analytical chemistry

Editors' Highlight

pubs.acs.org/ac

A View from Above: Cloud Plots to Visualize Global Metabolomic Data

Gary J. Patti,*[,†] Ralf Tautenhahn,[‡] Duane Rinehart,[‡] Kevin Cho,[†] Leah P. Shriver,[§,⊥] Marianne Manchester,[⊥] Igor Nikolskiy,[†] Caroline H. Johnson,[‡] Nathaniel G. Mahieu,[†] and Gary Siuzdak*[,‡]

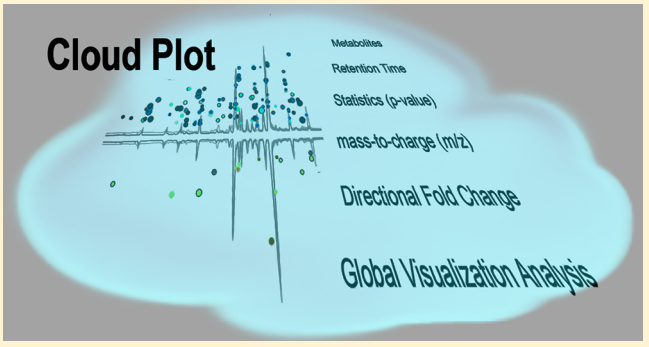
[†]Washington University, Departments of Chemistry, Genetics, and Medicine, Saint Louis, Missouri 63110, United States

[‡]The Scripps Research Institute, Departments of Chemistry, Molecular Biology, and Center for Metabolomics, La Jolla, California, United States

[§]University of Akron, Departments of Chemistry and Biology, Akron, Ohio 44325, United States

[⊥]University of California, San Diego, Skaggs School of Pharmacy and Pharmaceutical Sciences, La Jolla, California 92093, United States

ABSTRACT: Global metabolomics describes the comprehensive analysis of small molecules in a biological system without bias. With mass spectrometry-based methods, global metabolomic data sets typically comprise thousands of peaks, each of which is associated with a mass-to-charge ratio, retention time, fold change, *p*-value, and relative intensity. Although several visualization schemes have been used for metabolomic data, most commonly used representations exclude important data dimensions and therefore limit interpretation of global data sets. Given that metabolite identification through tandem mass spectrometry data acquisition is a time-limiting step of the untargeted metabolomic workflow, simultaneous visualization of these parameters from large sets of data could facilitate compound identification and data interpretation. Here, we present such a visualization scheme of global metabolomic data using a so-called "cloud plot" to represent multidimensional data from septic mice. While much attention has been dedicated to lipid compounds as potential biomarkers for sepsis, the cloud plot shows that alterations in hydrophilic metabolites may provide an early signature of the disease prior to the onset of clinical symptoms. The cloud plot is an effective representation of global mass spectrometry-based metabolomic data, and we describe how to extract it as standard output from our XCMS metabolomic software.



INTRODUCTION

Global Metabolomics and Metabolite Identification. With modern day mass spectrometers, particularly those interfaced with liquid chromatography, it is possible to detect thousands of peaks from the metabolic extract of a single biological sample.[1,2] These peaks correspond to a range of physiochemically distinct small molecules such as lipids, central carbon metabolites, sugars, amino acids, etc.[3,4] The objective of untargeted metabolomics is to comprehensively survey as many metabolites as possible and maximize the number of peaks detected to construct a global profile.[5] Although these global profiles are information rich, they are also exceedingly complex and therefore difficult to interpret intuitively.

While global metabolomic profiles consist of thousands of peaks, it is impractical with current methods to identify each feature since structural identification is a time-demanding and labor-intensive process requiring both accurate mass measurements and tandem mass spectrometry analysis. Although accurate mass measurements can be readily acquired for each peak and used to make putative metabolite assignments, accurate mass measurements alone are insufficient to structurally identify metabolites given the number of isobaric small molecules as well as the number of small molecules that cannot be resolved with the mass accuracy of many mass analyzers commonly used for metabolomics (Figure 1).[6,7] Tandem mass spectra provide structural data that are essential to increase the confidence of metabolite identifications. As supported by an analysis of the tandem mass spectra of metabolites included in the METLIN metabolite database, by matching the tandem mass spectra of research samples to the tandem mass spectra of model compounds, the number of false-positive metabolite assignments can be minimized (Figure 1).[8] Due to limitations in resources and acquisition speeds, however, it is currently impractical to acquire high-quality

Received: October 12, 2012
Accepted: December 3, 2012
Published: December 3, 2012

   dx.doi.org/10.1021/ac3029745 | *Anal. Chem.* 2013, 85, 798−804

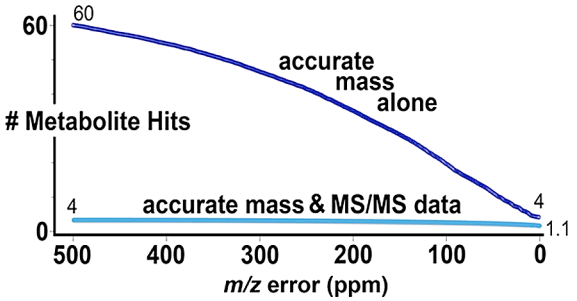


Figure 1. Number of metabolite hits from the METLIN database as a function of mass error alone and as a function of mass error combined with tandem mass spectrometry (MS/MS) data. The plot is based on 3 independent analyses using over 5000 MS^2 measurements and 40 000 accurate masses in the METLIN database.

tandem mass spectra for each of the thousands of peaks detected, and therefore, investigators typically select only a subset of compounds for structural identification. Often, the peaks are selected on the basis of statistical thresholds where peaks are chosen whose average integrated areas are changing significantly in one sample group compared to another. When the sample groups being compared are greatly different, however, the number of peaks changing can be large.[9] The latter may be especially true when comparing different cell types or when investigating a pathology that causes systemic metabolic imbalances.[10] In these instances, which peaks to structurally characterize might be ambiguous based on statistics alone and provide only a limited perspective of what is occurring in overall cellular metabolism.

This workflow for processing untargeted metabolomic data, as outlined above, limits interpretation of the results on a global scale. Accordingly, to represent the comprehensive nature of the data set, several visualizations are often used such as principal component analysis, heat maps, scatter plots, and volcano plots.[10−14] Principal component analysis (PCA) is a classical statistical technique that mathematically transforms data variables in an attempt to visually cluster sample groups; however, detailed information about individual features is not represented. Heat maps do show the relative intensity of features from each sample group, but generally, each row is individually normalized and fold changes therefore cannot be extracted. Additionally, a heat map for all of the features in a global metabolomic data set is typically too large to include in a publication or presentation. Scatter plots show the relative intensity of a feature in each sample of a pairwise comparison, from which the relative fold change can be deduced. Volcano plots are a variant of a scatter plot that also incorporate *p*-value into the representation in addition to fold change, yet volcano plots do not provide information about feature intensity and retention time. Thus, while each of these visualization techniques may provide some context-dependent insight, none of the approaches simultaneously represents all of the data parameters that are often important for the interpretation of metabolomic results. Moreover, these traditional visualizations are not effective in highlighting features that are of interest to target for further structural characterization.

Data characteristics of interest to metabolomic investigators are the *p*-value, fold change, retention time, mass-to-charge ratio, and signal intensity of features. Here, we introduce a visualization tool, called the cloud plot, which simultaneously represents all of these data characteristics. The utility of the cloud plot for interpreting untargeted metabolomic results is demonstrated for septic mice. Additionally, a method for generating a cloud plot from metabolomic data using the XCMS Online software is described.

Sepsis. Sepsis is a complex disorder with high mortality rates and is a leading cause of death in intensive care units.[15] Multiple signaling molecules have been implicated in the

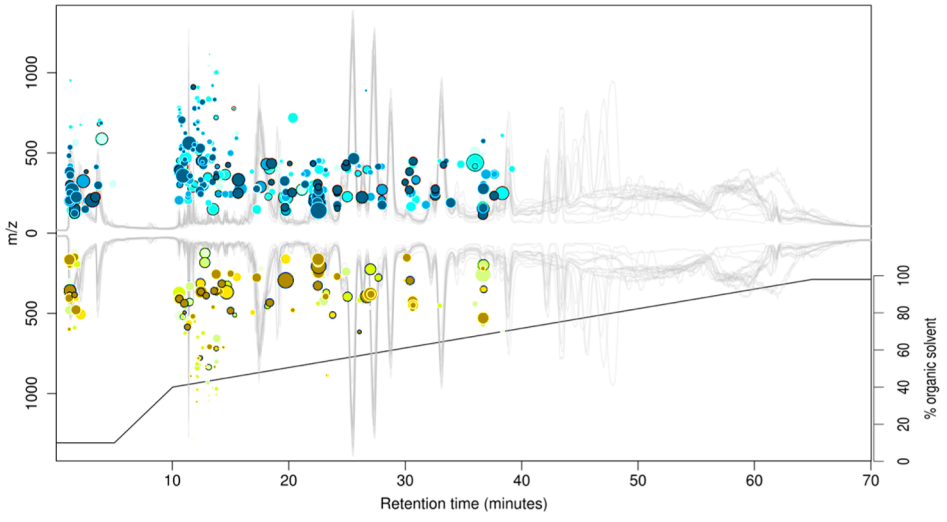


Figure 2. Cloud plot of sepsis data set, 487 features with *p*-value $\leq 1.0 \times 10^{-4}$ and fold change ≥ 3 includes visualization of the *p*-value, the directional fold change, the retention time, and the mass-to-charge ratio of features. Also shown are the total ion chromatograms for each sample and the time-dependent composition of the mobile phase. Features whose intensity is increased are shown on the top plot in blue, whereas features whose intensity is decreased are shown on the bottom plot in yellow. The size of each bubble corresponds to the log fold change of the feature: the larger the bubble, the larger the fold change. The statistical significance of the fold change, as calculated by a Welch *t* test with unequal variances, is represented by the intensity of the feature's color where features with low *p*-values are brighter compared to features with high *p*-values. The *y* coordinate of each feature corresponds to the mass-to-charge ratio of the compound as determined by mass spectrometry. Each feature is also color coded such that features that are shown with a black outline have database hits in METLIN, whereas features shown without a black outline do not have database hits.

dx.doi.org/10.1021/ac3029745 | *Anal. Chem.* 2013, 85, 798−804

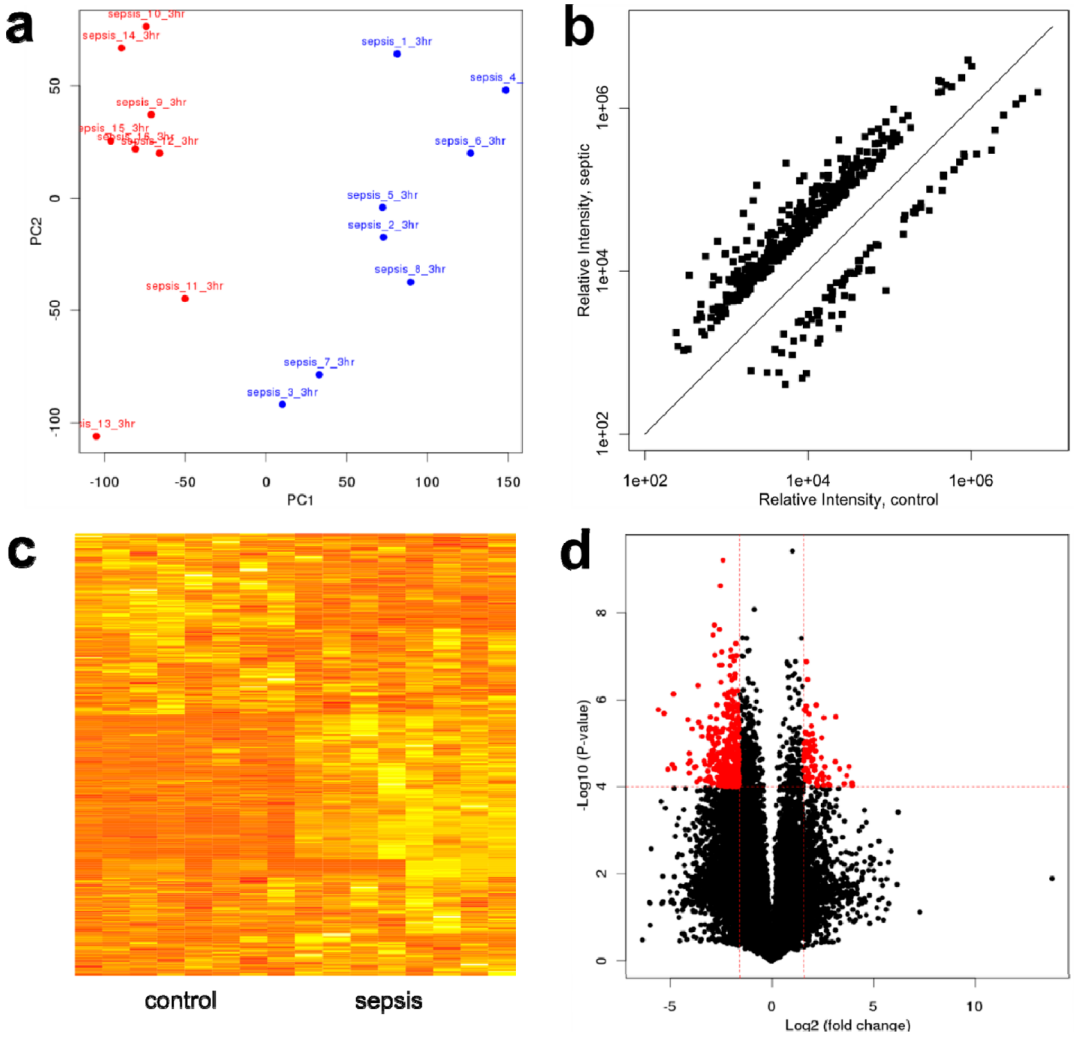


Figure 3. Other common visualizations of metabolomic data: principal component analysis (a), scatterplot (b), heatmap (c), and volcano plot (d). Here, these plots were constructed from the same metabolomic sepsis data visualized in Figure 2.

pathogenesis of sepsis. The initial stage of the condition is characterized by excessive immune activation, which leads to the release of pro-inflammatory cytokines such as tumor necrosis factor (TNF-α) and interleukin-1 (IL-1).[16] This early exaggerated pro-inflammatory response in sepsis is followed by profound immunosuppression characterized by release of anti-inflammatory cytokines such as interleukin-10 (IL-10) and transforming growth factor (TGF-β).[17] Dysregulatoin of systemic immune responses is also accompanied by activation of the hypothothalamic−pituitary−adrenal axis, which triggers the release of cortisol from the adrenal glands. Cortisol displays potent anti-inflammatory effects on macrophages, monocytes, and neutrophils, and activation of this system further suppresses innate immune responses to microbial infection.[18] Lipid mediators have also been associated with organ dysfunction and coagulation defects that occur during sepsis. Specifically, platelet activating factor (PAF) is a phospholipid-derived signaling molecule that exerts pro-inflammatory effects on immune cells such as neutrophils and platelets.[19,20] Release of PAF has been shown to increase vascular permeability in the lungs during endotoxin treatment that was dependent on the production of ceramide.[21]

The dysfunctional immune response characteristic of sepsis and septic shock are accompanied by profound alterations in cellular metabolism involving glucose, lipids, and proteins. Increased glucose utilization as well as the up-regulation of protein catabolism occurs in response to metabolic stress induced by severe infection.[22,23] Furthermore, increased lipids in blood, termed "lipemia of sepsis", results from TNF or catecholamine-stimulated lipolysis.[24,25] Taken together, these results suggest that metabolites dysregulated in response to altered energetic pathways may provide important diagnostic markers for disease prediction and patient stratification.

EXPERIMENTAL SECTION

C57BL/6 mice were maintained according to institutional animal care and use committee (IACUC) guidelines. Mice received one intraperitoneal injection of lipopolysaccharide (LPS) at a dose of 6 mg/kg. Mice were monitored for clinical symptoms, and blood was drawn for metabolomic analysis 3 h after injection.

Metabolites were extracted from blood with methanol using a defined protocol that has been shown to effectively isolate a range of metabolites, including water-soluble and lipophilic compounds.[26] First, 400 μL of cold methanol was added to 100 μL of sample and incubated at −20 °C overnight. After centrifugation at 13 000 rpm for 15 min, the supernatant was collected and dried in a SpeedVac at room temperature. The

dx.doi.org/10.1021/ac3029745 | *Anal. Chem.* 2013, 85, 798−804

Table 1. Characteristics of the Respective Visualization Tools for Metabolomic Data Sets

	Cloud Plot	Volcano Plot	Heat Map	Scatter Plot	PCA
Data analysis/global visualization tool	✓	✓	✓	✓	✓
Individual feature representation	✓	✓	✓	✓	
Statistical distrib. (e.g. p-value)	✓	✓	✓		
Directional fold change	✓	✓			
m/z	✓				
RT	✓				

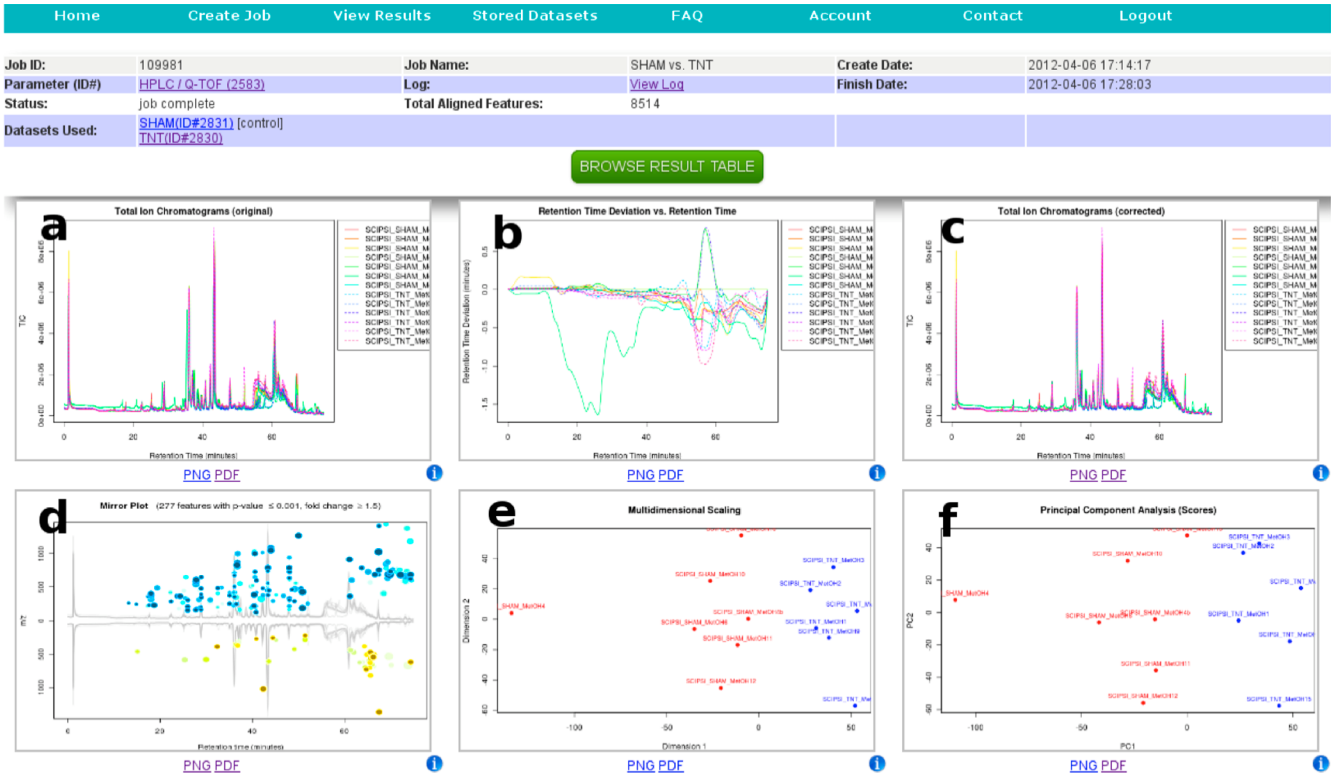


Figure 4. XCMS Online screen shot showing the overview of results from an untargeted analysis. The top display panels include text on general job information such data set names, parameter set names, and job date. The lower display panels are used to provide a global visualization of the experimental results (labeled d–f) in addition to plots that can be used for quality control (labeled a–c,e,f). The display shows the overlay of total ion chromatograms before (a) and after (c) retention time correction; the retention time correction curve (b); cloud plot (d); multidimensional scaling (e); and principal component analysis (f).

samples were then resuspended in 50 μL of 50:50 water/acetonitrile solution and transferred to liquid chromatography vials for mass spectrometry analysis.

Liquid chromatography was performed using a reversed-phase C18 column (Zorbax C18, Agilent, 5 μM, 150 × 0.5 mm diameter column) with a flow rate of 20 μL/min. Samples were analyzed in positive mode using electrospray ionization time-of-flight mass spectrometry (Agilent 6520 QTOF) with water/acetonitrile as mobile phases A/B, each containing 0.1% formic acid. The gradient consisted of the following linear changes in

dx.doi.org/10.1021/ac3029745 | *Anal. Chem.* 2013, 85, 798–804

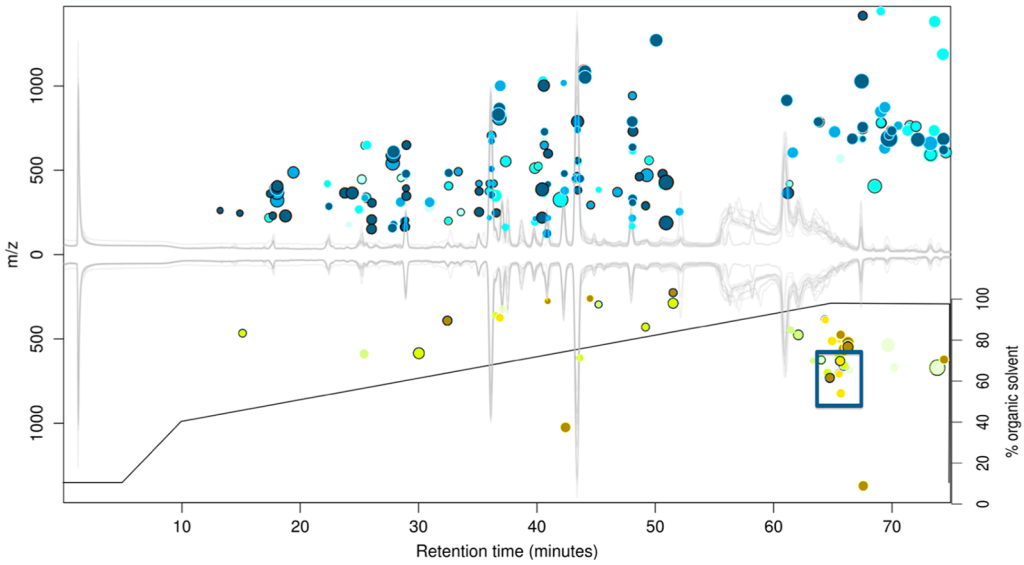


Figure 5. Cloud plot of pain data, 277 features with p-value of $\leq 1.0 \times 10^{-3}$ and fold change of ≥ 1.5. The cluster of dysregulated diacylglycerols, whose identities were confirmed by tandem mass spectrometry, is represented by the blue box.

mobile phase B composition with time: 0 min, 10% B; 5 min, 10% B; 10 min, 40% B; 65 min, 98% B; 70 min, 98% B. The order in which the samples were analyzed was randomized to reduce error due to instrument variability. Each sample analysis was followed by a wash to reduce possible carryover.

Data were analyzed using XCMS Online, which is freely available at https://xcmsonline.scripps.edu/. Metabolomic features are defined as ions with a unique m/z and retention time. The cloud plot included for the data was exported directly from XCMS Online.

RESULTS AND DISCUSSION

LPS treatment is known to activiate macrophages, and LPS injection into mice is a widely used preclinical animal model to study sepsis which was used here for investigation.[27,28] The model shares multiple similarities with human sepsis, including increased release of TNF and IL-1.[29,30] Furthermore, organ damage is recapitulated after administration of LPS and includes myocardial dysfunction in addition to lung injury.[31,32] The reproducibility of this model allows for the identification of metabolite markers associated with sepsis onset and progression.

In positive mode, we detected 29 920 features (i.e., ions with a unique m/z and retention time), 487 of which were altered between septic mice and healthy controls with fold changes greater than 3 and p-values less than 1.0×10^{-4}. Interestingly, at the 3 h time point studied, no features were found to be altered when the mobile phase consisted of more than 40% acetonitrile (Figure 2). In reversed-phase chromatography, compounds eluting at a high organic mobile phase are hydrophobic. Although thousands of features were detected in this region of the chromatogram, none of these features were found to be altered between septic mice and healthy controls with fold changes greater than 3 and p-values less than 1.0×10^{-4}.

Figure 2 provides a visualization of the data that includes the p-value, the directional fold change, the retention time, and the mass-to-charge ratio of features with fold changes greater than 3 and p-values less than 1.0×10^{-4}. The visualization also shows total ion chromatograms for each sample and the time-dependent composition of the mobile phase. Each feature with a fold change greater than 3 and a p-value less than 1.0×10^{-4} is represented as a circle. Features whose intensity is increased in septic mice relative to healthy mice are shown on the top plot in blue, whereas features whose intensity is decreased in septic mice relative to healthy mice are shown on the bottom plot in yellow. The size of each bubble corresponds to the log fold change of the feature. The statistical significance of the fold change, as calculated by a Welch t test with unequal variances, is represented by the intensity of the feature's color. That is, features with low p-values are brighter compared to features with high p-values. The y coordinate (specifically its absolute value) of each feature corresponds to the mass-to-charge ratio of the compound as determined by mass spectrometry. The x coordinate of each feature corresponds to the retention time that the compound elutes, which can be interpreted on the basis of mobile-phase composition as represented by the superimposed gradient plot. Each feature is also color coded. Features that are shown with a black outline have database hits in METLIN, whereas features shown without a black outline do not have database hits and are therefore likely to be more challenging to identify.

Sepsis, which is a leading cause of death in critically ill patients, involves an extremely complex chain of physiological events. Although much effort has recently focused on the role of lipid molecules as drivers of the pathogenesis of the syndrome and as potential biomarkers, our understanding of sepsis remains incomplete and available biomarkers have insufficient specificity and/or sensitivity to be used routinely in clinical practice. Here, we compared the metabolic profile of mice 3 h after they were challenged with LPS to healthy controls. At the 3 h time point after LPS treatment, the mice have yet to display clinical symptoms of sepsis. At this time point, however, we did not identify any lipids to be altered with statistical significance. Instead, as shown in Figure 2, we found many water-soluble metabolites to be dysregulated.

These data highlight the value of the cloud plot for visualizing untargeted metabolomic results. Other tools traditionally used to visualize metabolomic data do not similarly emphasize the potential importance of water-soluble com-

dx.doi.org/10.1021/ac3029745 | Anal. Chem. 2013, 85, 798−804

pounds in the condition (Figure 3). While PCA shows that the sample groups can be separated on the basis of the metabolomic results, the plot does not offer biochemical insight into the specific molecules involved. A heatmap provides improved biochemical specificity but is not effective at visualizing large data sets as shown here. A scatter plot and a volcano plot provide the most descriptive global visualizations but still exclude important data parameters such as m/z value and retention time from which lipophilicity can be deduced. The cloud plot includes these data while also showing the total ion chromatogram, the chromatographic gradient, and a representation of features that have database hits in METLIN. Further, in future plots, the color of the circles could be used to encode alternate information such as pathway inclusion or associated expression data. A comparison of the characteristics of these different plotting strategies is shown in Table 1. The cloud plot shown in Figure 2 was exported directly from XCMS Online. XCMS Online is a recently released web-based version of the widely used XCMS software that is freely available at https://xcmsonline.scripps.edu.[33] This software enables users to upload metabolomic data for processing with minimal training. The software performs feature detection, retention time correction, alignment, annotation, and statistical analysis and provides a cloud plot of the results for visualization (Figure 4). Using XCMS Online, the results can be browsed interactively within the cloud plot. That is, when users scroll their mouse over features with METLIN database hits, these putative assignments are displayed in a pop-up window. Cloud plots can be downloaded directly from XCMS Online as zip files for incorporation into publications, presentations, and grants.

Cloud plots also have the potential to facilitate the structural identification of some features. Some classes of metabolites, for example, will lie in characteristic zones of the cloud plot. The "clustering" of features into zones results from these classes of compounds having similar retention times and m/z values. As an example, we show a cloud plot from data we recently published.[34] Here, we compared spinal cord tissue from rats suffering from neuropathic pain to spinal cord tissue from control rats. As shown on the plot, we identified 8 diacylglycerols that were down-regulated in animals suffering from neuropathic pain (the identities of these compounds were verified with tandem mass spectrometry). These identified diacylglycerols were localized to a unique region of the cloud plot, which could prove useful for identifying diacylglycerols in future studies using the same C18 chromatography and gradient (Figure 5).

CONCLUSION

Cloud plots provide an effective visualization tool for global metabolomic data by representing the p-value, the directional fold change, the retention time, and the mass-to-charge ratio of features with fold changes and p-values within a defined threshold. Additionally, cloud plots show total ion chromatograms, chromatographic gradients, and features that have hits in the METLIN metabolite database (Table 1). Unlike other visualization tools traditionally used, cloud plots were designed specifically to represent global metabolomic data sets in a more intuitive scheme to facilitate interpretation and selection of features for structural characterization. Here, we have used a cloud plot to represent untargeted metabolomic results from septic mice compared to healthy controls. Despite the attention that has been dedicated to lipids, the data suggest that water-soluble metabolites are important in the early stages of sepsis pathogenesis and may serve as diagnostic markers. Cloud plots can be generated using our freely available, web-based XCMS Online software. The plots can be easily downloaded and used in publications, presentations, and grants.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gjpattij@wustl.edu (G.J.P.); siuzdak@scripps.edu (G.S.).

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779−787.
(2) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. *Anal. Chem.* **2011**, *83*, 2152−2161.
(3) Forsythe, I. J.; Wishart, D. S. Exploring Human Metabolites Using the Human Metabolome Database. *Current Protocols in Bioinformatics*; John Wiley and Sons, Inc.: Hoboken, New Jersey, 2009; 25:14.8.1−14.8.45.
(4) Patti, G. J.; Yanes, O.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 263−269.
(5) Baker, M. *Nat. Methods* **2011**, *8*, 117−121.
(6) Vinayavekhin, N.; Saghatelian, A. Untargeted Metabolomics. *Current Protocols in Molecular Biology*; John Wiley and Sons, Inc.: Hoboken, New Jersey, 2010, 90:30.1.1−30.1.24.
(7) Wang, Z.; Klipfell, E.; Bennett, B. J.; Koeth, R.; et al. *Nature* **2011**, *472*, 57−63.
(8) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; et al. *Nat. Biotechnol.* **2012**, *30*, 826−828.
(9) Weisenberg, S. A.; Butterfield, T. R.; Fischer, S. M.; Rhee, K. Y. *J. Sep. Sci.* **2009**, *32*, 2262−2265.
(10) Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; et al. *Nature* **2009**, *457*, 910−914.
(11) Krumsiek, J.; Suhre, K.; Illig, T.; Adamski, J.; Theis, F. J. *J. Proteome Res.* **2012**, *11*, 4120−4131.
(12) Griffin, J. L. *Philos. Trans. R. Soc., London B Biol. Sci.* **2006**, *361*, 147−161.
(13) Johansen, K. K.; Wang, L.; Aasly, J. O.; White, L. R.; et al. *PLoS One* **2009**, *4*, No. e7551.
(14) Bictash, M.; Ebbels, T. M.; Chan, Q.; Loo, R. L.; et al. *J. Clin. Epidemiol.* **2010**, *63*, 970−979.
(15) Gullo, A.; Foti, A.; Murabito, P.; Li Volti, G.; et al. *Front. Biosci. (Elite Ed.)* **2010**, *2*, 906−911.
(16) Block, M. I.; Berg, M.; McNamara, M. J.; Norton, J. A.; et al. *J. Exp. Med.* **1993**, *178*, 1085−1090.
(17) Bosmann, M.; Ward, P. A. *Trends Immunol.* **2012**, DOI: 10.1016/j.it.2012.09.004.
(18) Annane, D. *Curr. Pharm. Des.* **2008**, *14*, 1882−1886.
(19) Yost, C. C.; Cody, M. J.; Harris, E. S.; Thornton, N. L.; et al. *Blood* **2009**, *113*, 6419−6427.
(20) Lindemann, S.; Tolley, N. D.; Dixon, D. A.; McIntyre, T. M.; et al. *J. Cell Biol.* **2001**, *154*, 485−490.
(21) Goggel, R.; Winoto-Morbach, S.; Vielhaber, G.; Imai, Y.; et al. *Nat. Med.* **2004**, *10*, 155−160.
(22) Lang, C. H.; Dobrescu, C. *Endocrinology* **1991**, *128*, 645−653.
(23) Combaret, L.; Tilignac, T.; Claustre, A.; Voisin, L.; et al. *Biochem. J.* **2002**, *361*, 185−192.
(24) Banerjee, S.; Bhaduri, J. N. *Proc. Soc. Exp. Biol. Med.* **1959**, *101*, 340−341.
(25) Van der Poll, T.; Romijn, J. A.; Endert, E.; Borm, J. J.; et al. *Am. J. Physiol.* **1991**, *261*, E457−E465.
(26) Yanes, O.; Tautenhahn, R.; Patti, G. J.; Siuzdak, G. *Anal. Chem.* **2011**, *83*, 2152−2161.

(27) Bordbar, A.; Mo, M. L.; Nakayasu, E. S.; Schrimpe-Rutledge, A. C.; et al. *Mol. Syst. Biol.* **2012**, *8*, 558.
(28) Mofarrahi, M.; Sigala, I.; Guo, Y.; Godin, R.; et al. *PLoS One* **2012**, *7*, No. e47265.
(29) Remick, D. G.; Strieter, R. M.; Lynch, J. P., 3rd; Nguyen, D.; et al. *Lab. Invest.* **1989**, *60*, 766−771.
(30) Chensue, S. W.; Terebuh, P. D.; Remick, D. G.; Scales, W. E.; Kunkel, S. L. *Am. J. Pathol.* **1991**, *138*, 395−402.
(31) Balija, T. M.; Lowry, S. F. *Curr. Opin. Infect. Dis.* **2011**, *24*, 248−253.
(32) Sun, Q.; Chen, L.; Gao, M.; Jiang, W.; et al. *Int. Immunopharmacol.* **2012**, *12*, 88−93.
(33) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.* **2012**, *84*, 5035−5039.
(34) Patti, G. J.; Yanes, O.; Shriver, L. P.; Courade, J. P.; et al. *Nat. Chem. Biol.* **2012**, *8*, 232−234.

dx.doi.org/10.1021/ac3029745 | *Anal. Chem.* 2013, 85, 798−804