Clinical Immunology (2015) 157, 249–260

available at www.sciencedirect.com

Clinical Immunology

www.elsevier.com/locate/yclim

Automated flow cytometric analysis across large numbers of samples and cell types

Xiaoyi Chen [a,b], Milena Hasan [c], Valentina Libri [c], Alejandra Urrutia [c,d,e], Benoît Beitz [c], Vincent Rouilly [f], Darragh Duffy [c,d,e], Étienne Patin [i], Bernard Chalmond [g,h], Lars Rogge [c,i], Lluis Quintana-Murci [j,k], Matthew L. Albert [c,d,e,l,*], Benno Schwikowski [a,**]
for the Milieu Intérieur Consortium

[a] Systems Biology Lab, Institut Pasteur, Paris, France
[b] Laboratory of Analysis Geometry and Modeling, Department of Mathematics, University of Cergy-Pontoise, Ile de France, France
[c] Center for Human Immunology, Institut Pasteur, Paris France
[d] INSERM U818, France
[e] Laboratory of Dendritic Cell Immunobiology, Department of Immunology, Institut Pasteur, Paris France
[f] Center for Bioinformatics, Institut Pasteur, Paris France
[g] University of Cergy-Pontoise, France
[h] CLMA, ENS-Cachan, France
[i] Laboratory of Immunoregulation, Department of Immunology, Institut Pasteur, Paris France
[j] Unit of Human Evolutionary Genetics, Department of Genomes & Genetics, Institut Pasteur, Paris, France
[k] CNRS URA3012, France
[l] INSERM UMS20, France

Received 4 July 2014; accepted with revision 20 December 2014
Available online 7 January 2015

KEYWORDS
Flow cytometry;
Multidimensional analysis;
Population-based cohort;
Automation;
Standardization;
Algorithms;

Abstract Multi-parametric flow cytometry is a key technology for characterization of immune cell phenotypes. However, robust high-dimensional post-analytic strategies for automated data analysis in large numbers of donors are still lacking. Here, we report a computational pipeline, called FlowGM, which minimizes operator input, is insensitive to compensation settings, and can be adapted to different analytic panels. A Gaussian Mixture Model (GMM)-based approach was utilized for initial clustering, with the number of clusters determined using Bayesian Information Criterion. Meta-clustering in a reference donor permitted automated identification of 24 cell types across four panels. Cluster labels were integrated into FCS files, thus permitting

Abbreviations: BIC, Bayesian Information Criterion; CV, coefficient of variation; DC, dendritic cell; EM, Expectation Maximization; FSC, forward scatter; GMM, Gaussian Mixture Model; MFI, mean fluorescent intensity; SSC, side scatter

* Correspondence to: M. L. Albert, Unit of Dendritic Cell Immunobiology, Inserm U818, Institut Pasteur, 25, Rue du Dr. Roux, 75724 Paris Cedex 15, France. Fax: +33 1 45 68 85 48.
** Correspondence to: B. Schwikowski, Systems Biology Lab, Institut Pasteur, 25, rue du Dr. Roux, 75724 Paris Cedex 15, France. Fax: +33 1 40 61 37 01.
E-mail addresses: albertm@pasteur.fr (M.L. Albert), benno@pasteur.fr (B. Schwikowski).

http://dx.doi.org/10.1016/j.clim.2014.12.009
1521-6616/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

comparisons to manual gating. Cell numbers and coefficient of variation (CV) were similar between FlowGM and conventional gating for lymphocyte populations, but notably FlowGM provided improved discrimination of "hard-to-gate" monocyte and dendritic cell (DC) subsets. FlowGM thus provides rapid high-dimensional analysis of cell phenotypes and is amenable to cohort studies.
© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Flow cytometry is a key technology for the characterization of the cellular component of the immune system. Flow cytometers are able to simultaneously quantify different surface markers of single cells, allowing the identification and quantification of different immune cell subpopulations. In recent years, improvements in measurement speed and experimental automation have enabled comprehensive immunoprofiling of larger cohorts [1].

The gold standard for the analysis of raw flow cytometry data has until now remained "hand gating" (i.e., analysis through computer-assisted procedures for the classification of cells into single cell types using software tools such as FlowJo [2]). Each sample is analyzed by successively separating cell types by successive "gating" in a series of one- or two-dimensional projections. However, the manual operation is laborious and subject to biased visual inspection and gate adjustment. These concerns grow with increased numbers of measured phenotypic markers. Moreover, there is a major limitation in that information critical for accurate gating may not be present in the selected two-dimensional projections.

Here, we report a new method for analyzing multi-parametric flow cytometry, the need for which was motivated by the Milieu Intérieur study. This project aims at defining the genetic and environmental determinants of variable immunologic phenotypes in a healthy population [Thomas et al., co-submission]. Cell phenotyping constitutes one of the major data sets to be integrated into the data warehouse, and as such efforts were made to standardize each step of the sample collection, technical procedures and data analysis. A Companion paper highlights the pre-analytical semi-automated measures put in place for labeling and data generation [Hasan et al., co-submission]. This manuscript details the automated analytic workflow developed for the identification and analysis of 24 cell types across four 8-color cytometry panels.

Our work follows from a large number of computational approaches that have been developed for automated flow cytometry analysis. Recently, the FlowCAP study evaluated a range of approaches [13]. In all cases, however, the datasets used by these investigators were of a smaller scale than the ones in our study, in terms of samples studied (FlowCAP: up to 30 samples; here: 115 samples × 4 panels), and the number of events per experiment (FlowCAP: up to approximately 100,000 events; here: on average 300,000 events per FCS file). Due to these differences, we found that top-ranked FlowCAP approaches were inadequate to address the needs of our data sets. For example, the ADICyt approach [4] required more than 6 h for the analysis of a single sample. The FlowMeans software [5] was faster, but required manual assignment of cell types to each cluster in every single sample. The recent X-Cyt approach [3] was designed explicitly to efficiently address the problem of larger numbers of samples. However, X-Cyt still requires the definition of a "partitioning scheme", a series of mixture models whose sequence and parameters have to be manually configured and calibrated for each cell type of interest in any given analytic panel.

To support the analysis of the Milieu Intérieur cohort dataset, we developed a novel high-dimensional data analysis approach, which we refer to as FlowGM, utilizing fast algorithms that enable the standardized analysis of large numbers of samples. We describe its application to two representative 8-color panels with up to 11 cell populations classified per panel. Its principal feature is that, after the definition of global parameters in a reference sample (i.e., a one-time manual assignment of cell type labels to clusters), it is possible to automatically position and identify cell populations across the entire dataset. This approach will enable analysis of our large healthy donor cohort.

2. Materials and methods

2.1. Dataset

Four 8-color cytometry panels targeting major leukocyte populations across 115 individuals from different age groups and genders were designed to characterize the major immune cell populations (T cells, B cells, NK cells and monocytes), as well as subpopulations of T cells, dendritic cells (DC) and polymorphonuclear leukocytes (PMN). The standardized procedure of collection and treatment of the whole blood sample is described in [Hasan et al., co-submission]. For each of the four panels, technical replicates performed by five parallel blood samples obtained from three donors ("repeatability" studies from [Hasan et al., co-submission]) were generated to examine robustness of the experimental and computational protocols.

2.2. FlowGM cluster model

The input to FlowGM is a set of m sets of n quantitative measurements ("events"), formally, m n-dimensional vectors. Clustering is based on a multivariate Gaussian Mixture Model (GMM) [6], which has the form

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{j=1}^{k} \alpha_j \, \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_j, \Sigma_j)$$

A GMM thus corresponds to a set of k clusters, each described by a cluster weight α_j and an n-dimensional

Download English Version:

https://daneshyari.com/en/article/6087526

Download Persian Version:

https://daneshyari.com/article/6087526

Daneshyari.com