Research Techniques Made Simple: Mass Cytometry Analysis Tools for Decrypting the Complexity of Biological Systems

Tiago R. Matos[1,2,3], Hongye Liu[1,2] and Jerome Ritz[1,2]

Mass cytometry by time-of-flight experiments allow analysis of over 40 functional and phenotypic cellular markers simultaneously at the single-cell level. The data dimensionality escalation accentuates limitations, inherent to manual analysis, as being subjective, labor-intensive, slow, and often incapable of showing the detailed features of each unique cell within populations. The subsequent challenge of examining, visualizing, and presenting mass cytometry data has motivated continuous development of dimensionality reduction methods. As a result, an increasing recognition of the inherent diversity and complexity of cellular networks is emerging, with the discovery of unexpected cell subpopulations, hierarchies, and developmental pathways, such as those existing within the immune system. Here, we briefly review some frequently used and accessible mass cytometry data analysis tools, including principal component analysis (PCA); spanning-tree progression analysis of density-normalized events (SPADE); t-distributed stochastic neighbor embedding (t-SNE)–based visualization (viSNE); automatic classification of cellular expression by nonlinear stochastic embedding (ACCENSE); and cluster identification, characterization, and regression (CITRUS). Mass cytometry, used together with these innovative analytic tools, has the power to lead to key discoveries in investigative dermatology, including but not limited to identifying signaling phenotypes with predictive value for early diagnosis, prognosis, or relapse and a thorough characterization of intratumor heterogeneity and disease-resistant cell populations, that may ultimately unveil novel therapeutic approaches.

Journal of Investigative Dermatology (2017) **137,** e43–e51; doi:10.1016/j.jid.2017.03.002

CME Activity Dates: 12 April 2017
Expiration Date: 11 April 2018
Estimated Time to Complete: 1 hour

Planning Committee/Speaker Disclosure: All authors, planning committee members, CME committee members and staff involved with this activity as content validation reviewers have no financial relationship(s) with commercial interests to disclose relative to the content of this CME activity.

Commercial Support Acknowledgment: This CME activity is supported by an educational grant from Lilly USA, LLC.

Description: This article, designed for dermatologists, residents, fellows, and related healthcare providers, seeks to reduce the growing divide between dermatology clinical practice and the basic science/current research methodologies on which many diagnostic and therapeutic advances are built.

Objectives: At the conclusion of this activity, learners should be better able to:
- Recognize the newest techniques in biomedical research.
- Describe how these techniques can be utilized and their limitations.
- Describe the potential impact of these techniques.

CME Accreditation and Credit Designation: This activity has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education through the joint providership of William Beaumont Hospital and the Society for Investigative Dermatology. William Beaumont Hospital is accredited by the ACCME to provide continuing medical education for physicians.
William Beaumont Hospital designates this enduring material for a maximum of 1.0 *AMA PRA Category 1 Credit(s)*™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Method of Physician Participation in Learning Process: The content can be read from the Journal of Investigative Dermatology website: http://www.jidonline.org/current. Tests for CME credits may only be submitted online at https://beaumont.cloud-cme.com/RTMS-May17 – click 'CME on Demand' and locate the article to complete the test. Fax or other copies will not be accepted. To receive credits, learners must review the CME accreditation information; view the entire article, complete the post-test with a minimum performance level of 60%; and complete the online evaluation form in order to claim CME credit. The CME credit code for this activity is: 21310. For questions about CME credit email cme@beaumont.edu.

[1]*Division of Hematologic Malignancies, Dana-Farber Cancer Institute, Boston, Massachusetts, USA;* [2]*Harvard Medical School, Boston, Massachusetts, USA; and* [3]*Academic Medical Center, Department of Dermatology, University of Amsterdam, Amsterdam, The Netherlands*

Correspondence: Tiago R. Matos, Department of Dermatology, Room L3-119, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. E-mail: tiagoreismatos@gmail.com

Abbreviations: ACCENSE, automatic classification of cellular expression by nonlinear stochastic embedding; CITRUS, cluster identification, characterization, and regression; CyTOF, mass cytometry by time-of-flight mass spectrometry; FCS, Flow Cytometry Standard; PCA, principal component analysis; SPADE, spanning-tree progression analysis of density-normalized events; t-SNE, t-distributed stochastic neighbor embedding; viSNE, t-distributed stochastic neighbor embedding–based visualization

© 2017 The Authors. Published by Elsevier, Inc. on behalf of the Society for Investigative Dermatology.

SUMMARY POINTS

- New methods are being continuously developed to analyze and best represent multidimensional, complex CyTOF data.
- Principal component analysis (PCA) provides a visualization of the data in three-dimensional space and identifies the parameters with the most variance among the dataset.
- Spanning-tree progression analysis of density-normalized events (SPADE) clusters cells into a minimum-spanning hierarchical tree for two-dimensional visualization.
- In t-distributed stochastic neighbor embedding (t-SNE)–based visualization (viSNE) and automatic classification of cellular expression by nonlinear stochastic embedding (ACCENSE), each single cell data point has a unique location in a two-dimensional representation, reflecting the cells' immunophenotypic similarity or differences in high-dimensional space.
- Cluster identification, characterization, and regression (CITRUS) identifies cellular features that correlate to an experimental endpoint of interest.

INTRODUCTION

New methods are being developed to examine, visualize, and present the multidimensional complexity of cellular function and identity and the role of individual cells within biological systems. Mass cytometry by time-of-flight mass spectrometry (CyTOF)[1] currently has the capacity to allow investigation of 40 or more distinct parameters at the single-cell level (Figure 1). Although the technique has not yet been widely adopted within the field of investigative dermatology, it has potential to, for example, allow identification of cell signals for early diagnosis in cutaneous T-cell lymphoma, allow early detection or predict relapse in psoriasis and atopic dermatitis, and allow thorough characterization of drug-resistant cell populations in skin cancer, eventually unveiling new therapies. The large amount of data generated and potential of the technique to delineate rare cell subsets has driven the need to develop dimensionality reduction methods and analysis algorithms to best analyze and represent mass cytometry data. A significant limitation of traditional data clustering methods through biaxial plots and histograms, such as has been used to represent traditional flow cytometry data, is that pre-existing knowledge of the defining markers of each population is required. This limits the ability of researchers to discover unexpected cellular subsets and does not allow examination of system-level phenotypic diversity. Furthermore, manual analysis of individual markers and combinations of markers is a subjective, slow, and labor-intensive process, which results in a significant scalability restriction and can introduce several inherent biases. Although CyTOF technology and experimental methodology have been described in detail in previous reviews (Doan et al., 2015; Matos et al., 2017), comprehensive understanding is also required with respect to the tools available for analysis of high-dimensional datasets to make meaningful use of the results. In this short review, we focus on some of the most commonly used and accessible novel CyTOF data analysis tools, including principal component analysis (PCA), spanning-tree progression analysis of density-normalized events (SPADE), t-distributed stochastic neighbor embedding (t-SNE)–based visualization (viSNE), automatic classification of cellular expression by nonlinear stochastic embedding (ACCENSE), and cluster identification, characterization, and regression (CITRUS).

[1] The abbreviation "CyTOF", in addition to being the name of this technique, is also the name of a commercial product that enables researchers to use the method. The authors are in no way endorsing any specific commercial products.

DIMENSION REDUCTION AND VISUALIZATION ALGORITHMS

PCA

PCA is a well-established and widely used tool for visualizing multidimensional data that was adopted to analyze large mass cytometry datasets (Bendall et al., 2011; Jackson, 1991; Newell et al., 2012). PCA identifies those parameters among a certain dataset that present the most variance by generating linear combinations from a large list of parameters into new compound variables (principal components). As a result, the quotient of the relative variation of each principle component over the total variance gives an idea of the effectiveness of each component in separating out data points. In addition, PCA results in models that can be used to project new data points in linear time. For example, it allows graphical visualization of the expression intensity of several functional markers (y-axis) throughout the cell differentiation process (x-axis) (Figure 2). PCA also allows visualization of the data in three-dimensional space, often prominently displaying the first three data components of maximal variance. However, this feature can also be a limitation, because it may mask noteworthy biological differences that are more subtle variances in the data. Another constraint is the inherent assumption that the given data are parametric. PCA also represents the data through linear projections, which may not be representative of the inherent structure of the original data. To overcome this constraint, nonlinear methods such as t-SNE (described in following sections) were developed for high-dimensional data analysis. Newell at al. (2012) used PCA to represent simultaneously 25 markers from a single cell sample, hence quantifying the expression of functional markers among several $CD8^+$ T-cell subsets. This representation method displayed a greater phenotypic and functional complexity among $CD8^+$ T cells than previously appreciated (Figure 2). The holistic study of many functional and phenotypic markers and their expression levels through several differentiation subsets would not be possible by conventional manual analysis. This study also observed that subsets that develop in response to different viruses have distinct combinatorial patterns of cytokine expression, showing the remarkable

Download English Version:

https://daneshyari.com/en/article/5649486

Download Persian Version:

https://daneshyari.com/article/5649486

Daneshyari.com