Regulatory T cells in melanoma revisited by a computational clustering of FOXP3$^+$ T cell subpopulations[1]

Hiroko Fujii[*], Julie Josse[†], Miki Tanioka[*], Yoshiki Miyachi[*], François Husson[†], and Masahiro Ono[*‡¶2]

* Department of Dermatology, Graduate School of Medicine, Kyoto University, Kyoto, Japan

† Laboratoire de mathématiques appliquées, Agrocampus Ouest, Rennes, France

‡ Department of Life Sciences, Faculty of Natural Sciences, Imperial College London, Sir Alexander Fleming Building, Exhibition Road, London, SW7 2AZ, United Kingdom

¶ Immunobiology, UCL Institute of Child Health, 30 Guilford Street, London, WC1N 1EH, United Kingdom

1 This work was supported by a David Phillips Fellowship (BB/J013951/1) from the Biotechnology and Biological Sciences Research Council (BBSRC), and also partly by a Long-Term Fellowship from the Human Frontier Science Program.

2 Address correspondence and reprint requests to Dr. Masahiro Ono, Department of Life Sciences, Sir Alexander Fleming Building, Imperial College London, Imperial College Road, London, SW7 2AZ, United Kingdom. E-mail address: m.ono@imperial.ac.uk

Running title: Computational clustering of FOXP3$^+$ T cell subpopulations

Abbreviations used in this paper: Treg, regulatory T cells; HDDC, high-dimensional data clustering; MFI, mean fluorescence intensity; HC, healthy control.

Abstract

$CD4^+$ T cells that express the transcription factor FOXP3 ($FOXP3^+$ T cells) are commonly regarded as immunosuppressive regulatory T cells (Treg). $FOXP3^+$ T cells are reported to be increased in tumour-bearing patients or animals, and considered to suppress anti-tumour immunity, but the evidence is often contradictory. In addition, accumulating evidence indicates that FOXP3 is induced by antigenic stimulation, and that some non-Treg $FOXP3^+$ T cells, especially memory-phenotype $FOXP3^{low}$ cells, produce proinflammatory cytokines. Accordingly, the subclassification of $FOXP3^+$ T cells is fundamental for revealing the significance of $FOXP3^+$ T cells in tumour immunity, but the arbitrariness and complexity of manual gating have complicated the issue. Here we report a computational method to automatically identify and classify $FOXP3^+$ T cells into subsets using clustering algorithms. By analysing flow cytometric data of melanoma patients, the proposed method showed that the $FOXP3^+$ subpopulation that had relatively high FOXP3, CD45RO, and CD25 expressions was increased in melanoma patients, whereas manual gating did not produce significant results on the $FOXP3^+$ subpopulations. Interestingly, the computationally-identified $FOXP3^+$ subpopulation included not only classical $FOXP3^{high}$ Treg but also memory-phenotype $FOXP3^{low}$ cells by manual gating. Furthermore, the proposed method successfully analysed an independent dataset, showing that the same $FOXP3^+$ subpopulation was increased in melanoma patients, validating the method. Collectively, the proposed method successfully captured an important feature of melanoma without relying on the existing criteria of $FOXP3^+$ T cells, revealing a hidden association between the T cell profile and melanoma, and providing new insights into $FOXP3^+$ T cells and Treg.

Introduction

Regulatory T cells (Treg) are defined as the immunosuppressive T cells that suppress the activities of other T cells through undefined mechanisms, and are identified by the transcription factor FOXP3 (1). While Treg are reported to be increased in tumour-bearing patients or animals, and thereby suppress anti-tumour immunity (2-4), the evidence is in fact mixed (5): the increase of $FOXP3^+$ T cells is associated with poor prognosis in hepatocellular cancer (6), whereas it is related to good prognosis in colorectal cancer (7). The discrepancy may be explained by that $FOXP3^+$ T cells include not only regulatory but also non-regulatory T cells that produce proinflammatory cytokines (8). In fact, accumulating evidence indicates that FOXP3 is not the definitive marker for the immunosuppressive T cells in humans. The expression of FOXP3 can be induced in naïve T cells by conventional anti-CD3 stimulation (9, 10). In addition, some $FOXP3^+$ T cells, especially memory-phenotype $CD45RO^+FOXP3^{low}$ cells, produce effector cytokines, and are not suppressive by an *in vitro* assay, suggesting that they are enriched with effector and activated T cells (9).

Accordingly, the subclassification of $FOXP3^+$ T cells has been a major issue in human Treg research (8, 9, 11-17). It was proposed that $FOXP3^+$ T cells could be classified into three functionally different subpopulations: $CD45RO^+$(equivalent to $CD45RA^-$) $FOXP3^{high}$ T cells as classical Treg with suppressive activity (9, 11); $CD45RO^-$ (or $CD45RA^+$) $FOXP3^{low}$ naïve Treg (9, 12, 13); and $FOXP3^{low}CD45RO^+$ non-regulatory T cells (9, 14, 15). This classification has been used to analyse $FOXP3^+$ T cells in autoimmune diseases and cancers (8, 9, 16, 17). Unfortunately, however, the definition of $FOXP3^+$ subpopulations varies between studies, complicating the problem (18). Meanwhile, recently Abbas *et al* proposed not to use new terms for Treg subpopulations, until a new population has been extensively demonstrated to be unique, distinct from other populations and stable, because it is likely to lead to more confusion and the further 'jargonisation' of immunology (19). This opinion, however, ignores the fact that a clustering (classification) approach, whether manual or automatic gating, is indispensable for summarising and analysing flow cytometric data and thereby relating immunological profiles to biological response or disease status (20, 21).

Currently in experimental immunology, any cellular populations, including $FOXP3^+$ T cells, are almost always identified and analysed by *manual gating*, which is a process of identifying a cluster of cells by manually drawing regions, or *gates*, in two-dimensional (2D) graphical representations of the data (22-24). Obviously, manual gating is subjective and cannot fully use multidimensional flow

cytometry data, and therefore, the automatic gating using clustering methods has become an active research area of bioinformatics over the past several years (23, 25). In fact, a preceding study proposed a computational approach to identify a regulatory T cell population, precisely $CD4^+CD25^+DR^+FOXP3^+$ cells (26), but it did not address the immunological significance of the approach and that of the identified Treg population.

In this study, we aimed to establish a computational approach to identify and classify $FOXP3^+$ T cells into subpopulations, addressing the immunological and clinical significance of the method, and thereby to revisit the fundamental subclassification of $FOXP3^+$ T cells. In order to establish the method, we used a dataset of PBMC from melanoma patients and healthy controls, which previously identified that total $FOXP3^+$ T cells increased in melanoma patients, and that $FOXP3^{low}$ naïve Treg and $FOXP3^{low}$ non-regulatory T cells increased as the stage progressed (8). Furthermore, in the current study, we have newly obtained a flow cytometric dataset of $FOXP3^+$ T cells from melanoma patients and healthy controls (designated as the second dataset), in order to address the efficiency of the proposed method. Thus, we firstly show the clinical and immunological significance of a data-oriented clustering approach to the subclassification of the $FOXP3^+$ T cells.

Materials and Methods

Patient samples

All PBMC datasets were obtained from patients with malignant melanoma who were treated in the Department of Dermatology of Kyoto University Hospital. The first dataset analysed 23 individuals by FACS Calibur (BD Biosciences) (8). The second dataset analysed 19 individuals by LSR Fortessa (BD Biosciences). The patients' characteristics in the second dataset are summarised in **Table I**. We also obtained data from age- and sex-matched healthy controls (first dataset, n=28; second dataset, n=15). This study was approved by the Medical Ethics Committee of Kyoto University, and conducted in accordance with the principles of the Declaration of Helsinki. All participants provided written informed consent.

Flow cytometric analysis

PBMC were isolated with Ficoll-Isopaque (Lymphoprep™; Axis-Shield, Oslo, Norway) gradient centrifugation. All cells were freshly stained with the following monoclonal antibodies and analysed promptly as previously described (9): FITC-conjugated anti-CD45RO (UCHL1; BD Biosciences);

PE-conjugated anti-CD25 (M-A251; BD Biosciences); PerCP-Cy5.5-conjugated anti-CD4 (SK3; BD Biosciences); and biotinylated anti-FOXP3 (236A/E7; eBioscience, San Diego, CA) and allophycocyanin–streptavidin (BD Biosciences). The second dataset (**Table I**) was obtained by LSR Fortessa (BD Biosciences), using the following settings for voltage: FSC-A (319), SSC-A (335), FL1-A (CD45RO, 582), FL2-A (CD25, 478), FL3-A (CD4, 742), and FL4-A (FOXP3, 676). FlowJo (Tree Star) was used for manual gating.

Automatic gating of FOXP3$^+$ T cell subpopulations

For data preprocessing, boundary values were removed (i.e. 1000 >= or < 100 [FSC], 1000 >= [SSC], >= 4 or < 0.3 [fluorescence channels, logged] in the case of analogue data; 800 >= or < 100 [FSC], 1000 >= [SSC], and >= 3.5 or < 0.1 [fluorescence channels, logged] in the case of digital data), as they were considered meaningless events representing cellular debris or large non-lymphocytes, or noise (27). Subsequently, all fluorescence data were log-transformed, and each variable was normalised by the standardised scaling. In the established classification method (*the HKK clustering*), FOXP3$^+$ T cell subpopulations were identified by the following three steps: [1] CD4$^+$ T cells were clustered by a high-dimensional data clustering function, *hddc*, of a CRAN package, *HDclassif* (28) using FSC, SSC, and CD4 (k = 3). [2] FOXP3$^+$ T cells were clustered by a k-means clustering of FOXP3 values using *kmeans* of a CRAN package, *Stats* (29), and the cluster containing the centroid with the highest FOXP3 value was designated as FOXP3$^+$ T cells. The number of clusters (k = 3) was determined by examining the barplot of the loss-variability (30), and also, taking into account the identification of the FOXP3$^+$ cluster that has higher FOXP3 values than the FOXP3-negative cloud. [3] Finally, FOXP3$^+$ T cell subpopulations were identified by a k-means clustering, using CD45RO, CD25, and FOXP3 with k = 3, and subsequently assigned to the *Effector-Treg-like*, *Naïve-Treg-like*, and *Non-Treg-like* clusters as follows: [1] compute the centroid of each cluster and designate the cluster containing the centroid with the highest value for FOXP3 as *effector-Treg-like*, [2] among the two other clusters, the one with the smallest value for CD45RO as *naïve-Treg-like*, and the last one as *non-Treg-like*. All computational analyses were done using a laptop with an Intel Core i5-3360 M CPU - 2.80 GHz or a Mac desktop with 3.5 GHz Intel Core i5, OS10.10.4.

Statistical analysis

A Mann-Whitney-Wilcoxon test was used for analysing two groups, testing the null hypothesis that the two statuses (i.e. HC or melanoma) have equal medians. A Kruskal-Wallis test, a non-parametric alternative to ANOVA, was used for analysing more than two groups, testing the null hypothesis that the medians are equal across the groups (HC and disease stages), followed by pairwise comparisons using a Mann-Whitney-Wilcoxon test. P-values were adjusted by a Bonferroni procedure for multiple comparisons in all the analyses.

Results

1. Manual gating approach to FOXP3$^+$ T subpopulations

This section shows how the standard approach, *manual gating*, identified and classified FOXP3$^+$ T cells into three subsets: CD45RO$^+$FOXP3^{high} effector (memory) Treg ("*effector-Treg population*" hereafter), CD45RO$^-$FOXP3^{low} naïve Treg ("*naïve-Treg population*"), and CD45RO$^+$FOXP3^{low} non-regulatory T cells ("*non-Treg population*") (8, 9) (**Fig. 1a**). These subpopulations were identified in a sequential manner using the following four gates: [1] the lymphocyte gate using Forward Scatter (FSC), and Side Scatter (SSC, **Fig. 1b**); [2] the CD4$^+$ gate using CD4 and SSC (**Fig. 1c**); and [3] the FOXP3$^+$ gate using FOXP3 and CD45RO to visually identify a "groove" of the FOXP3 distribution (**Fig. 1d**); [4] the FOXP3/CD45RO gate to identify the FOXP3$^+$ T cell subpopulations (**Fig. 1e**). The level of FOXP3 by which FOXP3^{high} and FOXP3^{low} cells were separated was determined so that CD45RO$^-$FOXP3^{high} cells were < 0.2% of CD4$^+$ lymphocytes using healthy controls (**Fig. 1e**).

2. Data-oriented clustering (automatic gating) of FOXP3$^+$ T cell subpopulations

We aimed to establish an automated clustering method for identifying FOXP3$^+$ T cells and classifying them into three subpopulations, and thereby to revisit the immunological significance of the FOXP3$^+$ T cell classification. Importantly, there is no major controversy regarding the identification of the total FOXP3$^+$ CD4$^+$ T cells, while there are multiple ways to classify FOXP3$^+$ T cells, which we aimed to address in this study. Thus, we used the following approach to identify the FOXP3$^+$ T cell subpopulations: [1] to identify FOXP3$^+$CD4$^+$ T cells, and [2] to classify FOXP3$^+$CD4$^+$ T cells into three subpopulations without using the manual gating strategy. We used the flow cytometric dataset in our previous report, which was obtained by a FACS Calibur (8) ("*the first dataset"*), in order to establish a clustering method.

Automatic gating of FOXP3$^+$ CD4$^+$ T cells

The aim of this step is to efficiently and robustly identify $FOXP3^+$ $CD4^+$ T cells as described above. $CD4^+$ cells are distinct from $CD4^-$ cells in the space of FSC, SSC, and CD4 (Fig. 1b, 1c), while the distribution of FOXP3 expression is continuous in $CD4^+$ T cells (Fig. 1d). In fact, a preliminary analysis showed that $CD4^+$ T cells were efficiently identified using a model of high-dimensional data clustering (HDDC, (28)), but not by a common clustering method, k-means. HDDC is based on Gaussian mixtures with restricted covariance matrices (28), and can efficiently identify elliptic populations such as $CD4^+$ T cells in the space of FSC, SSC, and CD4. On the other hand, $FOXP3^+$ cells are not as discrete as $CD4^+$ T cells, and it was not obvious what method was suitable.

Thus, we compared different combinations of the clustering methods using a resampling approach, addressing the sensitivity and the accuracy of the methods. Here we compared the following methods: [1] *HK clustering*: $CD4^+$ T cell selection by HDDC, followed by $FOXP3^+$ T cell selection by k-means; [2] *HH clustering*: $CD4^+$ T cell selection by HDDC, followed by $FOXP3^+$ T cell selection by HDDC; and [3] *one-step H clustering*: $FOXP3^+CD4^+$ T cell selection by one-step HDDC. Using several random number seeds, the HK clustering showed the highest sensitivities and accuracies across different cell numbers comparing to the other methods (**Fig. 2a-2g**).

Automatic classification of the three $FOXP3^+$ T cell subpopulations

Next, we aimed to establish a method that subclassifies the $FOXP3^+CD4^+$ T cells into three subpopulations without relying on the manual gating criteria, and thereby to readdress the significance of $FOXP3^+CD4^+$ T cell subpopulations. We compared k-means and HDDC ($k = 3$; designated as *HKK and HKH clustering methods*, respectively) by a resampling approach, in order to identify a method that consistently assigns similar cells to each cluster, using CD45RO, CD25, and FOXP3. The resampling experiment showed that the HKK clustering had smaller variations in both the percentage and the mean fluorescence intensities (MFI) of CD45RO, CD25, and FOXP3 of each subpopulation, than the HKH clustering (**Fig. 3a-3d**). Thus, the HKK clustering has been chosen as the method for identifying and classifying the $FOXP3^+$ T cell subpopulations.

The three clusters identified by the HKK clustering were partially overlapped with the three subpopulations identified by manual gating. In order to make the clusters and the subpopulations comparable, the cluster with the highest FOXP3 expressions was designated as *effector-Treg-like*, and those with low and high CD45RO expressions were designated as *naïve-Treg-like* and *non-Treg-*

like clusters, respectively (**Fig. 3e-3f**). Obviously, the effector-Treg-like cluster contained not only the cells that were identified as effector-Treg but also some of the $FOXP3^{low}$ cells that were identified as non-Treg by manual gating (**Fig. 3e-3f**). The HKK clustering was not computationally expensive: it took 45 seconds per patient on average to perform all the 3 clustering steps using a conventional laptop.

Statistical comparisons of $FOXP3^{+}$ T cell subpopulations between manual and automatic gating

We compared the subpopulations that were identified by the HKK clustering with those by the manual gating, in order to understand the similarities and dissimilarities of the two approaches. Obviously, the HKK clustering included more cells in the effector-Treg-like cluster than manual gating, while the latter included more in the non-Treg-like cluster (**Fig. 4a-4d**). The effector-Treg-like and the non-Treg-like clusters showed relatively low correlations with their corresponding manually-gated populations ($r = 0.6236$ and 0.7782, respectively) (**Fig. 4e, 4f**). On the other hand, the naïve-Treg-like cluster and all $FOXP3^{+}$ T cells had high correlations with their corresponding manually-gated populations ($r = 0.9168$ and 0.8947, respectively, **Fig. 4g, 4h**). All results were statistically significant ($p < 0.0001$).

Next, the HKK clustering and the manual gating approaches were compared for association with melanoma. As we previously reported, the manually-gated three subpopulations, effector-Treg, naïve-Treg, and non-Treg, showed a significant difference between HC and melanoma patients ($p < 0.05$) (8). However, assuming that the three subpopulations might be related to each other, when p-values were adjusted for multiple comparisons using a Bonferroni method, all the adjusted p-values exceeded 0.05, and thus, the manually-identified subpopulations did not show significant difference using the conservative approach. On the other hand, among the automatically-identified clusters, the effector-Treg-like cluster was significantly increased in melanoma patients comparing to controls (adjusted p-value = 0.027, **Fig. 5**). This result suggested that the effector-Treg-like cluster by the HKK clustering more efficiently captured the characteristics of melanoma patients.

Application of automatic gating to an independent dataset

Lastly, we attempted to apply the established method to an independent dataset. We have generated a new dataset using a different flow cytometer, *LSR Fortessa*, analysing 12 healthy controls and 19 melanoma patients ("*the second dataset*", **Table I**). When p-values were adjusted for the multiple

comparisons of the three clusters, the effector-Treg-like cluster only showed a significant increase in melanoma patients comparing to HC by a Mann-Whitney-Wilcoxon test (adjusted p-value = 0.0015, **Fig. 6**), confirming the results of the first dataset. Furthermore, a Kruskal-Wallis test showed that the effector-Treg-like cluster only was significantly different across HC and three disease stages (adjusted p-value = 0.0014, **Fig. 6a**). Pairwise comparisons of the 4 different statuses using a Mann-Whitney-Wilcoxon test showed that the effector-Treg-like cluster was significantly increased in stage III and stage IV comparing to HC (adjusted p-values = 0.0127 and 0.0325, respectively; **Fig. 6b**).

Discussion

This study has proposed to use a clustering approach to reveal the immunological profiles of $FOXP3^+$ T cells without invoking the concept of the immunosuppressive phenotype of $FOXP3^+$ T cells (19), and thereby to correlate them with disease phenotype or biological response. The current dogma of Treg (i.e. *the lineage perspective*(31)) considers that a Treg population can be defined only when it has been shown to be unique and distinct from other T cells and stable as a lineage (19). To be compatible with the lineage perspective, the threshold level of FOXP3 for defining the effector-Treg subpopulation by manual gating has been determined based on the result of a suppressive assay (9), the gold standard for assessing the immunosuppressive function of Treg (32, 33). Alarmingly, however, recent studies indicate that the "suppressive activity" by the assay is mostly explained by the absorption of IL-2 in the culture by CD25 (IL-2 receptor α chain) on the surface of anergic Treg (34, 35). Since FOXP3 has positive correlations with CD25 and anergy (36), it is not surprising that $FOXP3^{high}CD45RO^+CD25^{high}$ T cells (i.e. effector-Treg by manual gating) show a high suppressive activity in the *in vitro* assay.

The proposed method revealed that the effector-Treg-like cluster only was significantly increased in melanoma patients in both of the datasets, suggesting that this cluster captured an important immunological feature. These results encourage the computational clustering approach to reveal the immunological features of T cells in tumour-bearing patients. On the other hand, the computationally identified effector-Treg-like cluster included some memory-phonotype $CD45RO^+FOXP3^{low}$ non-Treg cells by manual gating (Fig. 3e, 3f, 4a-4d) (8, 9). While the result is difficult to be interpreted by the lineage perspective, our recently proposed model of Treg and FOXP3, *feedback control perspective* (31), may be useful for reconciling the results in this study with findings on Treg in the literature. Under this new perspective, the increase of the effector-Treg-like cluster in melanoma patients is

interpreted that FOXP3 was more frequently induced in tumour-bearing patients as a consequence of antigen recognition and a negative feedback mechanism of T cell activation, which also explains the memory phenotype of the effector-Treg-like cluster. Interestingly, the second dataset analysis showed that the effector-Treg-like cluster was increased in higher stages of melanoma (stage III and IV, Fig. 6), which is interpreted by the new model that reactive $FOXP3^+$ cells accumulated in the immune system as a result of prolonged chronic stimulation by cancer cells. Although the present study is not conclusive as to which perspective should be used, certainly the new perspective allows a more flexible interpretation of clinical and experimental data, because it does not assume stable and distinct lineages but is more concerned with the dynamics at the cellular and molecular levels. In fact, a recent study demonstrated an extensive TCR overlap between $FOXP3^+$ cells and $CD25^+FOXP3^-$ cells at the site of inflammation (37), further confirming that T cells may dynamically change the expression of FOXP3, especially in disease conditions. Importantly, this view leads to a question whether the negative feedback mechanism is stronger than the positive feedback mechanism in melanoma patients, encouraging further investigations on the latter in future studies.

The proposed clustering method provided reproducible results between two independent datasets. Note that the first dataset is in FCS2.0 format, obtained by an analogue system, FACS Calibur, while the second dataset is in FCS3.0 format, and was obtained by a digital acquisition system, LSR Fortessa (38). While data normalisation method for such different datasets is yet to be established and is a big issue in flow cytometric data analysis (39), the present study encourages the use of the proposed method or similar clustering methods for the analysis of complex flow cytometric data. In our analysis, $CD4^+$ T cell selection was almost identical between the manual and the automatic approaches (Fig. 2a). While the manual gating commonly creates a lymphocyte gate using FSC and SSC, and subsequently identifies $CD4^+$ T cells (Fig. 1), our investigation showed that it was more efficient to identify $CD4^+$ T cells by a one-step HDDC clustering approach using all FSC, SSC and CD4. This indicates that $CD4^+$ T cells are the most distinct when all the three dimensions are used. In fact, it is a common practice in manual gating to use either FSC or SSC to create a CD4-gate (e.g. Fig. 1c). This result confirms that flow cytometric data analysis in a higher dimensional space enables more efficient analysis (23, 25), which is widely accepted but yet-to-be further demonstrated by addressing real immunological problems. On the other hand, there is a small discordance in $FOXP3^+$ T cells by the manual gating and by the automatic clustering (Fig. 2b, left panel), although the overall correlation was high ($r = 0.8947$, Fig. 4h). This small discordance may be related to either

or both of the inherent arbitrariness of the manual gating approach to the $FOXP3^+$ selection and the variations by k-means. Since FOXP3 is measured by intranuclear staining and therefore its autofluorescence background is high (Fig 1, (9, 40)), it is experimentally difficult to precisely determine the boundary between the negative cloud and positive cells without using FOXP3-deficient T cells from mutant humans (41), which are practically difficult to be obtained in most laboratories. In addition, even using such negative controls, it is not obvious where to set the boundary. This is why the manual gating approach visually identifies a 'groove' between the negative cloud and positive cells and sets it as a boundary for $FOXP3^+$ cells (Fig. 1d). On the other hand, k-means is a method to determine clusters by minimising the within-cluster sum of squares (42), and thus can be affected by the distribution of cells and also by outliers, which can introduce variations. The small discordance of the boundary for $FOXP3^+$ cells may have contributed to those of the non-Treg-like and naïve-Treg-like clusters as well, as both clusters faced the boundary of FOXP3 positive and negative cells, while it presumably did not directly affect the effector-Treg-like cluster (Fig. 3e). However, considering that the proposed automatic gating only identified a significant increase of the $FOXP3^+$ T cell subpopulation in melanoma patients, and that the manual gating did not produce any significant results on the analysed subpopulations including non-Treg and naïve-Treg, the immunological feature of melanoma patients most probably is in the cells with higher FOXP3 expressions, and the discordance in the FOXP3 boundary is probably not important in the setting. Yet, it is hoped that future studies will develop data analytic and modelling methods to better deal with the problem of where to set the boundary between negative and positive cells in continuous distributions.

Disclosures

The authors have no financial conflicts of interest.

Acknowledgements

We thank Ms A. Bradley for proofreading the manuscript.

Reference

1. Sakaguchi, S., T. Yamaguchi, T. Nomura, and M. Ono. 2008. Regulatory T cells and immune tolerance. *Cell* 133: 775-787.
2. Gallimore, A., and A. Godkin. 2008. Regulatory T cells and tumour immunity - observations in mice and men. *Immunology* 123: 157-163.

3. Zou, W. 2006. Regulatory T cells, tumour immunity and immunotherapy. *Nat Rev Immunol* 6: 295-307.
4. Motz, G. T., and G. Coukos. 2013. Deciphering and reversing tumor immune suppression. *Immunity* 39: 61-73.
5. deLeeuw, R. J., S. E. Kost, J. A. Kakal, and B. H. Nelson. 2012. The prognostic value of FoxP3$^+$ tumor-infiltrating lymphocytes in cancer: a critical review of the literature. *Clinical cancer research : an official journal of the American Association for Cancer Research* 18: 3022-3029.
6. Wang, F., X. Jing, G. Li, T. Wang, B. Yang, Z. Zhu, Y. Gao, Q. Zhang, Y. Yang, Y. Wang, P. Wang, and Z. Du. 2012. Foxp3$^+$ regulatory T cells are associated with the natural history of chronic hepatitis B and poor prognosis of hepatocellular carcinoma. *Liver Int* 32: 644-655.
7. Salama, P., M. Phillips, F. Grieu, M. Morris, N. Zeps, D. Joseph, C. Platell, and B. Iacopetta. 2009. Tumor-infiltrating FOXP3$^+$ T regulatory cells show strong prognostic significance in colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 27: 186-192.
8. Fujii, H., A. Arakawa, A. Kitoh, M. Miyara, M. Kato, S. Kore-eda, S. Sakaguchi, Y. Miyachi, M. Tanioka, and M. Ono. 2011. Perturbations of both nonregulatory and regulatory FOXP3$^+$ T cells in patients with malignant melanoma. *Br J Dermatol* 164: 1052-1060.
9. Miyara, M., Y. Yoshioka, A. Kitoh, T. Shima, K. Wing, A. Niwa, C. Parizot, C. Taflin, T. Heike, D. Valeyre, A. Mathian, T. Nakahata, T. Yamaguchi, T. Nomura, M. Ono, Z. Amoura, G. Gorochov, and S. Sakaguchi. 2009. Functional delineation and differentiation dynamics of human CD4$^+$ T cells expressing the FoxP3 transcription factor. *Immunity* 30: 899-911.
10. Buckner, J. H., and S. F. Ziegler. 2008. Functional analysis of FOXP3. *Ann N Y Acad Sci* 1143: 151-169.
11. Vukmanovic-Stejic, M., Y. Zhang, J. E. Cook, J. M. Fletcher, A. McQuaid, J. E. Masters, M. H. Rustin, L. S. Taams, P. C. Beverley, D. C. Macallan, and A. N. Akbar. 2006. Human CD4$^+$ CD25^{hi} Foxp3$^+$ regulatory T cells are derived by rapid turnover of memory populations in vivo. *J Clin Invest* 116: 2423-2433.
12. Seddiki, N., B. Santner-Nanan, J. Martinson, J. Zaunders, S. Sasson, A. Landay, M. Solomon, W. Selby, S. I. Alexander, R. Nanan, A. Kelleher, and B. Fazekas de St Groth. 2006. Expression of interleukin (IL)-2 and IL-7 receptors discriminates between human regulatory and activated T cells. *J Exp Med* 203: 1693-1700.
13. Valmori, D., A. Merlo, N. E. Souleimanian, C. S. Hesdorffer, and M. Ayyoub. 2005. A peripheral circulating compartment of natural naïve CD4 Tregs. *J Clin Invest* 115: 1953-1962.
14. Hall, B. M., N. D. Verma, G. T. Tran, and S. J. Hodgkinson. 2011. Distinct regulatory CD4+T cell subsets; differences between naïve and antigen specific T regulatory cells. *Curr Opin Immunol* 23: 641-647.
15. Feuerer, M., J. A. Hill, D. Mathis, and C. Benoist. 2009. Foxp3$^+$ regulatory T cells: differentiation, specification, subphenotypes. *Nat Immunol* 10: 689-695.
16. Matsuki, F., J. Saegusa, K. Nishimura, Y. Miura, M. Kurosaka, S. Kumagai, and A. Morinobu. 2014. CD45RA$^-$Foxp3^{low} non-regulatory T cells in the CCR7$^-$CD45RA$^-$ CD27$^+$CD28$^+$ effector memory subset are increased in synovial fluid from patients with rheumatoid arthritis. *Cell Immunol* 290: 96-101.
17. Marwaha, A. K., S. Q. Crome, C. Panagiotopoulos, K. B. Berg, H. Qin, Q. Ouyang, L. Xu, J. J. Priatel, M. K. Levings, and R. Tan. 2010. Cutting edge: Increased IL-17-secreting T cells in children with new-onset type 1 diabetes. *J Immunol* 185: 3814-3818.

18. Jacobs, J. F., S. Nierkens, C. G. Figdor, I. J. de Vries, and G. J. Adema. 2012. Regulatory T cells in melanoma: the final hurdle towards effective immunotherapy? *Lancet Oncol* 13: e32-42.
19. Abbas, A. K., C. Benoist, J. A. Bluestone, D. J. Campbell, S. Ghosh, S. Hori, S. Jiang, V. K. Kuchroo, D. Mathis, M. G. Roncarolo, A. Rudensky, S. Sakaguchi, E. M. Shevach, D. A. Vignali, and S. F. Ziegler. 2013. Regulatory T cells: recommendations to simplify the nomenclature. *Nat Immunol* 14: 307-308.
20. Maecker, H. T., J. P. McCoy, and R. Nussenblatt. 2012. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol* 12: 191-200.
21. Rossin, E., T. I. Lin, H. J. Ho, S. J. Mentzer, and S. Pyne. 2011. A framework for analytical characterization of monoclonal antibodies based on reactivity profiles in different tissues. *Bioinformatics* 27: 2746-2753.
22. Mosmann, T. R., I. Naim, J. Rebhahn, S. Datta, J. S. Cavenaugh, J. M. Weaver, and G. Sharma. 2014. SWIFT-scalable clustering for automated identification of rare cell populations in large, high-dimensional flow cytometry datasets, part 2: biological evaluation. *Cytometry A* 85: 422-433.
23. Ge, Y., and S. C. Sealfon. 2012. flowPeaks: a fast unsupervised clustering for flow cytometry data via K-means and density peak finding. *Bioinformatics* 28: 2052-2058.
24. Maecker, H. T., A. Rinfret, P. D'Souza, J. Darden, E. Roig, C. Landry, P. Hayes, J. Birungi, O. Anzala, M. Garcia, A. Harari, I. Frank, R. Baydo, M. Baker, J. Holbrook, J. Ottinger, L. Lamoreaux, C. L. Epling, E. Sinclair, M. A. Suni, K. Punt, S. Calarota, S. El-Bahi, G. Alter, H. Maila, E. Kuta, J. Cox, C. Gray, M. Altfeld, N. Nougarede, J. Boyer, L. Tussey, T. Tobery, B. Bredt, M. Roederer, R. Koup, V. C. Maino, K. Weinhold, G. Pantaleo, J. Gilmour, H. Horton, and R. P. Sekaly. 2005. Standardization of cytokine flow cytometry assays. *BMC Immunol* 6: 13.
25. Cron, A., C. Gouttefangeas, J. Frelinger, L. Lin, S. K. Singh, C. M. Britten, M. J. Welters, S. H. van der Burg, M. West, and C. Chan. 2013. Hierarchical modeling for rare event detection and cell subset alignment across flow cytometry samples. *PLoS Comput Biol* 9: e1003130.
26. Pyne, S., X. Hu, K. Wang, E. Rossin, T. I. Lin, L. M. Maier, C. Baecher-Allan, G. J. McLachlan, P. Tamayo, D. A. Hafler, P. L. De Jager, and J. P. Mesirov. 2009. Automated high-dimensional flow cytometric data analysis. *Proc Natl Acad Sci U S A* 106: 8519-8524.
27. Donnenberg, D., and V. Donnenberg 2008. Understanding Clinical Flow Cytometry. In *Handbook of Human Immunology*. M. O'Gorman, and A. Donnenberg, eds. CRC Press.
28. Bergé, L., C. Bouveyron, and S. Girard. 2012. HDclassif: An R Package for Model-Based Clustering and Discriminant Analysis of High-Dimensional Data. *Journal of Statistical Software* 46: 1-29.
29. R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/
30. Husson, F., S. Le, and J. Pages. 2011. Clustering. In *Explanatory Multivariate Analysis by Example Using R*. CRC Press. 169-204.
31. Ono, M., and R. J. Tanaka. 2015. Controversies concerning thymus-derived regulatory T cells: fundamental issues and a new perspective. *Immunol Cell Biol*.
32. Shevach, E. M. 2000. Regulatory T cells in autoimmmunity. *Annu Rev Immunol* 18: 423-449.
33. Sakaguchi, S. 2004. Naturally arising $CD4^+$ regulatory T cells for immunologic self-tolerance and negative control of immune responses. *Annu Rev Immunol* 22: 531-562.
34. Busse, D., M. de la Rosa, K. Hobiger, K. Thurley, M. Flossdorf, A. Scheffold, and T. Hofer. 2010. Competing feedback loops shape IL-2 signaling between helper and regulatory T lymphocytes in cellular microenvironments. *Proc Natl Acad Sci U S A* 107: 3058-3063.

35. Smith, K. A., and Z. Popmihajlov. 2008. The quantal theory of immunity and the interleukin-2-dependent negative feedback regulation of the immune response. *Immunol Rev* 224: 124-140.
36. Ziegler, S. F. 2006. FOXP3: of mice and men. *Annu Rev Immunol* 24: 209-226.
37. Bending, D., E. Giannakopoulou, H. Lom, and L. R. Wedderburn. 2015. Synovial Regulatory T Cells Occupy a Discrete TCR Niche in Human Arthritis and Require Local Signals to Stabilize FOXP3 Protein Expression. *J Immunol*.
38. Seamer, L. C., C. B. Bagwell, L. Barden, D. Redelman, G. C. Salzman, J. C. Wood, and R. F. Murphy. 1997. Proposed new data file standard for flow cytometry, version FCS 3.0. *Cytometry* 28: 118-122.
39. Le Meur, N. 2013. Computational methods for evaluation of cell-based data assessment--Bioconductor. *Current opinion in biotechnology* 24: 105-111.
40. Ono, M., J. Shimizu, Y. Miyachi, and S. Sakaguchi. 2006. Control of autoimmune myocarditis and multiorgan inflammation by glucocorticoid-induced TNF receptor family-related protein(high), Foxp3-expressing $CD25^{+}$ and $CD25^{-}$ regulatory T cells. *J Immunol* 176: 4748-4756.
41. McMurchy, A. N., J. Gillies, M. C. Gizzi, M. Riba, J. M. Garcia-Manteiga, D. Cittaro, D. Lazarevic, S. Di Nunzio, I. S. Piras, A. Bulfone, M. G. Roncarolo, E. Stupka, R. Bacchetta, and M. K. Levings. 2013. A novel function for FOXP3 in humans: intrinsic regulation of conventional T cells. *Blood* 121: 1265-1275.
42. Aghaeepour, N., R. Nikolic, H. H. Hoos, and R. R. Brinkman. 2011. Rapid cell population identification in flow cytometry data. *Cytometry Part A* 79A: 6-13.

Figure Legends

Fig. 1. Manual gating approach to classify the FOXP3$^+$ T cell subpopulations.

The manual gating approach to identify and classify FOXP3$^+$ T cell subpopulations is depicted. (**a**) Representative flow cytometric data showing the three subsets of FOXP3$^+$ T cells in PBMC: CD45RO$^+$FOXP3^{high} effector Treg; CD45RO$^-$FOXP3^{low} naïve Treg; and CD45RO$^+$FOXP3^{low} non-regulatory T cells (non-Treg). These subpopulations are classified in a sequential manner using the following three gates: (**b**) the lymphocyte gate on the window displaying FSC and SSC; (**c**) the CD4$^+$ gate using CD4 and SSC; (**d**) the FOXP3$^+$ gate using FOXP3 and CD45RO; and (**e**) the FOXP3$^+$ T cell subpopulation gate using FOXP3 and CD45RO. The level of FOXP3^{high} and FOXP3^{low} cells was determined so that CD45RO$^-$FOXP3^{high} cells were < 0.2% of CD4$^+$ T cells.

Fig. 2. Automatic gating of FOXP3$^+$CD4$^+$ T cells.

Three clustering methods were compared for identifying FOXP3$^+$CD4$^+$ T cells: (1) CD4$^+$ T cell selection by High-Dimensional Data Clustering (HDDC), followed by FOXP3$^+$ T cell selection by k-means ("HK clustering"); CD4$^+$ T cell selection by HDDC, followed by FOXP3$^+$ T cell selection by HDDC ("HH clustering"); and (3) FOXP3$^+$CD4$^+$ T cell selection by one step-HDDC ("one-step H clustering"). Various random number seeds were used to resample events from a flow cytometric data, and resampling was repeated 100 times for each random number seed, in order to address the robustness and the efficiency of the three clustering methods. Sensitivities and accuracies were calculated by assuming that the manual gating provides a gold standard. (**a - c**) Sensitivities and accuracies of HK and HH: (**a**) Sensitivities and accuracies for identifying CD4$^+$ T cells by HDDC (shared by HK and HH). (**b - c**) Sensitivities and accuracies of (**b**) k-means and (**c**) HDDC for identifying FOXP3$^+$ T cells from the identified CD4$^+$ T cell cluster (HK and HH, respectively). (**d**) Sensitivities and accuracies of HDDC for identifying FOXP3$^+$ T cells from all cells (one-step H). (**e - g**) Representative plots of automatically gated FOXP3$^+$CD4$^+$ T cells by the HK clustering method for (**e, f**) CD4$^+$ T cells and (**g**) FOXP3$^+$ T cells. The clustered cells are shown by black dots.

Fig. 3. Automatic clustering of FOXP3$^+$ CD4$^+$ T cell subpopulations.

K-means and HDDC were used for classifying the computationally clustered FOXP3$^+$ T cells into three subpopulations, and compared them for stability using a resampling approach, which was repeated 100 times. (**a - d**) Boxplots showing (**a**) the percentages and the mean fluorescence intensities (MFI) of (**b**) CD45RO, (**c**) CD25, or (**d**) FOXP3 of each FOXP3$^+$ T cell subcluster in CD4$^+$ T cells by either k-means or HDDC in the 100 resampled samples (i.e. HKK or HKH,

respectively). (**e - f**) Representative plots of FOXP3$^+$ T cell subpopulations (**e**) by the HKK clustering (automatic) and (**f**) by manual gating.

Fig. 4. Comparison of the FOXP3$^+$ T cell clusters/populations identified by the automatic and manual gating approaches.

(**a-d**) The automatic (the HKK clustering) and manual gating approaches were compared by Spaghetti plots of the percentage of the cells that were classified as (**a**) effector-Treg(-like), (**b**) non-Treg(-like) and (**c**) naïve-Treg(-like), and (**d**) all FOXP3$^+$ T cells, in all samples including healthy controls and melanoma patients. (**e-h**) Scatter plots showing the percentages of each FOXP3$^+$ T cell cluster/population by automatic (the HKK clustering) and manual gating: (**e**) effector-Treg(-like), (**f**) non-Treg(-like) and (**g**) naïve-Treg(-like), and (**h**) all FOXP3$^+$ T cells. Closed and open circles represent melanoma and healthy control samples, respectively. All percentages are in CD4$^+$ T cells. Pearson's correlation coefficient (r) was calculated using all the samples for each cluster.

Fig. 5. Effector-Treg-like cluster in HC and melanoma patients by the automatic HKK clustering or manual gating.

(**a**) Boxplot showing %(*effector-Treg-like*)/CD4 of HC and melanoma patients by the automatic HKK clustering. * Adjusted $p < 0.05$. (**b**) Boxplot showing %(*effector-Treg*)/CD4 of HC and melanoma patients by manual gating.

Fig. 6. Application of the automatic gating approach to an independent dataset.

The established automatic gating method, the HKK clustering, was applied to an independent dataset (the second dataset, see **Table I**). (**a**) Boxplots showing %(*effector-Treg-like*)/CD4 in healthy control (HC) and melanoma patients (*** adjusted $p = 0.0015$). (**b**) Boxplots showing %(*effector-Treg-like*)/CD4 in healthy control (HC) and different disease stages of melanoma patients (I-II, III, or IV). A Kruskal-Wallis test showed that the percentages were significantly different across different statuses (adjusted $p = 0.0014$). Pairwise comparisons were done by a Mann-Whitney-Wilcoxon test, showing significant differences (adjusted $p < 0.05$) as indicated by (*).

Table I. Patient characteristics in the second dataset

Patient	Age(years)/sex	Type*	Stage	TNM classification	Previous treatment**
1	82/M	SSM	IA	pT1aN0M0	-
2	62/F	SSM	IB	T1bN0M0	-
3	52/M	SSM	IB	T2aN0M0	-
4	59/F	SSM	IB	pT2aN0M0	-
5	63/F	ALM	IIA	pT2bN0M0	-
6	60/M	ALM	IIC	pT4bN0M0	OP, Rec
7	78/F	SSM	IIC	T4bN0M0	-
8	64/M	SSM	IIIA	pT4N2aM0	-
9	53/F	SSM	IIIA	pT2aN1aM0	-
10	54/F	SSM	IIIB	pT1bN1aM0	OP, Rec
11	63/M	NM	IIIB	pT4bN1aM0	-
12	71/F	SSM	IIIB	T1aN1bM0	-
13	33/F	ALM	IIIB	pT3bN2aM0	-
14	78/M	ALM	IIIC	pT4bN3M0	-
15	71/M	ALM	IV	T4bN3M1c	OP, CT
16	62/M	MU	IV	pTxN0M1c	OP, CT
17	56/F	MU	IV	pTXN0M1c	OP, CT
18	39/F	SSM	IV	pTxNxM1c	OP, CT, RT
19	59/F	ALM	IV	TxN3M1c	OP

* Type: ALM, acral lentiginous melanoma; LMM, lentigo maligna melanoma; NM, nodular melanoma; MU, mucosal melanoma; SSM, superficial spreading melanoma

** OP: operation; CT: chemotherapy; RT: radiotherapy; Rec: recurrence.

FIGURE 1

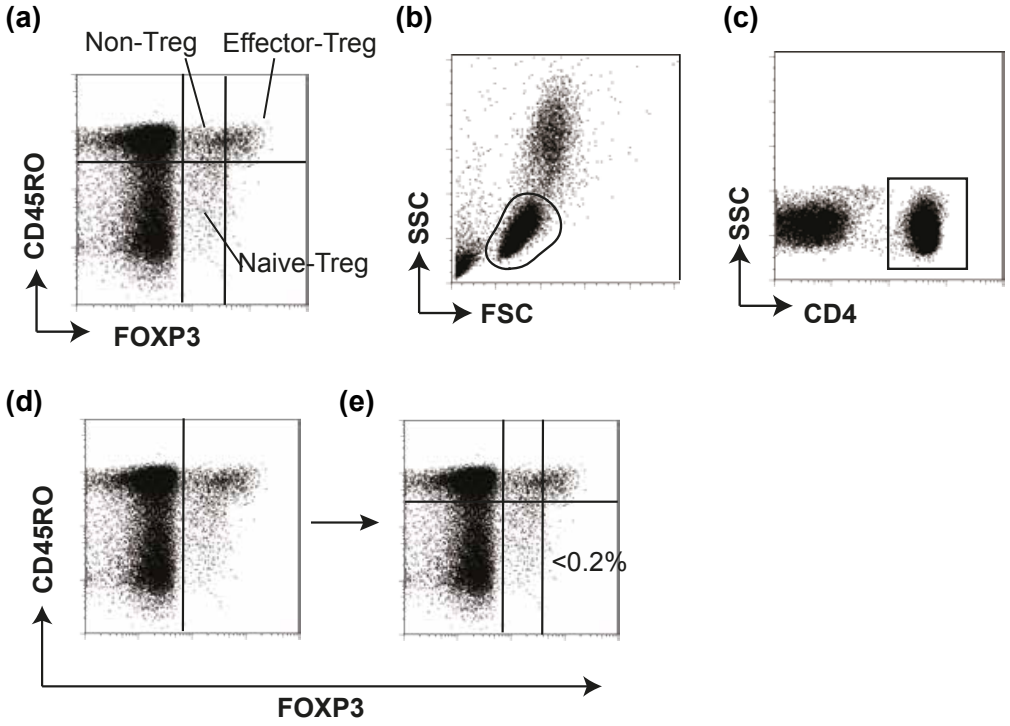

(a)
Non-Treg
Effector-Treg
Naive-Treg
CD45RO
FOXP3
(b)
SSC
FSC
(c)
SSC
CD4
(d)
CD45RO
(e)
<0.2%
FOXP3

Figure 2

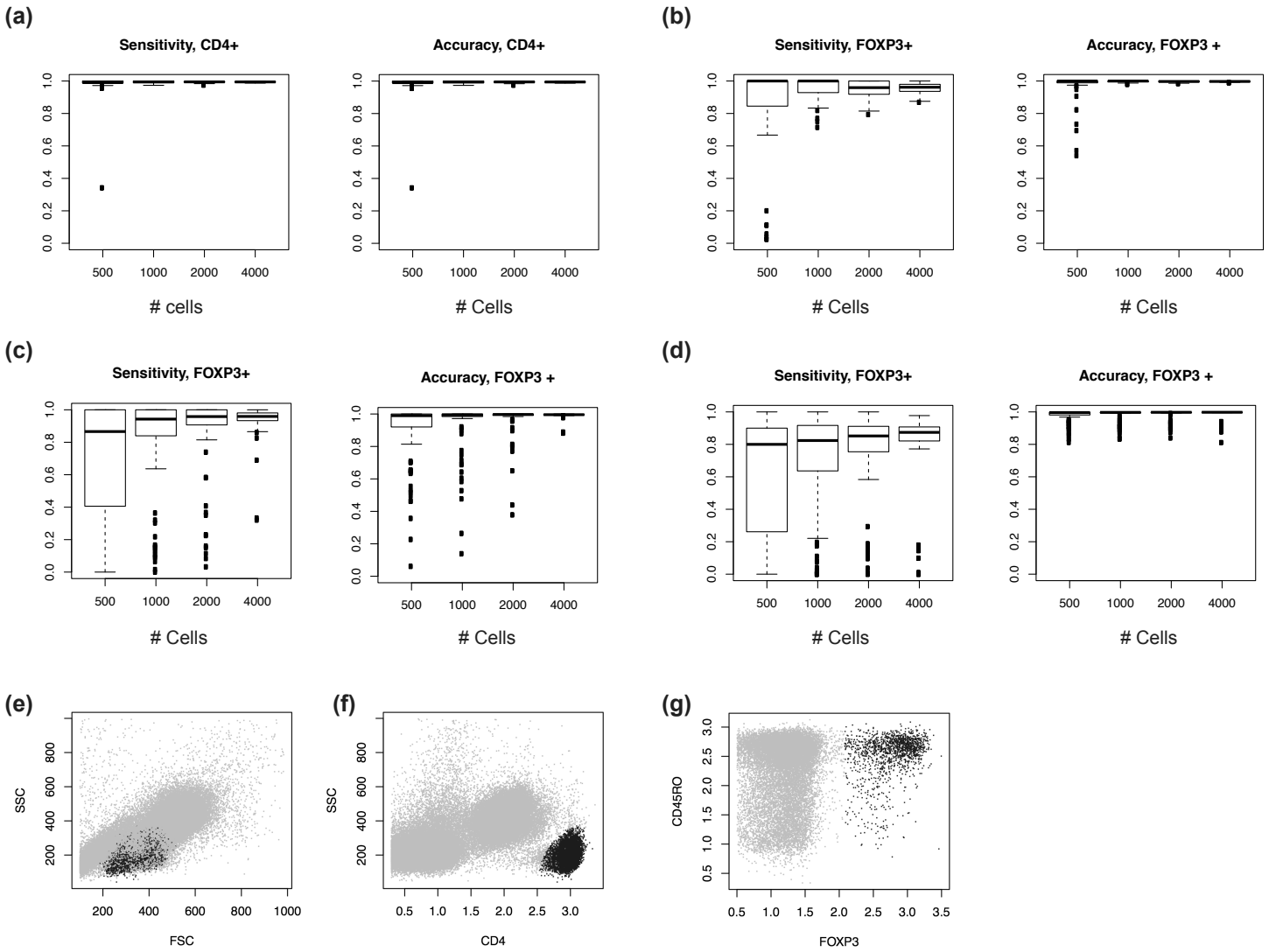

(a)
Sensitivity, CD4+
Accuracy, CD4+
cells
Cells
(b)
Sensitivity, FOXP3+
Accuracy, FOXP3 +
Cells
(c)
Sensitivity, FOXP3+
Accuracy, FOXP3 +
Cells
(d)
Sensitivity, FOXP3+
Accuracy, FOXP3 +
Cells
(e)
SSC
FSC
(f)
SSC
CD4
(g)
CD45RO
FOXP3

FIGURE 3

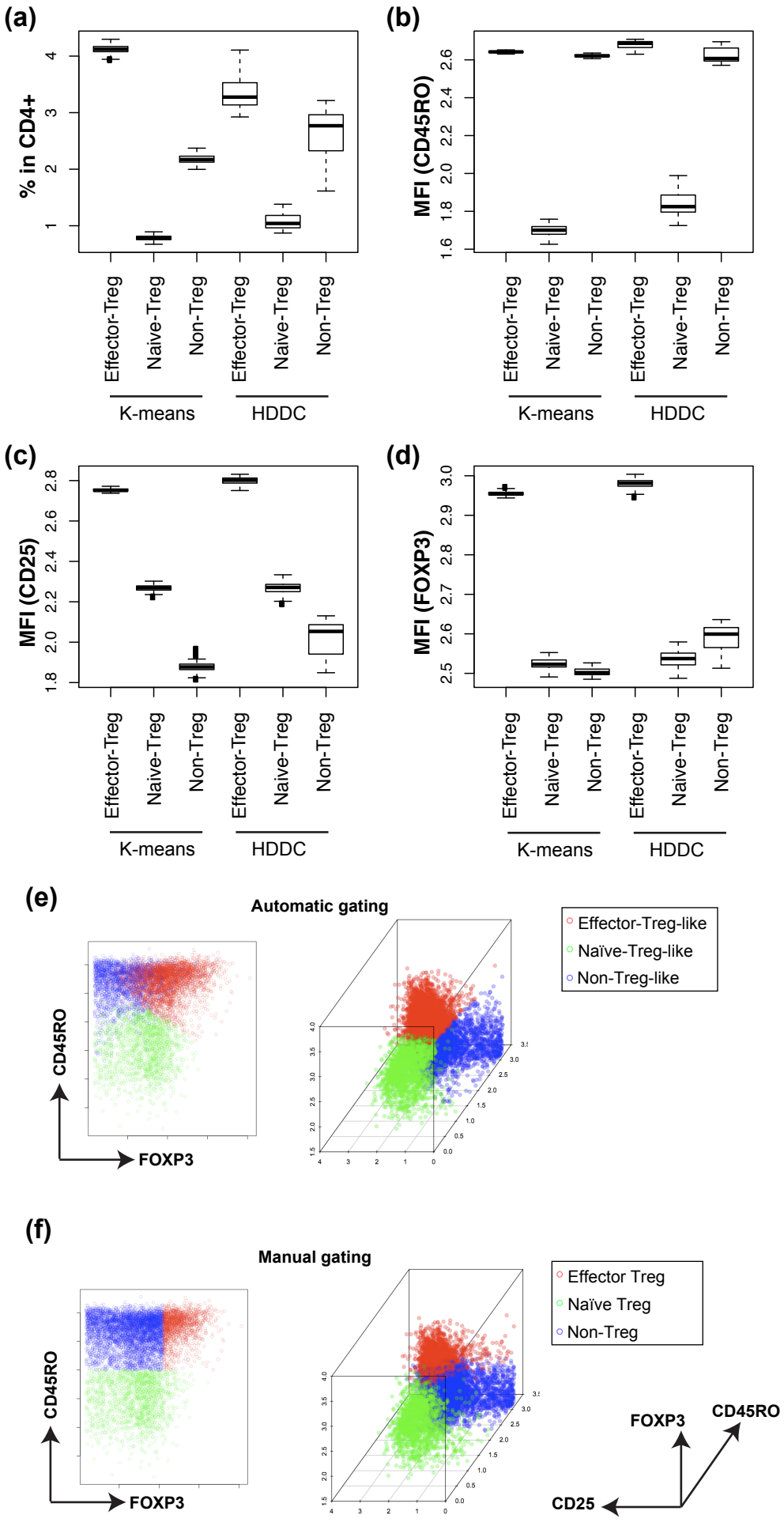

(a)
% in CD4+
(b)
MFI (CD45RO)
(c)
MFI (CD25)
(d)
MFI (FOXP3)
Effector-Treg
Naive-Treg
Non-Treg
K-means
HDDC
(e)
Automatic gating
Effector-Treg-like
Naïve-Treg-like
Non-Treg-like
CD45RO
FOXP3
(f)
Manual gating
Effector Treg
Naïve Treg
Non-Treg
CD25

FIGURE 4

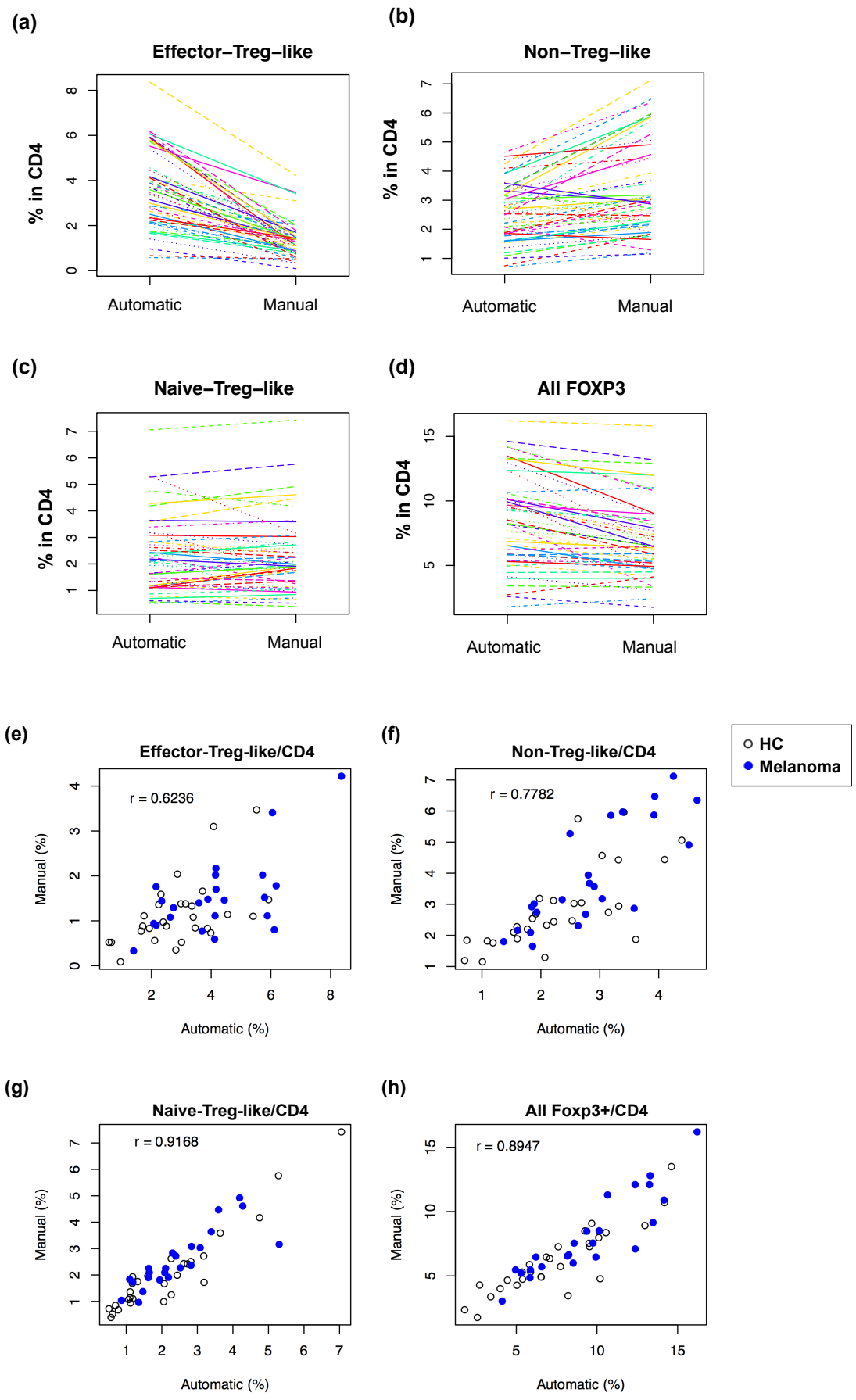

(a)
Effector−Treg−like
% in CD4
Automatic
Manual
(b)
Non−Treg−like
% in CD4
Automatic
Manual
(c)
Naive−Treg−like
% in CD4
Automatic
Manual
(d)
All FOXP3
% in CD4
Automatic
Manual
(e)
Effector-Treg-like/CD4
r = 0.6236
Manual (%)
Automatic (%)
(f)
Non-Treg-like/CD4
r = 0.7782
Manual (%)
Automatic (%)
HC
Melanoma
(g)
Naive-Treg-like/CD4
r = 0.9168
Manual (%)
Automatic (%)
(h)
All Foxp3+/CD4
r = 0.8947
Manual (%)
Automatic (%)

FIGURE 5

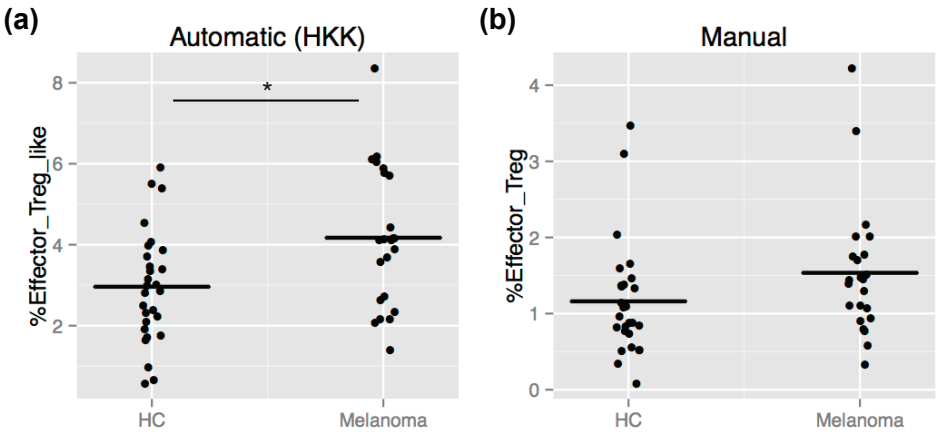
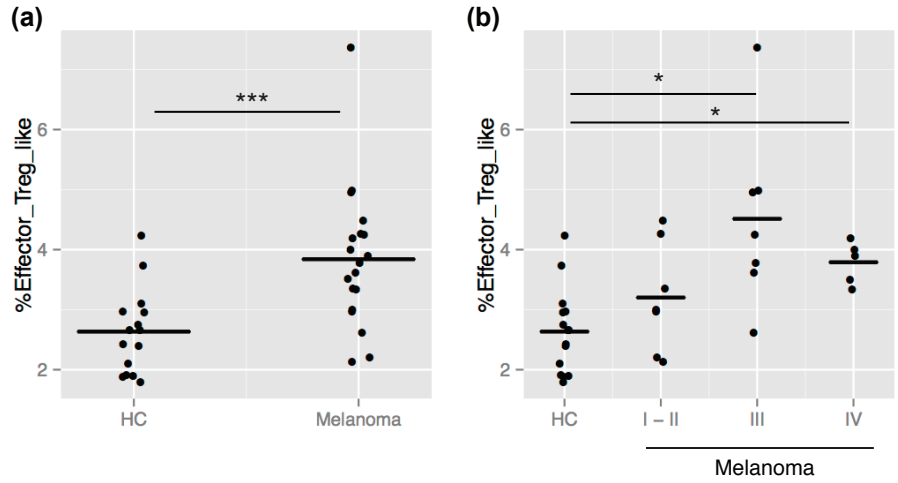

(a)
Automatic (HKK)
%Effector_Treg_like
8
6
4
2
*
HC
Melanoma
(b)
Manual
%Effector_Treg
4
3
2
1
0
HC
Melanoma

FIGURE 6


(a)
%Effector_Treg_like
6
4
2
***
HC
Melanoma
(b)
%Effector_Treg_like
6
4
2
*
*
HC
I – II
III
IV
Melanoma