMIST assays. The protein expressions data are merged from all the sepsis mice that survived the CLP surgery for at least 7 days. Black bars indicate the mean value of the scatter plots.

often compromised in commercial platforms like Isoplexis chips that use a much larger size of barcode arrays.

Single T cells from septic mice have been profiled by scMIST with 46-plex protein detection (the protein panel is provided in the Supporting Information Table S1). CLP surgery is a gold-standard method to generate septic shock in mice.[31] Blood samples were withdrawn periodically from septic mice as long as they were still alive (up to 7 days). The fresh samples were quickly processed to purify T cells and analyzed by scMIST. It is found that CD4 and CD8 cells are clearly separated (Figure 3a) into two groups in the dot plot and UMAP plot as expected. We also investigated the dynamics of protein expression in the mice that were still alive after 7 days. The single-cell data for those mice are pooled together in the dot plots (Figure 3b). Ten representative proteins uniformly show elevated expression after CLP surgery and fall back on day 7, which indicates inflammation being involved and T cells being activated (other expression dynamics of top featured proteins are provided in the Supporting Information Figure S2). Many markers such as Granzyme B are temporarily in high expression on day 3, possibly due to the adaptive immune system being triggered. The pro-apoptotic protease Granzyme B is critical for elimination of infected cells and invading pathogens, and it is also known for association with increased disease severity and a higher risk of death.[39] Likewise, expression of CD194, a marker of severe systemic inflammatory response to infection, is also augmented on day 3 together with CD11b, which served as a crucial adaptor for the migration of leukocytes from the bloodstream to sites of infection.[40] Meanwhile, anti-inflammatory activity represented by IL-10 is also increased on day 3. After 7 days, the protein expression levels are relatively similar to those of the control data. Consistent with clinical signs, the immune systems in those mice might have reached homeostasis after 7 days. The scMIST assay results suggest that septic conditions upregulate the expression of surface markers on T cells, which could represent early markers for assessment of sepsis prognosis. Additionally, the top-featured biomarkers identified may be considered potential therapeutic targets for sepsis treatment. For example, drugs that block the interaction between CD194 or Granzymes and its ligands have been shown to reduce the severity of sepsis in animal models.[41,42]

Since sepsis death is related with unbalanced inflammation and anti-inflammation, we have investigated the inflammation-related proteins from T cells and also in blood from the mice that died on day 1 after sepsis onset (Figure 4a). The dying mice obviously have lower cytokine production than the T cells, which might have lost activity due to a shortage of energy supply right before the mouse death. However, the cytokine levels in those mice are much higher in the blood sera than those in the survival group (Figure 4b). For instance, the pro-inflammatory factor IL-6 and anti-inflammatory cytokine IL-10 are both at least four times higher in the dying mice. Elevated IL-17 indicates neutrophils were significantly active during the sepsis onset, which can cause severe damage to organs. Together with TNFα, these three cytokines are often used as

https://doi.org/10.1021/acs.analchem.3c01157
Anal. Chem. XXXX, XXX, XXX−XXX

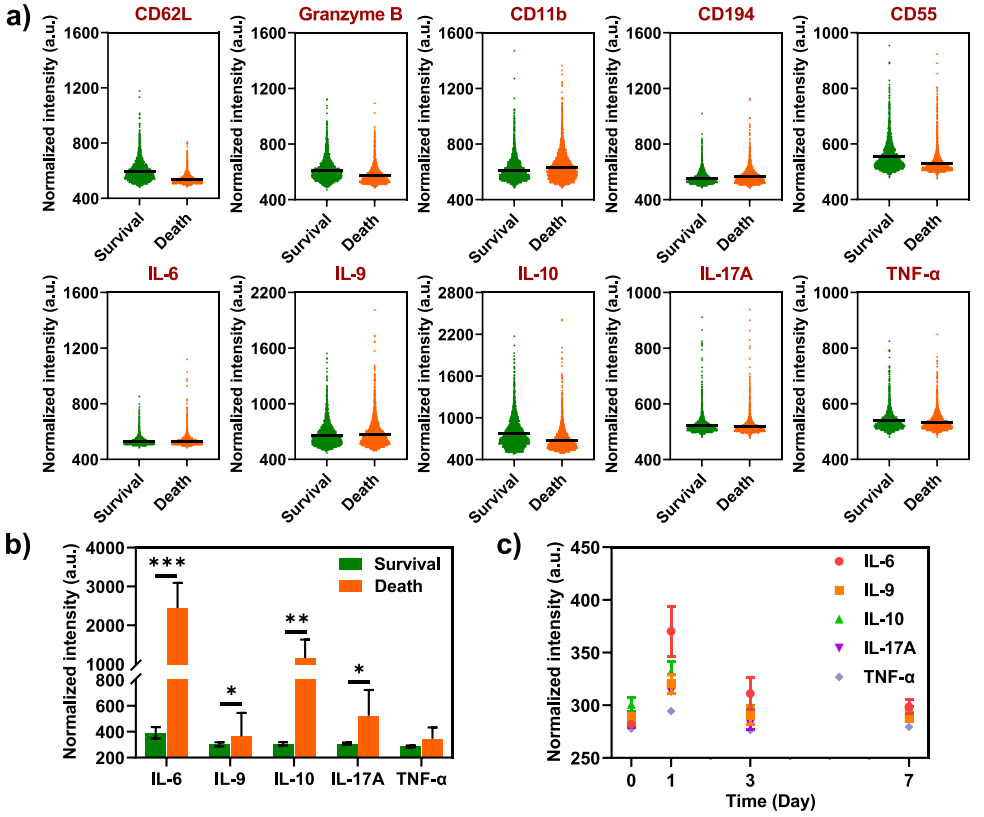


Figure 4. Differential expression of selected proteins in single T cells and in blood sera of survival and death mouse groups. (a) Protein levels in the survival and death groups where the aggregated scMIST data on day 1 after CLP surgery is used in the scatter plots. Black bars indicate the mean value of the scatter plots. (b) The intensity of 5 selected cytokines from the plasma of the survival and dead mice measured by conventional sandwich ELISA. (c) The dynamics of 5 selected cytokines from the sepsis mice before CLP surgery and on the 1st, 3rd, and 7th day after the surgery. Data are presented as mean values ± SD of more than three independent experiments, and error bars are within symbol size if not shown.

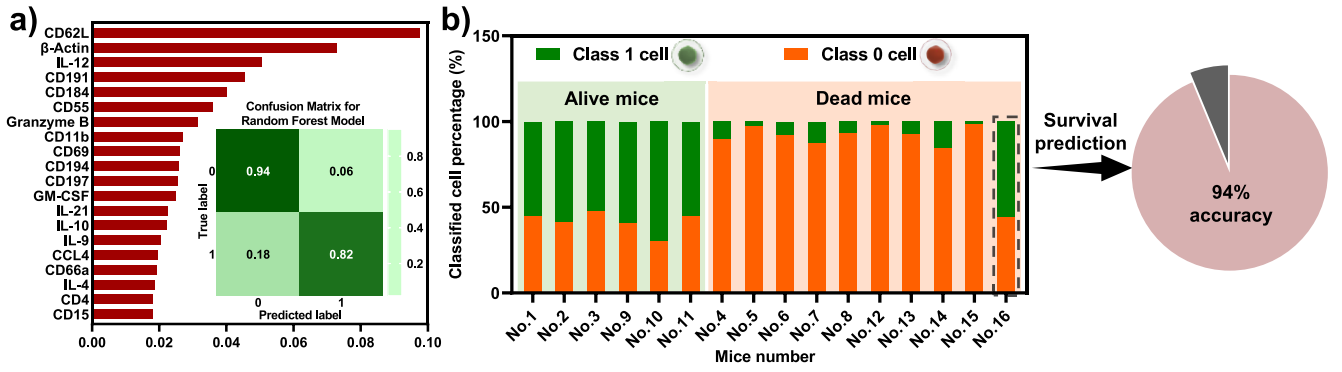


Figure 5. Cell classification and sepsis outcomes predicted by Random Forest algorithm. (a) Ranked importance of top features (proteins) in discriminating mouse groups using day 1 scMIST data. Inset describes the confusion matrix of the machine learning model. (b) Prediction result from the optimized model using top 20 features. Survival group (green) on the left and death group (orange) on the right are group truth for each mouse. Color label on each bar indicates the percentage of cells in two groups predicted by the algorithm, while the overall group of a mouse is determined by majority rule of cells from that mouse.

biomarkers of sepsis in an intensive care unit (ICU).[43] Our data show that early loss of T cell function could be an early indicator of severe sepsis, and cytokine profiles of T cells are departed far from cytokines in the blood. That might be because they were produced by other immune cells and the cytokine recycling in the blood was delayed. Our study found that T cell function is timely reflecting the current physiological status. For those mice that survived sepsis, the cytokine profile indicates the peak of inflammation appears on day 1, and the cytokine levels were attenuated after day 3 and fully recovered on day 7 (Figure 4c). In comparison with T cell function dynamics, a cytokine storm showed up in the early time of infection and sepsis, which should be mainly contributed by the innate immune system.

A machine learning algorithm random forest has been used to facilitate accurate prediction of sepsis death from T cell functions. Different from conventional approaches, we labeled every cell as either survival (1) or death (0), while a mouse

https://doi.org/10.1021/acs.analchem.3c01157
Anal. Chem. XXXX, XXX, XXX−XXX

typically has at least 500 such labeled cells. The training and prediction were also applied to single cells instead of individual mice. The training result shows the importance of proteins in discriminating the two groups (Figure 5a). CD62L, which is the naïve T cell marker, has been found to be the most significant protein in the trained model. By following the majority rule, the dominant cell group determines the predicted mouse fate. As shown in Figure 5b, such a classification approach results in 94% accuracy for totally 16 mice in the study. Probably the most interesting discovery is that, in the 9/10 dead mice (ground truth), the predicted class 0 cells almost occupy 90% of the population, while such a clear distinction is not observed in the survival group of mice. It is likely that the T cell functions in dying mice are more similar to each other, whereas they are more diversified in the alive mice. The only mouse that is not correctly predicted is mouse #16, which received exactly the same surgery together with another 3 mice in a batch that survived for 7 days. It accidently died early maybe because of larger incision and lack of food feeding.

CONCLUSION

In summary, we have developed a method in combination of single-cell experiments and machine learning algorithm to predict immunological disease outcomes through mapping of T cell markers. The scMIST assays have captured the T cell activity status and dynamics over the course of inflammation. While the overall profiles of T cell cytokines and surface makers are similar between the alive and dying sepsis mice, the selected markers such as CD62L and IL10 already show highly different expression on the first day after sepsis onset. It is interesting to observe the deep discrepancy between T cell cytokine levels and the cytokine amount in the blood serum. A random forest algorithm was employed to predict the outcomes of sepsis based on T cell markers. Model training found ranked importance of top features (proteins) in discriminating mouse groups. Based on the top 20 features, we classified the single cells and accordingly the mice by majority rule. Surprisingly, the T cells in dying mice are uniformly pointing to the same group, showing that they either lack diversity or have similar activity. With this method, we found sepsis outcome can be predicted with 94% accuracy which surpasses many conventional methods based on multiplex assays of serum cytokines. In the future, more mouse samples involved in the study could enhance the robustness and prediction value of our method. More rapid processing of cells will be helpful to minimize the change of molecular features. For example, the cell sorting and loading as well as removal of red blood cells could be achieved on chip in one step through the attachment of magnetic beads to the targeted cells and the design of microwells to capture single cells. In all, this T cell marker-based machine learning represents a completely new approach in the diagnosis and prognosis of immunological diseases and should be broadly applicable to human diseases with further development.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c01157.

Schematic illustration of scMIST process; protein expression dynamics by scMIST assay; the information on protein panel, their functions, oligos, and decoding design for scMIST technology (PDF)

AUTHOR INFORMATION

Corresponding Author

Jun Wang − *Multiplex Biotechnology Laboratory, Department of Biomedical Engineering, State University of New York at Stony Brook, Stony Brook, New York 11794, United States;* orcid.org/0000-0002-8781-8248; Email: jun.wang.5@stonybrook.edu

Authors

Liwei Yang − *Multiplex Biotechnology Laboratory, Department of Biomedical Engineering, State University of New York at Stony Brook, Stony Brook, New York 11794, United States;* orcid.org/0000-0001-7834-0249

Pratik Dutta − *Department of Biomedical Informatics, State University of New York at Stony Brook, Stony Brook, New York 11794, United States*

Ramana V. Davuluri − *Department of Biomedical Informatics, State University of New York at Stony Brook, Stony Brook, New York 11794, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c01157

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health R21AG072076 (J.W.), R21AG075087 (J.W.). and R01GM128984 (J.W.). The authors also thank Rachel Brownlee and Marissa Mastromarino in the Division of Laboratory Animal Resources at Stony Brook School of Medicine for the help of CLP surgery with sepsis mice model.

REFERENCES

(1) Heath, J. R.; Ribas, A.; Mischel, P. S. *Nat. Rev. Drug Discov* **2016**, *15* (3), 204−216.

(2) Dutta, A. K.; Alberge, J. B.; Sklavenitis-Pistofidis, R.; Lightbody, E. D.; Getz, G.; Ghobrial, I. M. *Nat. Rev. Clin Oncol* **2022**, *19* (4), 223−236.

(3) Mondal, M.; Liao, R. J.; Guo, J. *Chem.-Eur. J.* **2018**, *24* (28), 7083−7091.

(4) Grun, D.; van Oudenaarden, A. *Cell* **2015**, *163* (4), 799−810.

(5) Bendall, S. C.; Simonds, E. F.; Qiu, P.; Amir, E. A. D.; Krutzik, P. O.; Finck, R.; Bruggner, R. V.; Melamed, R.; Trejo, A.; Ornatsky, O. I.; Balderas, R. S.; Plevritis, S. K.; Sachs, K.; Pe'er, D.; Tanner, S. D.; Nolan, G. P. *Science* **2011**, *332* (6030), 687−696.

(6) Bjornson, Z. B.; Nolan, G. P.; Fantl, W. J. *Curr. Opin Immunol* **2013**, *25* (4), 484−494.

(7) Ma, C.; Fan, R.; Ahmad, H.; Shi, Q. H.; Comin-Anduix, B.; Chodon, T.; Koya, R. C.; Liu, C. C.; Kwong, G. A.; Radu, C. G.; Ribas, A.; Heath, J. R. *Nat. Med.* **2011**, *17* (6), 738−U133.

(8) Lu, Y.; Xue, Q.; Eisele, M. R.; Sulistijo, E. S.; Brower, K.; Han, L.; Amir, E. D.; Pe'er, D.; Miller-Jensen, K.; Fan, R. *P Natl. Acad. Sci. USA* **2015**, *112* (7), E607−E615.

(9) Deng, J.; Ji, Y. H.; Zhu, F. J.; Liu, L. N.; Li, L. M.; Bai, X.; Li, H. B.; Liu, X. M.; Luo, Y.; Lin, B. C.; Lu, Y. *Proc. Natl. Acad. Sci. USA* **2022**, *119* (44), e2200944119.

(10) Wenzel, R. P.; Edmond, M. B. *New Engl J. Med.* **2012**, *366* (22), 2122−2124.

(11) Marshall, J. C. *Trends Mol. Med.* **2014**, *20* (4), 195−203.

(12) Lorente, J. A.; Marshall, J. C. *Shock* **2005**, *24* (Supp1), 107−19.

(13) Bosmann, M.; Ward, P. A. *Trends Immunol* **2013**, *34* (3), 129−136.
(14) Biron, B. M.; Ayala, A.; Lomas-Neira, J. L. *Biomark Insights* **2015**, *10*, 7−17.
(15) Buchegger, P.; Preininger, C. *Anal. Chem.* **2014**, *86* (6), 3174−3180.
(16) Chousterman, B. G.; Swirski, F. K.; Weber, G. F. *Semin Immunopathol* **2017**, *39* (5), 517−528.
(17) Crapnell, R. D.; Jesadabundit, W.; Garcia-Miranda Ferrari, A.; Dempsey-Hibbert, N. C.; Peeters, M.; Tridente, A.; Chailapakul, O.; Banks, C. E. *Anal. Chem.* **2021**, *93* (14), 5931−5938.
(18) Munford, R. S.; Pugin, J. *Am. J. Resp Crit Care* **2001**, *163* (2), 316−321.
(19) Boomer, J. S.; To, K.; Chang, K. C.; Takasu, O.; Osborne, D. F.; Walton, A. H.; Bricker, T. L.; Jarman, S. D.; Kreisel, D.; Krupnick, A. S.; Srivastava, A.; Swanson, P. E.; Green, J. M.; Hotchkiss, R. S. *Jama-J. Am. Med. Assoc* **2011**, *306* (23), 2594−2605.
(20) Xiao, W. Z.; Mindrinos, M. N.; Seok, J.; Cuschieri, J.; Cuenca, A. G.; Gao, H.; Hayden, D. L.; Hennessy, L.; Moore, E. E.; Minei, J. P.; Bankey, P. E.; Johnson, J. L.; Sperry, J.; Nathens, A. B.; Billiar, T. R.; West, M. A.; Brownstein, B. H.; Mason, P. H.; Baker, H. V.; Finnerty, C. C.; Jeschke, M. G.; Lopez, M. C.; Klein, M. B.; Gamelli, R. L.; Gibran, N. S.; Arnoldo, B.; Xu, W. H.; Zhang, Y. P.; Calvano, S. E.; McDonald-Smith, G. P.; Schoenfeld, D. A.; Storey, J. D.; Cobb, J. P.; Warren, H. S.; Moldawer, L. L.; Herndon, D. N.; Lowry, S. F.; Maier, R. V.; Davis, R. W.; Tompkins, R. G. *J. Exp Med.* **2011**, *208* (13), 2581−2590.
(21) Hotchkiss, R. S.; Monneret, G.; Payen, D. *Nat. Rev. Immunol* **2013**, *13* (12), 862−874.
(22) Amodio, M.; van Dijk, D.; Srinivasan, K.; Chen, W. S.; Mohsen, H.; Moon, K. R.; Campbell, A.; Zhao, Y. J.; Wang, X. M.; Venkataswamy, M.; Desai, A.; Ravi, V.; Kumar, P.; Montgomery, R.; Wolf, G.; Krishnaswamy, S. *Nat. Methods* **2019**, *16* (11), 1139.
(23) Ma, Q.; Xu, D. *Nat. Rev. Mol. Cell Bio* **2022**, *23* (5), 303−304.
(24) Ravindra, N.; Sehanobish, A.; Pappalardo, J. L.; Hafler, D. A.; van Dijk, D. In *Disease state prediction from single-cell data using graph attention networks,* Proceedings of the ACM conference on health, inference, and learning; ACM Digital Library, 2020; pp 121−130 DOI: 10.1145/3368555.3384449.
(25) He, B.; Thomson, M.; Subramaniam, M.; Perez, R.; Ye, C. J.; Zou, J. *Pac Symp. Biocomput* **2022**, *27*, 337−348.
(26) Hu, Z. C.; Bhattacharya, S.; Butte, A. J. *Front. Immunol.* **2022**, *12*, No. 787574.
(27) Montante, S.; Brinkman, R. R. *Int. J. Lab Hematol* **2019**, *41*, 56−62.
(28) Cheng, L. J.; Karkhanis, P.; Gokbag, B.; Liu, Y. Z.; Li, L. *Plos Comput. Biol.* **2022**, *18* (4), e1008885.
(29) Yang, L. W.; Ball, A.; Liu, J.; Jain, T.; Li, Y. M.; Akhter, F.; Zhu, D. H.; Wang, J. *Nat. Commun.* **2022**, *13* (1), No. 3548.
(30) Zhang, M. Q.; Macala, K. F.; Fox-Robichaud, A.; Mendelson, A. A.; Lalu, M. M. *Shock* **2021**, *56* (2), 178−187.
(31) Toscano, M. G.; Ganea, D.; Gamero, A. M. *Jove-J. Vis. Exp.* **2011**, No. 51, No. e2860.
(32) Schuerwegh, A. J.; Stevens, W. J.; Bridts, C. H.; De Clerck, L. S. *Cytometry* **2001**, *46* (3), 172−176.
(33) Miller, M. L.; Mashayekhi, M.; Chen, L.; Zhou, P.; Liu, X.; Michelotti, M.; Tramontini Gunn, N.; Powers, S.; Zhu, X.; Evaristo, C.; Alegre, M.-L.; Molinero, L. L. *P Natl. Acad. Sci. USA* **2014**, *111* (20), 7397−7402.
(34) Diaz-Uriarte, R.; Alvarez de Andres, S. *Bmc Bioinformatics* **2006**, *7*, No. 3.
(35) Datta, S. *Stat Appl. Genet Mol.* **2008**, *7* (2), 1544−6115.
(36) Manilich, E. A.; Ozsoyoglu, Z. M.; Trubachev, V.; Radivoyevitch, T. *J. Bioinf Comput. Biol.* **2011**, *9* (2), 251−267.
(37) Pal, S.; Bi, Y.; Macyszyn, L.; Showe, L. C.; O'Rourke, D. M.; Davuluri, R. V. *Nucleic Acids Res.* **2014**, *42* (8), e64.
(38) Shilpi, A.; Kandpal, M.; Ji, Y.; Seagle, B. L.; Shahabi, S.; Davuluri, R. V. *JCO Clin Cancer Inform* **2019**, *3*, 1−9.
(39) Velotti, F.; Barchetta, I.; Cimini, F. A.; Cavallo, M. G. *Front. Immunol.* **2020**, *11*, No. 587581.
(40) Muller Kobold, A. C.; Tulleken, J. E.; Zijlstra, J. G.; Sluiter, W.; Hermans, J.; Kallenberg, C. G. M.; Cohen Tervaert, J. W. *Intens Care Med.* **2000**, *26* (7), 883−892.
(41) Beck, T. C.; Beck, K. R.; Holloway, C. B.; Hemings, R. A.; Dix, T. A.; Norris, R. A. *Front Pharmacol* **2020**, *11*, No. 1253.
(42) Garzon-Tituana, M.; Arias, M. A.; Sierra-Monzon, J. L.; Morte-Romea, E.; Santiago, L.; Ramirez-Labrada, A.; Martinez-Lostao, L.; Pano-Pardo, J. R.; Galvez, E. M.; Pardo, J. *Front Immunol* **2020**, *11*, No. 1054.
(43) Ahmed Ali, M.; Mikhael, E. S.; Abdelkader, A.; Mansour, L.; El Essawy, R.; El Sayed, R.; Eladawy, A.; Mukhtar, A. *Eur. J. Trauma Emerg S* **2018**, *44* (4), 621−626.

https://doi.org/10.1021/acs.analchem.3c01157
Anal. Chem. XXXX, XXX, XXX−XXX