Cytometry Part A 71A:334–344 (2007)

Subject Classification Obtained by Cluster Analysis and Principal Component Analysis Applied to Flow Cytometric Data

Enrico Lugli,[1] Marcello Pinti,[1] Milena Nasi,[1] Leonarda Troiano,[1] Roberta Ferraresi,[1] Chiara Mussi,[2] Gianfranco Salvioli,[2] Valeri Patsekin,[3] J. Paul Robinson,[3] Caterina Durante,[4] Marina Cocchi,[4] and Andrea Cossarizza[1]*

[1]Department of Biomedical Sciences, Chair of Immunology, University of Modena and Reggio Emilia, via Campi 287, 41100 Modena, Italy
[2]Chair of Geriatrics and Gerontology, Nuovo Ospedale S. Agostino-Estense (NOCSE) a Baggiovara, University of Modena and Reggio Emilia, Modena, Italy
[3]Purdue University Cytometry Laboratories, Purdue University, West Lafayette, Indiana, 47907
[4]Department of Chemistry, University of Modena and Reggio Emilia, via Campi 183, 41100 Modena, Italy

Received 25 August 2006; Revision received 3 December 2006; Accepted 7 January 2007

Background: Polychromatic flow cytometry (PFC) allows the simultaneous determination of multiple antigens in the same cell, resulting in the generation of a high number of subsets. As a consequence, data analysis is the main difficulty with this technology. Here we show the use of cluster analysis (CA) and principal component analyses (PCA) to simplify multicolor data visualization and to allow subjects' classification.
Methods: By eight-colour cytofluorimetric analysis, we investigated the T cell compartment in donors of different age (young, middle-aged, and centenarians). T cell subsets were identified by combining positive and negative expression of antigens. The resulting data set was organized into a matrix and subjected to CA and PCA.
Results: CA clustered people of different ages on the basis of cytofluorimetric profile. PCA of the cellular subsets identified centenarians within a different cluster from young donors, while middle-aged donors were scattered between these groups. These approaches identified T cell phenotypes that changed with increasing age. In young donors, memory T cell subsets tended to be CD127+ and CD95− whereas CD127−, CD95+ phenotypes were found at higher frequencies in people with advanced age.
Conclusions: Our data suggest the use of bioinformatic approaches to analyze large data-sets generated by PFC and to obtain the rapid identification of key populations that best characterize a group of subjects. © 2007 International Society for Analytical Cytology

Key terms: polychromatic flow cytometry; data analysis; cluster analysis; principal component analysis; subject classification

Understanding the fine characteristics of the immune system is important for developing new strategies regarding the generation of immune-based therapies and production of vaccines. Many tools are now available that allow the analysis of thousands of parameters even in a relatively small cell population, such as microarrays or proteomic techniques. Flow cytometry has the same potential, since it allows the simultaneous detection of several surface or intracellular antigens in the same cell (1,2). This powerful technology has been used for more than 30 years to analyze a variety of immunological parameters, as well as functional activities of cells under investigation (3,4). Currently, thanks to upcoming technologies and to the availability of new, bright fluorochromes such as tandem conjugates and quantum dots, it is possible to stain cells with up to 17 colors simultaneously (5). Theoretically, considering only the antigens that are clearly expressed or not in a given cell, such a large number of fluorescence signals would allow for the identification of 2^{17} different cell populations in a few microliter of blood. Clearly, the analysis of results deriving from such a large number of antigen combinations is not feasible with classical approaches, i.e. those that analyze the combinations of a maximum of three fluorochromes per event on the x, y, and z axes.

Grant sponsors: MIUR Cofin 2004, ISS (Programma Nazionale di Ricerca sull'AIDS 2006).
*Correspondence to: Andrea Cossarizza, MD PhD, Department of Biomedical Sciences, Section of General Pathology, via Campi, 287, 41100 Modena, Italy.
E-mail: cossarizza.andrea@unimore.it
Published online 12 March 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/cyto.a.20387

To overcome this problem, and have the possibility to fully appreciate the importance of each single cell subset that can be recognized by using polychromatic flow cytometry, we have identified two new approaches based upon CA and PCA. By multiple color flow cytometry, among CD4+ or CD8+ T lymphocytes we have analyzed the expression of markers associated with T cell differentiation (CD45RA and CCR7) (6), with cell survival (CD127) (7) or with activation (CD38 and CD95) (8) in subjects of different age, including people such as centenarians that approached the maximum human lifespan.

Through combination of positive, dim, and negative antigen expression, we generated all the possible subpopulations which were indicated as percentages of CD3+, CD4+ or CD3+, CD8+ T cells and which were subjected to hierarchical clustering and PCA. These approaches allowed us to identify sophisticated characteristics of T cell dynamics during aging of the immune system. In particular, CD4+ and CD8+ T cell flow cytometric profile based on the markers analyzed was sufficient to cluster subjects of different ages. Moreover, PCA of the obtained large data set allowed us to report phenotype-to-subject associations. Finally, a few key populations of cells were identified to be responsible for individual differences among subjects and the dynamics of these few subsets was sufficient to explain more than 93% of the total variance of the entire data set.

MATERIALS AND METHODS

Subjects and Blood Samples

We analyzed a total of 24 subjects: young (n = 9, four males and five females, mean age = 21.3 years, range 20–25), middle-age (n = 8, three males and five females, mean age: 60.0 years, range: 58–62), and centenarians (n = 7, one male and six females). All young and middle-aged donors were selected according to the criteria described in the SENIEUR protocol (9). Centenarians were defined as individuals older than 99 years (10). All participating subjects did not display relevant acute or chronic disease affecting the immune system and gave their informed consent, according to Italian laws.

Polychromatic Analysis of Peripheral Blood Mononuclear Cells

Peripheral blood mononuclear cells (PBMCs) were separated from freshly collected blood according to standard procedures; after PBMCs were extensively washed with Hanks' balanced salt solution, 1×10^6 cells were resuspended in 100 μl PBS and stained with the following antibodies (1 μg each): CD45RA FITC, CCR7 PE-Cy7, and CD8 APC-Cy7 (BD Pharmingen, Franklin Lakes, NJ), CD127 PE and CD4 APC (R&D Systems, Minneapolis, MN), CD3 PE-Texas-Red (Caltag, Burlingame, CA), CD95 PE-Cy5 and CD38 APC-Cy5.5 (e-Bioscience, San Diego, CA). Cells were incubated for 30 min at 4°C, then washed with PBS as above, and resuspended in PBS for flow cytometric analysis. At least 150,000 events were acquired for each sample. Cells were analyzed using a 16-parameter Cyflow ML (Partec GmbH, Münster, Germany), equipped with a solid-state blue laser (emitting at 488 nm, 200 mW, kept at 50 mW, for detection of FITC, PE, PE-Texas Red, PE-Cy5, PE-Cy7), a red diode laser (635 nm, 25 mW, for detection of APC, APC-Cy5.5, APC-Cy7), a UV mercury lamp HBO (100 long life, 100 W, not used in this study), a solid-state green laser (532 nm, 50 mW, not used in this study), and a CCD camera. Fluorescence compensation was performed using single-color staining controls. Data were analyzed by Partec Flomax 3.0 for Windows and FlowJo 6.3 for MacOSX.

One Way Analysis of Variance

Mean differences of antigen expression in T_N, T_{CM}, T_{EM}, and T_{EMRA} CD4+ and CD8+ T cells were computed by unpaired one way analysis of variance (for repeated measures when required) with Dunn's post-test under Graph Pad Prism 3.03 or JMP 5.1, both under Windows XP. Differences were considered significant when $P < 0.05$.

Generation of Data Sets, Cluster Analysis, and Principal Component Analysis of CD4+ and CD8+ Subpopulations

Each cellular subpopulation obtained by the combination of cell surface antigen expression was indicated as the percentage of total CD3+, CD4+ or CD3+, CD8+ T cells (each of these main subsets was considered as 100%) and was inserted in a matrix whose rows correspond to T cell subsets and columns to the subjects we studied. Subsequently, CA (11) and PCA (12) were performed. CD4+ and CD8+ T cell subsets were analyzed as two separated data sets.

In the case of CA, the percentages were normalized to the mean of values derived from young subjects; CA (using "complete linkage" and "correlation similarity" as parameters) was performed with "Cluster" software (11) and then visualized with "Treeview" software (11) (both freely available at http://rana.lbl.gov/EisenSoftware.htm). All those cell populations that were absent in all subjects were automatically eliminated from the analysis.

PCA, performed with the software PLS Toolbox 3.5 (Eigenvector Research, Wenatchee, WA), was used to capture the relevant information and visualize major trends and structure inherent in the data. PCA is an unsupervised dimension-reduction method that generates a new set of decorrelated variables (called principal components) as linear combinations of the original variables (in our case, represented by T cell subsets). The majority of the variation of flow cytometric datasets (subjects) can be captured by the most dominant principal components. An additional advantage of expressing the data in terms of the leading principal components is their robustness to noise. For each data set, PCA was carried out on a matrix with the 48 populations on the rows and the subjects on the columns. We used 3D plots with the first three principal components to display the 48 T cell subsets and the subjects simultaneously. The overlay of the 3D plot of the scores (subjects) with the 3D plot of the loadings (combination of antigens) allows the identification of the variables that most contribute to the characterization of a

specific subject since they lie in the same area of the graphs. To validate the PCA results, a crossvalidation method called "leave one out" (contained in PLS Toolbox 3.5) was applied to the data sets under investigation. Briefly, the software creates a PCA model by randomly considering all but one subject (the training set). The remaining subject (the test set) is inserted in the model to test his classification on the basis of the considered variables (i.e. combinations of antigens). This approach has been carried out for some of the subjects included in the study both for CD4+ and CD8+ T cells.

RESULTS

Strategy for the Identification of the Subpopulations Present Within CD4+ and CD8+ T Cell Subset

The phenotypic features of centenarians, as far as the percentages of naïve (CD45RA+, CCR7+, defined T_N), central memory (CD45RA−, CCR7+, defined T_{CM}), effector memory (CD45RA−, CCR7−, defined T_{EM}), and terminally differentiated (CD45RA+, CCR7−, defined T_{EMRA}) T cells are concerned, have been previously described in details (13). The donors included in this study had almost identical T cell differentiation patterns of those above quoted (13). Data regarding the fine identification of the T cell populations were processed as shown in Figure 1. To simplify the analysis, we first gated on T lymphocytes (on the basis of their physical parameters and CD3 expression), then on CD4 or CD8 to identify the main T cell populations. In this study, double-positive (CD4+, CD8+) or double-negative (CD4−, CD8−) CD3+ T lymphocytes were not considered. Other gates were subsequently set on CD45RA, CCR7, CD127, CD95, and CD38 on the basis of their positivity and negativity. In the case of CD38, a further distinction was made between dim and bright expression. These gates were combined using Boolean gating strategy contained in FlowJo 6.3. A total of 48 populations were so obtained both for CD3+, CD4+, CD8−, and for CD3+, CD4−, CD8+ T cells.

Expression of CD127, CD95, and CD38 in Virgin and Memory CD4+ and CD8+ T Cells and Their Changes with Aging

In CD3+, CD4+, CD8−, and CD3+, CD−, CD8+ T cells from young donors, middle-aged donors, and centenarians, we determined the expression of antigens involved in cell survival (CD127), cell death (CD95), and immune activation (CD38) in peripheral naïve (dots in green), central memory (in black), effector memory (in yellow), and terminally differentiated (in red) T cells. Thus, we could quantify the 12 possible phenotypes generated by the combination of CD127 (negative or positive), CD95 (negative or positive), and CD38 (negative, dim, and bright).

Figures 2a and 2b, reporting representative examples, show that CD127 (the IL-7 receptor α chain) was present in almost all CD4+ and CD8+ T_N and T_{CM} from young donors, and was progressively down-regulated as cells proceeded to T_{EM} and T_{EMRA}. Note that this loss is greater for CD8+ than for CD4+ T cells. On the other hand, CD95, which was poorly expressed on T_N, was present on T_{CM} and T_{EM}. Unexpectedly, we found that the CD4+ T_{EMRA} subset, in contrast to the CD8+ counterpart, lost CD95.

We detected significant differences in CD38 expression between CD4+ and CD8+ T cells: indeed, in young subjects, most CD4+ cells were dim for CD38, whose expression did not change with peripheral differentiation. On the contrary, several CD8+ cells were CD38−, and its expression was modulated by the differentiation status. The same pattern was found in middle-aged donors, who did not differ for the expression of these antigens both on CD4+ and CD8+ T cells. On the contrary, centenarians were characterized by a higher amount of CD4+ and CD8+ naïve and memory cells expressing CD95, while CD127 was down-regulated on CD4+ T_{CM} and T_{EMRA} and on CD8+ T_N; CD38 expression did not change with age.

We have then analyzed the subsets in which statistical significance was present among the three groups of subjects (Table 1). We found that in the case of CD4+ T cells, age-dependent changes were found mainly in the T_N and T_{EMRA} subsets that preferentially expanded CD127−, CD95+ cells, either CD38dim or CD38−, in spite of CD127+, CD95− cells (not shown). To a major extent, this phenomenon was also observed for CD8+ T cells, since all CD8+ T cell subsets are mainly characterized by contraction of cells expressing CD127 and lacking CD95 and CD38, in favor of cells lacking CD127 but expressing CD95 and CD38.

Cluster Analysis of CD4+ and CD8+ T Cell Subpopulations

Cluster analysis for CD4+ T cells allowed the identification of three populations of subjects, as revealed by the three branches of the tree at the top of Figure 3a. Based on the CD4+ T cell flow cytometric profile, most young donors (indicated in red), one middle-aged subject (green), and one centenarian (blue) were clustered together (right branch). Centenarians were also found in the middle branch, along with middle-aged donors, as well as in the left branch, together with four middle-aged and three young donors. This suggests that as far as the lymphocyte subsets we analyzed were concerned, a high heterogeneity was present among CD4+ T cells. We were able to identify two main clusters of cell populations (as highlighted by blue- and orange-depicted branches): the first was poorly represented in young donors and expanded with age (upper part of the panel, orange branch), the second had an opposite trend (lower part of the panel, blue branch). However, CA of the CD4+ T cell flow cytometric profile failed to identify a particular trend in subset dynamics during aging.

Cluster analysis of CD8+ T cells allowed recognition of two groups of donors. As indicated in Figure 3b, all but one centenarian clustered on one branch of the tree, along with some middle-aged donors; all young donors, four middle-aged subjects and one centenarian clustered

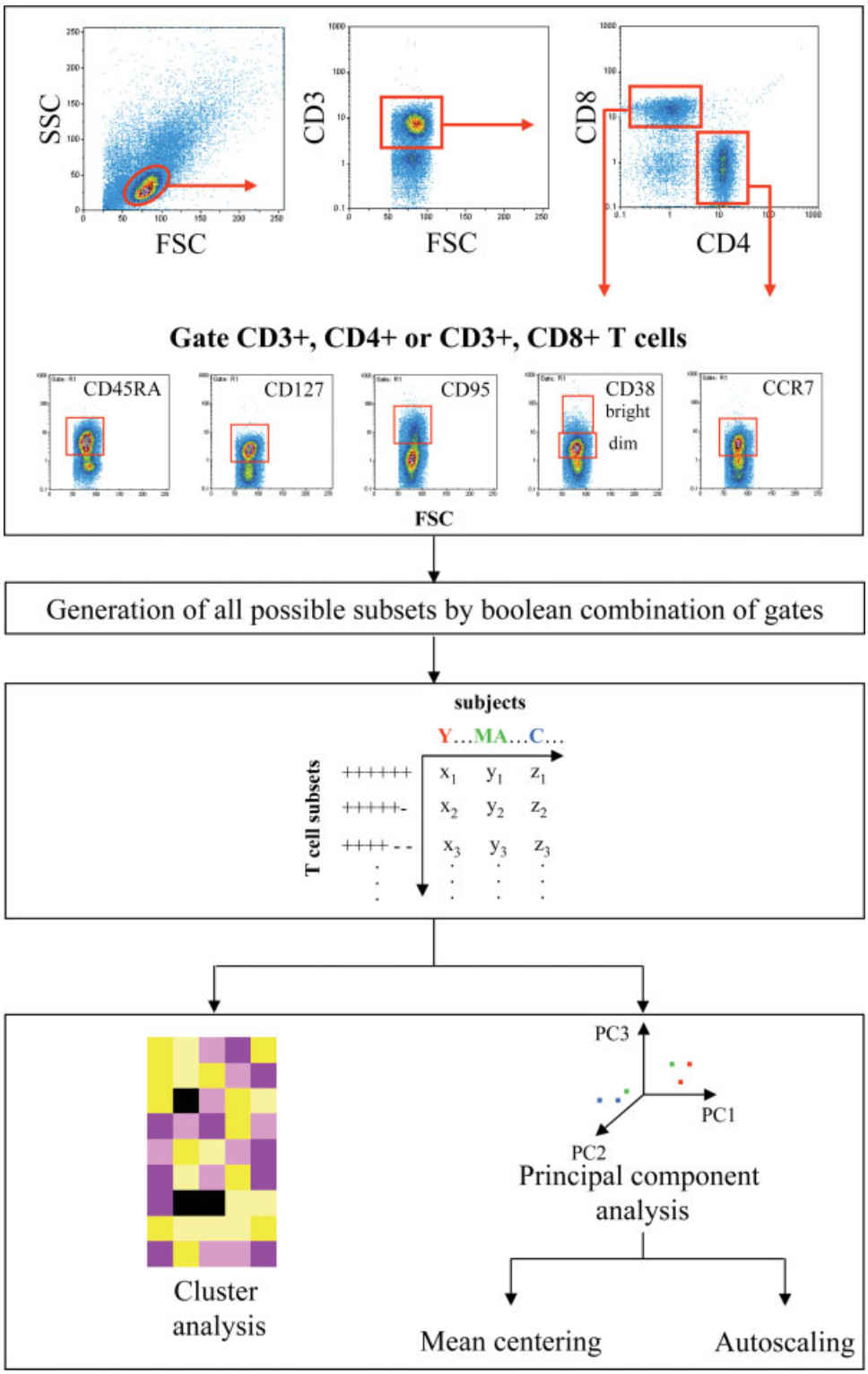

FIG. 1. Schematic representation of the methodology used for the analysis of cytometric data, which was performed as follows. First, T cells were identified by gating on lymphocytes on the basis of FSC and SSC, then by gating on CD3+, CD4+ or CD3+, CD8+ cells. Multiple gates were subsequently drawn on the basis of positivity and negativity of each indicated antigen. In the case of CD38, the expression was further distinguished between dim and bright. All possible phenotypes were next generated by performing automatic Boolean combinations. The data obtained were put in a matrix whose rows indicate T cell subsets and columns indicate subjects. The entire data sets were analyzed by CA and PCA. In the latter case, both mean centering and autoscaling were tested.

together in another branch. As observed for CD4+ T cells, two clusters of cell populations were obtained. In the upper part of the cluster array (identified by the branch indicated in blue), different T_N or memory subsets with preferential expression of CD127 but not CD95 or CD38 were found to be more frequent in young donors than in centenarians. Conversely, in the lower part of the cluster array (identified by the branch indicated in orange), memory subsets with expression of CD95 and lack of CD127 were more frequent in centenarians than in young sub-

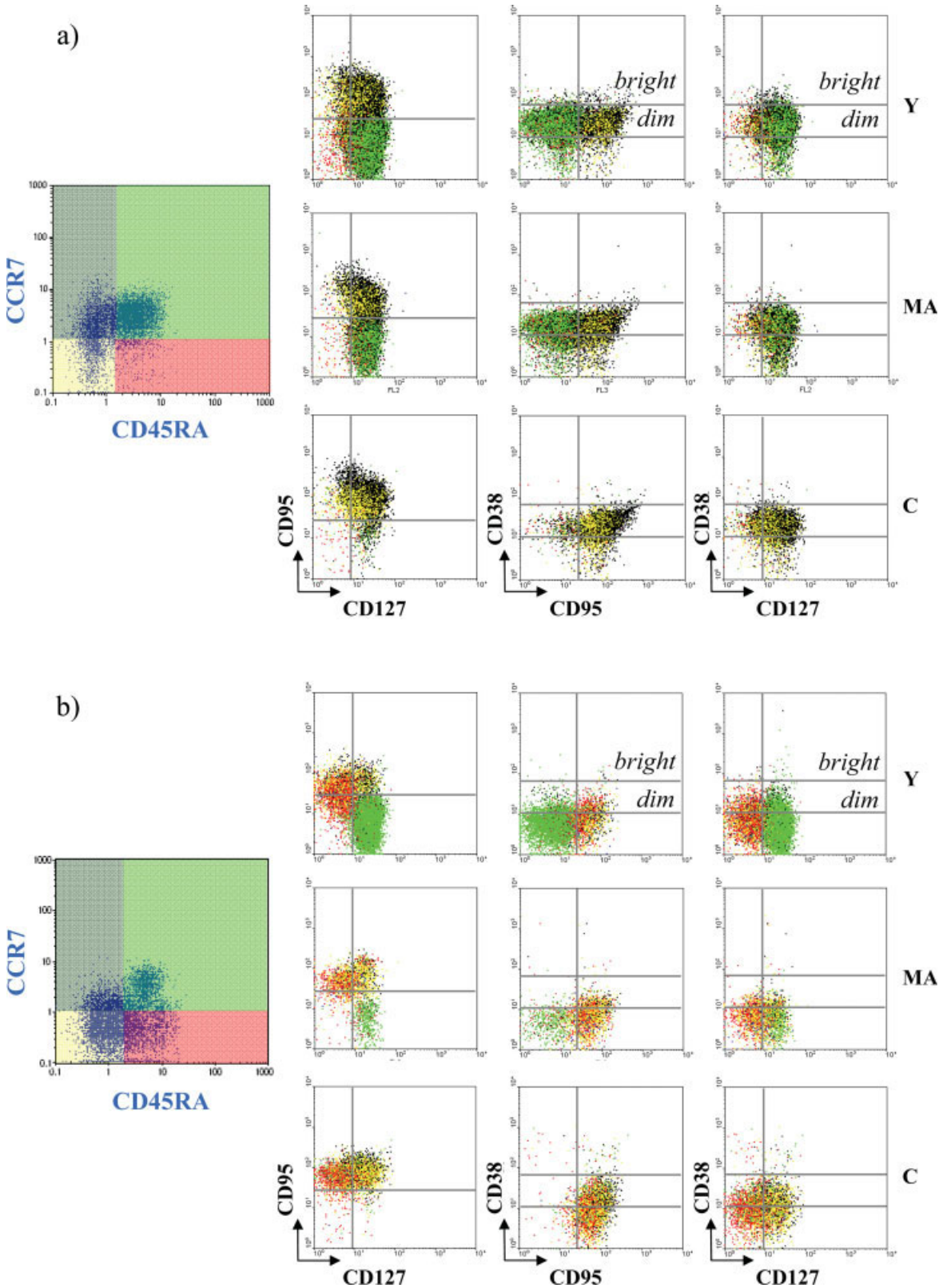

FIG. 2. Bivariate dot plots indicating the expression of CD127 (IL-7 receptor α chain), CD95 (Fas), and CD38 in CD4+ (**a**) and CD8+ (**b**) T cells from three representative subjects of different age: young (upper lane), middle-aged (middle lane), and centenarian (lower lane) donors. Naïve T cells (T_N: CD45RA+, CCR7+) were depicted in green, central memory (T_{CM}: CD45RA−, CCR7+) in black, effector memory (T_{EM}: CD45RA−, CCR7−) in yellow, and terminal effectors (T_{EMRA}: CD45RA+, CCR7−) in red. Y, young; MA, middle-age; C, centenarian.

Table 1
Lymphocyte Subsets that Showed a Statistically Significant Difference Among the Three Groups of Subjects

Antigen combination	Parental subset	Young	Middle-aged	Centenarians	*P* value
127 + 95 − 38dim + 38br−	CD4+ T_N	74.4 ± 5.7	67.8 ± 6.5	48.0 ± 7.1	<0.05
127 + 95 + 38dim + 38br−	CD4+ T_N	3.1 ± 0.3	6.6 ± 1.2	13.4 ± 2.2	<0.001
127 + 95 + 38dim − 38br+	CD4+ T_N	0.4 ± 0.1	0.9 ± 0.3	2.6 ± 1.0	<0.001
127 + 95 + 38dim − 38br−	CD4+ T_N	0.6 ± 0.3	1.6 ± 0.7	3.6 ± 1.5	<0.05
127 − 95 + 38dim + 38br−	CD4+ T_N	0.4 ± 0.0	1.4 ± 0.4	5.7 ± 2.6	<0.001
127 − 95 + 38dim − 38br+	CD4+ T_N	0.1 ± 0.0	0.3 ± 0.1	2.6 ± 1.5	<0.001
127 + 95 + 38dim + 38br−	CD4+ T_{EMRA}	5.9 ± 0.7	10.7 ± 3.5	14.3 ± 3.3	<0.05
127 + 95 + 38dim − 38br+	CD4+ T_{EMRA}	0.2 ± 0.1	0.6 ± 0.3	0.8± 0.1	<0.01
127 + 95 − 38dim + 38br−	CD4+ T_{EMRA}	48.1 ± 4.1	36.7 ± 6.5	19.4 ±6.0	<0.01
127 − 95 + 38dim + 38br−	CD4+ T_{EMRA}	3.6 ± 0.5	14.2 ± 4.6	25.6 ± 8.1	<0.001
127 − 95 + 38dim − 38br+	CD4+ T_{EMRA}	1.2 ± 0.5	1.2 ± 0.4	2.8 ± 0.4	<0.05
127 − 95 + 38dim − 38br−	CD4+ T_{EMRA}	1.2 ± 0.3	3.8 ± 1.0	8.0 ± 4.4	<0.05
127 − 95 − 38dim + 38br−	CD4+ T_{EMRA}	18.0 ± 2.4	11.8 ± 1.7	9.3 ± 2.1	<0.05
127 + 95 − 38dim + 38br−	CD8+ T_N	11.7 ± 2.8	7.4 ± 2.4	1.9 ± 0.6	<0.01
127 + 95 − 38dim − 38br−	CD8+ T_N	68.3 ± 5.0	59.3 ± 8.7	9.0 ± 3.0	<0.01
127 + 95 + 38dim + 38br−	CD8+ T_N	1.3 ± 0.4	2.9 ± 1.0	11.0 ± 3.0	<0.01
127 + 95 + 38dim − 38br+	CD8+ T_N	0.3 ± 0.1	0.2 ± 0.0	1.9 ± 0.7	<0.01
127 + 95 + 38dim − 38br−	CD8+ T_N	3.2 ± 0.5	7.1 ± 1.7	13.7 ± 2.3	<0.01
127 − 95 + 38dim + 38br−	CD8+ T_N	1.5 ± 0.4	3.5 ± 1.5	17.9 ± 4.3	<0.01
127 − 95 + 38dim − 38br+	CD8+ T_N	0.3 ± 0.1	0.2 ± 0.1	2.8 ± 1.6	<0.05
127 − 95 + 38dim − 38br−	CD8+ T_N	5.3 ± 1.7	14.6 ± 6.3	37.0 ± 8.0	<0.01
127 − 95 − 38dim + 38br−	CD8+ T_{CM}	4.1 ± 1.0	1.6 ± 0.5	0.5 ± 0.1	<0.01
127 − 95 − 38dim − 38br−	CD8+ T_{CM}	11.2 ± 1.9	8.5 ± 3.0	1.2 ± 0.3	<0.01
127 + 95 + 38dim + 38br−	CD8+ T_{CM}	6.8 ± 1.7	6.5 ± 1.0	13.8 ± 2.4	<0.05
127 − 95 + 38dim − 38br−	CD8+ T_{CM}	3.8 ± 0.8	1.0 ± 0.3	1.7 ± 0.8	<0.05
127 + 95 − 38dim − 38br−	CD8+ T_{EM}	7.0 ± 1.5	3.1 ± 1.4	0.7 ± 0.1	<0.01
127 − 95 + 38dim + 38br−	CD8+ T_{EM}	6.0 ± 0.9	7.3 ± 1.3	15.4 ± 2.5	<0.05
127 − 95 − 38dim − 38br−	CD8+ T_{EM}	7.8 ± 1.3	3.2 ± 0.9	1.8 ± 0.4	<0.01
127 + 95 − 38dim + 38br−	CD8 T_{EMRA}	2.2 ± 0.6	1.3 ± 0.6	0.1 ± 0.0	<0.01
127 + 95 − 38dim − 38br−	CD8 T_{EMRA}	19.8 ± 5.9	12.9 ± 6.2	0.7 ± 0.2	<0.01
127 − 95 + 38dim + 38br−	CD8 T_{EMRA}	7.2 ± 1.7	8.1 ± 1.6	22.2 ± 4.5	<0.05
127 − 95 − 38dim + 38br−	CD8 T_{EMRA}	3.4 ± 0.7	1.2 ± 0.4	0.8 ± 0.2	<0.01
127 − 95 − 38dim − 38br−	CD8 T_{EMRA}	16.6 ± 2.0	6.0 ± 1.5	2.3 ± 0.6	<0.01

Numbers in the first column indicate the cluster designation (CD) of the antigens under investigation; br stands for bright. Numbers in the column related to young, middle aged and centenarians indicate the percentage of cells with a given phenotype (in the first column of each row). T_N, naïve T cells; T_{CM}, central memory; T_{EM}, effector memory; T_{EMRA}, terminally differentiated (effector memory expressing CD45RA).

jects. Moreover, particular subsets of memory cells expressing CD95 but retaining CD45RA and CCR7 were found at higher frequency in centenarians. A high heterogeneity among middle-aged donors was observed since they were equally distributed between the two clusters comprising centenarians and young donors.

We also tested whether rare subsets (frequency $< 1\%$ in the majority of the patients) could influence the CA-based classification. Removal of these subsets from the analysis (both for CD4+ and CD8+ T cells) gave almost identical arrays (data not shown).

Principal Component Analysis of CD4+ and CD8+ T Cell Subpopulations

To obtain visualization of the entire data set generated by combination of cell-surface antigen expression, we applied PCA to reduce the dimensionality of the complex data set. We performed this type of analysis in order to test whether subject classification was possible on the basis of the T cell flow cytometric profile by considering all the identified variables at the same time. Score and loading plots for the first three components obtained by PCA of the CD4+ T cell data set after column mean centering are shown in Figures 4a and 4b, respectively. The first three principal components accounted for >93% of the total data variance, indicating that young donors (red triangles) were well distinguished from centenarians (blue squares). Indeed, the latter subjects had lower PC1 values and higher PC2 and PC3 values in comparison with young donors. Middle-age donors fell between these two groups, and most of them remained near the PC-axes origin, meaning that they did not differ from the average values of all three subject populations studied. Analysis of loadings (Fig. 4b) showed the cell populations that were mostly responsible for these differences and that were more frequent in the three groups of subjects. The majority of young subjects were characterized by the presence of CD45RA+, CCR7+, CD95− naïve cells expressing CD127 but not CD38, or with CD38 at dim levels, while centenarians were characterized by T_{CM} or T_{EM} cells, either expressing CD127 or not. Note that the majority of the subsets were found near the axis origin, indicating that they do not contribute to subject classification.

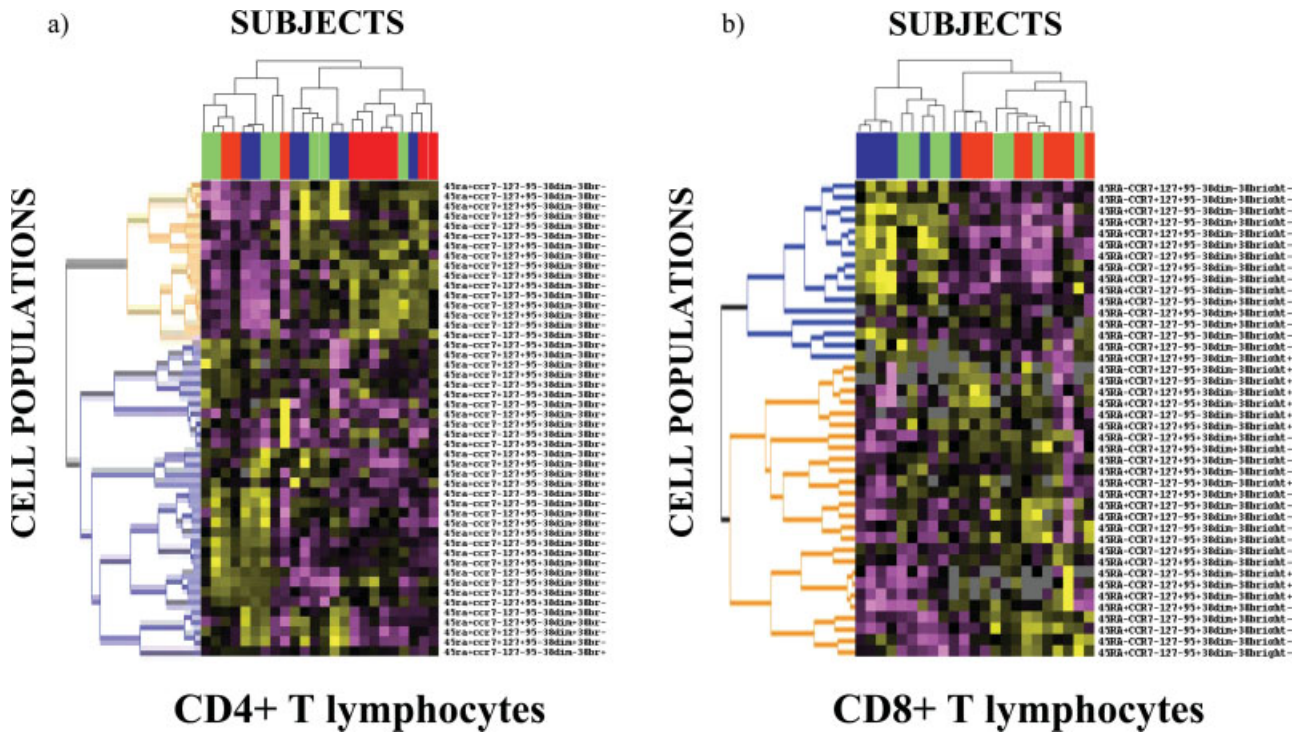

FIG. 3. Hierarchical clustering of CD4+ and CD8+ T cell subpopulations. Phenotypic combinations of antigens are in rows while subjects are in columns. Young subjects are indicated in red, middle-aged subjects in green, and centenarians in blue. Variable (cell population) coloring is based on the percentages normalized to the mean of variables of young donors. The color scale ranges from saturated yellow for log ratios of −3.0 and below to saturated purple for log ratios of +3.0 and above. Variables with missing values are depicted in grey. In the case of the CD8+ T cells, five cell populations (all comprising CD38bright cells) were not included in the matrix because they were absent in the majority of the subjects analyzed.

To strengthen the aforementioned results, PCA was performed on the same data but after autoscaling. The first three principal components accounted for >62% of the total variance, and young donors were distinguished from centenarians mainly on the basis of PC2 and partly by PC3 values rather than by PC1 values (Fig. 4c). This type of scaling, which assigns to each variable the same standard deviation (so giving to each variable the equal chance of contributing to the PCA model), showed that young subjects were preferentially characterized by naïve or memory CD127+, CD95−, CD38dim subsets, while memory T cells with a CD127−, CD95+ phenotype were mainly found in correspondence with centenarians (Fig. 4d).

In the case of CD8+ T cells, the first three principal components, calculated on mean-centered data, accounted for >92% of the total data variance. As with CD4+ T cells, young donors were characterized by higher PC1 values than centenarians while middle-aged subjects were scattered between them, indicating the extreme variability of these groups (Fig. 5a). Analysis of loadings indicates that the main feature of immunological aging of CD8+ T cells is the loss of a particular naïve (CD45RA+, CCR7+) subset expressing CD127 and lacking CD38 and CD95 in favor of CD45RA−, CCR7+ effector memory or CD45RA+, CCR7− terminal effector cells expressing CD95 and lacking CD127 and CD38 (Fig. 5b). As for CD4+ T cells, most of the subsets lay near the axis origin and did not drive subject classification. PCA of autoscaled data is presented in Figure 5c: the first three principal components accounted for >63% of the total variance. Young subjects were separated from centenarians on the basis of PC2 and partly PC3, and as for CD4+ T lymphocytes, young donors were mainly characterized by naïve cell subsets and to a lesser degree by CD127+, CD95−, CD38dim memory cells, centenarians by memory cells with a CD127−, CD95+ phenotype with different expression of CD38 (negative or dim, Fig. 5d).

Together, these results indicate that, during aging, CD4+ and CD8+ T cells undergo lineage-specific modifications (expansion of a particular differentiation state) although they share similar phenotypes (CD127−, CD95+).

PCA results were crossvalidated by applying the "leave one out" approach: some subjects were randomly selected by the software and inserted in the previously created model. In all cases, the subjects used for validation were plotted in their parental group (data not shown). As for CA, elimination of rare subsets from the analysis did not influence PCA-dependent classification of subjects, after either mean centering or autoscaling (data not shown).

DISCUSSION

The advanced technologies nowadays available permit the simultaneous analysis of multiple parameters in a particular cell population. However, the large amount of data

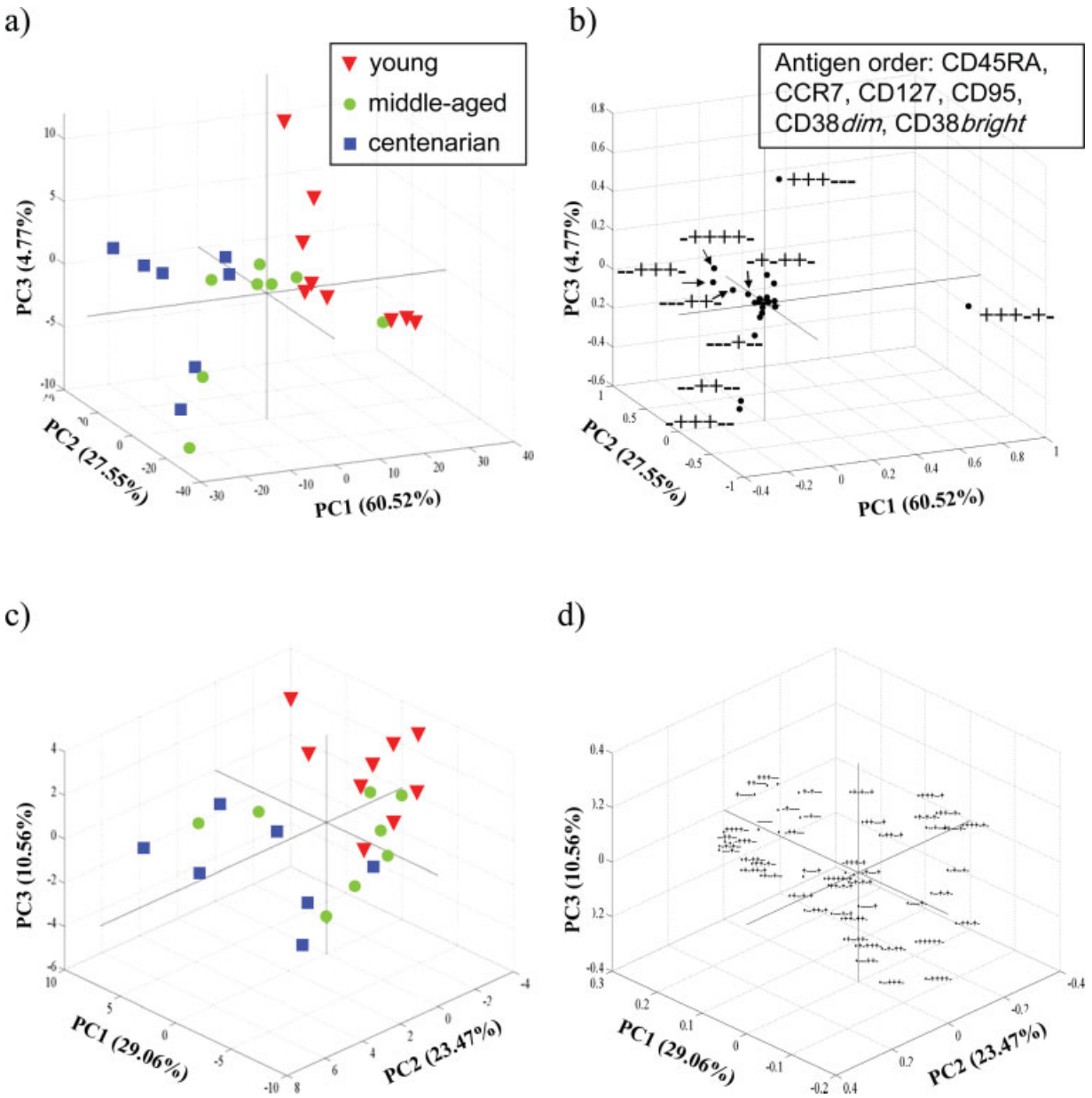

FIG. 4. PCA of CD4+ T cell subpopulations after mean centering (**a**,**b**) and autoscaling (**c**,**d**). (a,c): 3D score plot of the first three principal components (PC): young donors are represented by red triangles, middle-age donors by green circles, and centenarians by blue squares. (b,d): 3D loading plot of the first three PCs: cell populations most contributing to the PCA output are indicated as combinations of + and − that represent, respectively, the presence or absence of surface antigens in this order: CD45RA, CCR7, CD127, CD95, CD38dim, CD38bright.

generated by these techniques is difficult to analyze and thus computer assistance and mathematical algorithms are absolutely necessary for data interpretation. This problem is enhanced in polychromatic flow cytometry because the analysis is performed on single cells; as a consequence, hundreds of subsets can be identified in a specific cell population and each of them can bear a different function (5,14,15) or play a specific role in the development of diseases (16–20). A general approach to the visualization of flow cytometric data is the utilization of bivariate plots. However, in multicolor experiments, a significant amount of information is lost by this type of representation, since each cell is identified by multiple fluorescent values which correspond to the parameters analyzed. In this article we show that CA and PCA, widely used in microarray experiments, can overcome this problem and facilitate the understanding of complex flow cytometric data. Further, they permit subject classification on the basis of the T cell flow cytometric profile by considering all the parameters (i.e. surface markers) under investigation. As far as CD4+ and CD8+ T cells are concerned, analysis of the surface expression of five markers involved in extrathymic differ-

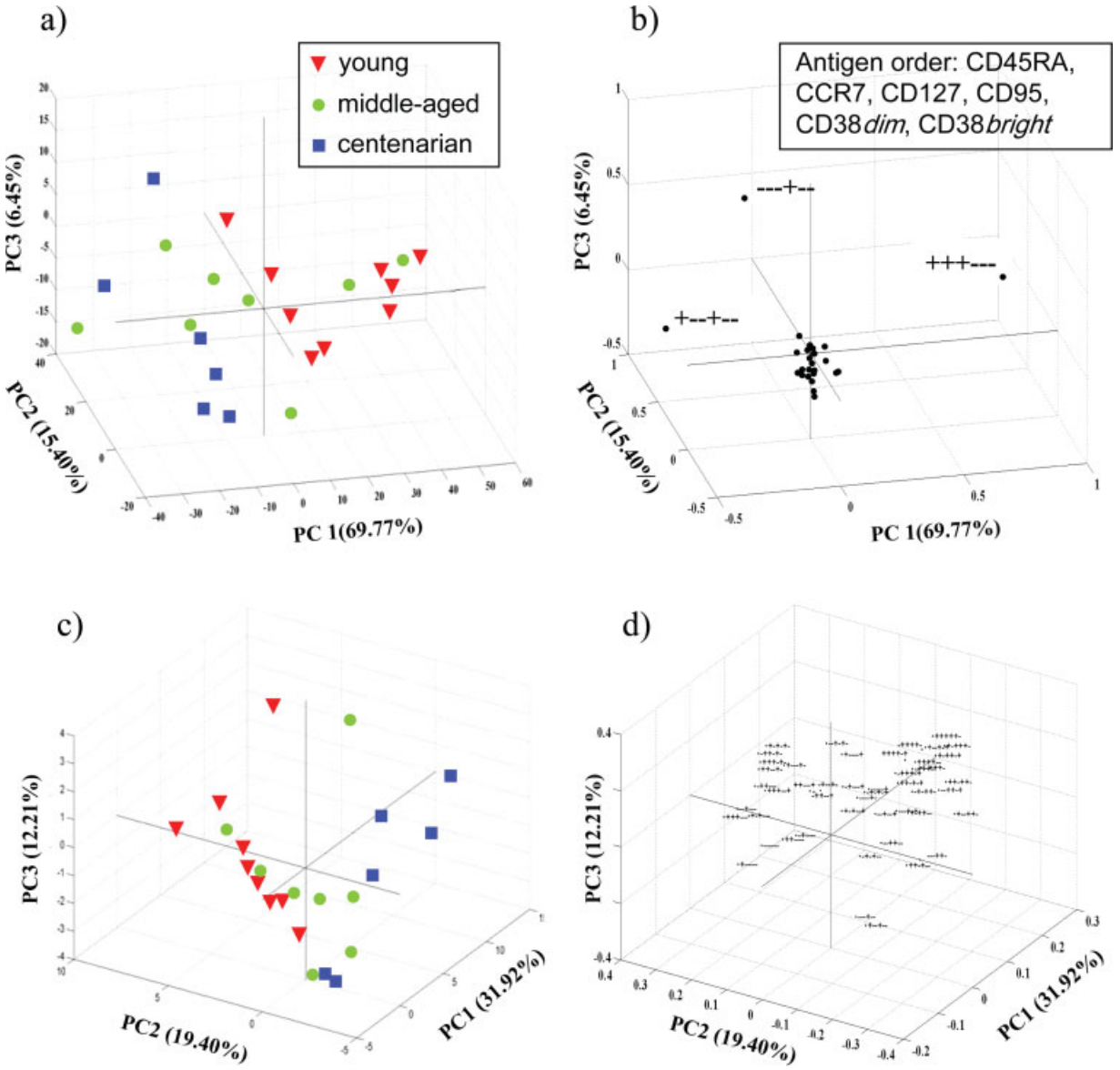

FIG. 5. PCA of CD8+ T cell subpopulations after mean centering (**a**,**b**) and autoscaling (**c**,**d**). (a,c): 3D score plot of the first three PCs. (b,d): 3D loading plot of the first three PCs. Subjects and variables (cell populations) are indicated as in Figure 4.

entiation and lymphocyte function by eight-color flow cytometry led us to identify a total of 48 possible populations within each subset. This type of subdivision of a given number of events into distinct populations reduces the number of events per population; as a consequence, rare subsets could display higher variability if a relative small number of events are acquired per sample, thus leading to false positive or false negative results. Subject classification obtained by CA and PCA is not influenced by this limit, since the elimination of rare subsets from the analysis did not alter the results obtained with the complete data set. This has been also recently described by Hofmann and Zerwes (21) who applied DNA-chip analysis software to multicolor data.

Here we report that while the majority of subsets identified by PFC is likely not fundamental for the comprehension of T cell dynamics during the aging process, some crucial populations arose from the analyses. If specific (CD127+, CD95−, CD38dim or CD127+, CD95−, CD38−) T_N subsets are the main feature of CD4+ and CD8+ T cells from young donors, particular memory subsets, preferentially displaying CD127−, CD95+ phenotypes, were found highly represented in people of advanced age. The application of CA allowed us to distinguish centenarians quite clearly from young donors on the basis of the CD8+ and (to a minor extent) CD4+ T cell flow cytometric profile. Indeed, CA grouped the T cell subsets undergoing the same expansion (memory

cells with preferential expression of CD95 but lacking CD127) or contraction (naïve T cell subsets) during immunological aging. PCA confirmed these results. In particular, PCA, which is able to compress very complex data sets to a three-dimensional space without loss of information, allowed the representation of a particular donor in this new system of coordinates (the principal components) by considering all the collected variables (in this case a particular phenotype/combination of antigens). This is of great advantage since complex flow cytometric data can be analyzed in their entirety. In this way, PCA led us to identify the phenotypes most responsible for the differences among groups of different age. For both CD4+ and CD8+ T cells, young donors and centenarians clustered in two separate groups, while middle-aged subjects were scattered between them. This classification was mainly driven by the naïve-to-memory conversion. In particular, in the case of CD4+ T cells, we observed naïve T cell depletion in favor of a memory pool comprising CD95+ T_{CM} or T_{EM} cells with or without CD127; in the case of CD8+ T cells, centenarians mainly expand T_{EM} or T_{EMRA} subsets, both characterized by the expression of CD95 and by the absence of CD127 and CD38. This was well evidenced by the application of PCA to autoscaled data, a method that assigns the same standard deviation to all variables, allowing all of the variables to contribute equally to the PCA model. Such an assay confirmed that memory subsets displaying the aforementioned characteristics are more frequent in centenarians than in young subjects.

Besides the technical observations, these data suggest a remodeling of the memory T cell pool in centenarians because of the different expression of the molecules regarding T cell survival and death, CD127 and CD95 respectively. Alterations in CD127 signaling in terminally differentiated CD8+ T_{EMRA} cells has been reported in aged people (22). It is possible, however, that although the production of IL-7 is maintained (13), dysfunctions affecting its receptor could influence the survival of CD8+ T cells in the periphery. Moreover, the increased expression of CD95 in most of the T cell subsets identified in centenarians suggests a different tendency of senescent T cells to undergo apoptosis, although there are contrasting reports on this idea: if some groups have shown a lower rate of activation-induced cell death in older subjects (23–25), others, included us, have suggested a remodeling of the survival-death balance (26–28).

In conclusion, the data presented here indicate that among several different subpopulations of CD4+ and CD8+ T cells, there are a few that expand during the aging process, and that can be recognized by a bioinformatic approach to cytometric analysis. In particular, CA and PCA provide a unique tool to identify cellular dynamics during multiple conditions (in this case the aging of the immune system), in addition to the simplification of the analysis of large and complex data sets. Finally, they allow subject classification on the basis of the flow cytometric profile under consideration and could be very helpful to identify the cellular subsets associated to a group of patients with specific features, such as a different clinical outcome or response to therapy.

ACKNOWLEDGMENTS

We thank Prof. Wolfgang Göhde (University of Münster, Germany) for critical reading of the manuscript and helpful comments, and Dr. Luca Cicchetti (Space Import Export, Milan, Italy) for continuous support. Dr. Gabriele Marcotullio is acknowledged for precious editorial assistance.

LITERATURE CITED

1. De Rosa SC, Herzenberg LA, Roederer M. 11-color, 13-parameter flow cytometry: Identification of human naive T cells by phenotype, function, and T-cell receptor diversity. Nat Med 2001;7:245–248.
2. De Rosa SC, Brenchley JM, Roederer M. Beyond six colors: A new era in flow cytometry. Nat Med 2003;9:112–117.
3. Kern F, LiPira G, Gratama JW, Manca F, Roederer M. Measuring Ag-specific immune responses: Understanding immunopathogenesis and improving diagnostics in infectious disease, autoimmunity and cancer. Trends Immunol 2005;26:477–484.
4. Suni MA, Maino VC, Maecker HT. Ex vivo analysis of T-cell function. Curr Opin Immunol 2005;17:434–440.
5. Perfetto SP, Chattopadhyay PK, Roederer M. Seventeen-colour flow cytometry: Unravelling the immune system. Nat Rev Immunol 2004; 4:648–655.
6. Sallusto F, Lenig D, Forster R, Lipp M, Lanzavecchia A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature 1999;401:708–712.
7. Marrack P, Kappler J. Control of T cell viability. Annu Rev Immunol 2004;22:765–787.
8. Cossarizza A, Stent G, Mussini C, Paganelli R, Borghi V, Nuzzo C, Pinti M, Pedrazzi J, Benatti F, Esposito R, Rosok B, Nagata S, Vella S, Franceschi C, De Rienzo B. Deregulation of the CD95/CD95L system in lymphocytes from patients with primary acute HIV infection. AIDS 2000;14:345–355.
9. Ligthart GJ, Corberand JX, Fournier C, Galanaud P, Hijmans W, Kennes B, Muller-Hermelink HK, Steinmann GG. Admission criteria for immunogerontological studies in man: The SENIEUR protocol. Mech Ageing Dev 1984;28:47–55.
10. Gueresi P, Troiano L, Minicuci N, Bonafé M, Pini G, Salvioli G, Carani C, Ferrucci L, Spazzafumo L, Olivieri F, Cavrini G, Valentini D, Franceschi C. The MALVA (MAntova LongeVA) study: An investigation on people 98 years of age and over in a province of Northern Italy. Exp Gerontol 2003; 38:1189–1197.
11. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–14868.
12. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom Intell Lab Syst 1987;2:37–52.
13. Nasi M, Troiano L, Lugli E, Pinti M, Ferraresi R, Monterastelli E, Mussi C, Salvioli G, Franceschi C, Cossarizza A. Thymic output and functionality of the IL-7/IL-7 receptor system in centenarians: Implications for the neolymphogenesis at the limit of human life. Aging Cell 2006;5:167–175.
14. Perez OD, Nolan GP. Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. Nat Biotechnol 2002;20:155–162.
15. Chattopadhyay PK, Yu J, Roederer M. A live-cell assay to detect antigen-specific CD4+ T cells with diverse cytokine profiles. Nat Med 2005;11:1113–1117.
16. Champagne P, Ogg GS, King AS, Knabenhans C, Ellefsen K, Nobile M, Appay V, Rizzardi GP, Fleury S, Lipp M, Forster R, Rowland-Jones S, Sekaly RP, McMichael AJ, Pantaleo G. Skewed maturation of memory HIV-specific CD8 T lymphocytes. Nature 2001;410:106–111.
17. Brenchley JM, Karandikar NJ, Betts MR, Ambrozak DR, Hill BJ, Crotty LE, Casazza JP, Kuruppu J, Migueles SA, Connors M, Roederer M, Douek DC, Koup RA. Expression of CD57 defines replicative senescence and antigen-induced apoptotic death of CD8+ T cells. Blood 2003;101:2711–2720.
18. Curiel TJ, Coukos G, Zou L, Alvarez X, Cheng P, Mottram P, Evdemon-Hogan M, Conejo-Garcia JR, Zhang L, Burow M, Zhu Y, Wei S, Kryczek I, Daniel B, Gordon A, Myers L, Lackner A, Disis ML, Knutson KL, Chen L, Zou W. Specific recruitment of regulatory T cells in ovarian carcinoma fosters immune privilege and predicts reduced survival. Nat Med 2004;10:942–949.

19. Betts MR, Nason MC, West SM, De Rosa SC, Migueles SA, Abraham J, Lederman MM, Benito JM, Goepfert PA, Connors M, Roederer M, Koup RA. HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T-cells. Blood 2006;107:4781-4789.
20. Chattopadhyay PK, Price DA, Harper TF, Betts MR, Yu J, Gostick E, Perfetto SP, Goepfert P, Koup RA, De Rosa SC, Bruchez MP, Roederer M. Quantum dot semiconductor nanocrystals for immunophenotyping by polychromatic flow cytometry. Nat Med 2006;12:972–977.
21. Hofmann M, Zerwes HG. Identification of organ-specific T cell populations by analysis of multiparameter flow cytometry data using DNA-chip analysis software. Cytometry A 2006;69A:533–540.
22. Kim HR, Hong MS, Dan JM, Kang I. Altered IL-7Rα expression with aging and the potential implications of IL-7 therapy on CD8+ T-cell immune responses. Blood 2006;107:2855–2862.
23. Spaulding C, Guo W, Effros RB. Resistance to apoptosis in human CD8+ T cells that reach replicative senescence after multiple rounds of antigen-specific proliferation. Exp Gerontol 1999;34:633–644.
24. Bryl E, Gazda M, Foerster J, Witkowski JM. Age-related increase of frequency of a new, phenotypically distinct subpopulation of human peripheral blood T cells expressing lowered levels of CD4. Blood 2001;98:1100–1107.
25. Bryl E, Vallejo AN, Weyand CM, Goronzy JJ. Down-regulation of CD28 expression by TNF-α. J Immunol 2001;167:3231–3238.
26. Salvioli S, Capri M, Scarcella E, Mangherini S, Faranca I, Volterra V, De Ronchi D, Marini M, Bonafe M, Franceschi C, Monti D. Age-dependent changes in the susceptibility to apoptosis of peripheral blood CD4+ and CD8+ T lymphocytes with virgin or memory phenotype. Mech Ageing Dev 2003;124:409–418.
27. Pinti M, Troiano L, Nasi M, Bellodi C, Ferraresi R, Mussi C, Salvioli G, Cossarizza A. Balanced regulation of mRNA production for Fas and Fas ligand in lymphocytes from centenarians: How the immune system starts its second century. Circulation 2004;110:3108–3114.
28. Ginaldi L, De Martinis M, Monti D, Franceschi C. Chronic antigenic load and apoptosis in immunosenescence. Trends Immunol 2005; 26:79–84.