



### The Journal of Immunology

This information is current as of December 16, 2015.

### Categorical Analysis of Human T Cell Heterogeneity with One-Dimensional Soli-Expression by Nonlinear Stochastic Embedding

Yang Cheng, Michael T. Wong, Laurens van der Maaten and Evan W. Newell

*J Immunol* published online 14 December 2015 http://www.jimmunol.org/content/early/2015/12/11/jimmun ol.1501928

Supplementary Material	http://www.jimmunol.org/content/suppl/2015/12/11/jimmunol.150192 8.DCSupplemental.html
Subscriptions	Information about subscribing to <i>The Journal of Immunology</i> is online at: http://jimmunol.org/subscriptions
Permissions	Submit copyright permission requests at: http://www.aai.org/ji/copyright.html
Author Choice	Freely available online through <i>The Journal of Immunology</i> Author Choice option
Email Alerts	Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/cgi/alerts/etoc



# Categorical Analysis of Human T Cell Heterogeneity with One-Dimensional Soli-Expression by Nonlinear Stochastic Embedding

### Yang Cheng,\* Michael T. Wong,\* Laurens van der Maaten,<sup>†</sup> and Evan W. Newell\*

Rapid progress in single-cell analysis methods allow for exploration of cellular diversity at unprecedented depth and throughput. Visualizing and understanding these large, high-dimensional datasets poses a major analytical challenge. Mass cytometry allows for simultaneous measurement of >40 different proteins, permitting in-depth analysis of multiple aspects of cellular diversity. In this article, we present one-dimensional soli-expression by nonlinear stochastic embedding (One-SENSE), a dimensionality reduction method based on the *t*-distributed stochastic neighbor embedding (*t*-SNE) algorithm, for categorical analysis of mass cytometry data. With One-SENSE, measured parameters are grouped into predefined categories, and cells are projected onto a space composed of one dimension for each category. In contrast with higher-dimensional *t*-SNE, each dimension (plot axis) in One-SENSE has biological meaning that can be easily annotated with binned heat plots. We applied One-SENSE to probe relationships between categories of human T cell phenotypes and observed previously unappreciated cellular populations within an orchestrated view of immune cell diversity. The presentation of high-dimensional cytometric data using One-SENSE showed a significant improvement in distinguished T cell diversity compared with the original *t*-SNE algorithm and could be useful for any high-dimensional dataset. *The Journal of Immunology*, 2016, 196: 000–000.

he development of fluorescence-based flow cytometry (FACS) (1) and the ability to probe single-cell protein expression with high throughput has been instrumental in laying the foundations of cellular immunology. More recently, high-dimensional polychromatic flow cytometry (2), mass cytometry (3), and high-throughput single-cell transcriptomics (4–6) approaches allow for identification of multiple subpopulations of cells and the ability to probe relationships between expression levels of large numbers of proteins or genes simultaneously (7). Hierarchical subgating based on biaxial plots offers a straightforward approach to analyze flow cytometry data, but this becomes impractical when interpreting the immense detail generated from mass cytometry (8) or other high-dimensional approaches.

To circumvent difficulties in visualizing high-dimensional data (9), numerous platforms based on clustering algorithms and/or

The online version of this article contains supplemental material.

This article is distributed under The American Association of Immunologists, Inc., Reuse Terms and Conditions for Author Choice articles.

Copyright © 2015 by The American Association of Immunologists, Inc. 0022-1767/15/\$30.00

dimensionality reduction have been developed to dissect mass cytometry (10-14) and other single-cell analysis data (4, 5). Principal component analysis (PCA) is a widely used dimensionreduction technique that constructs new summary parameters by linearly combining all data parameters to maximally explain variance in the data. Other methods that account for nonlinear relationships between parameters can analyze data with considerably higher resolution, allowing segregation of rare or subtly distinct populations. One such method, called t-distributed stochastic neighbor embedding (t-SNE) (9), performs exceptionally well on mass cytometry data (10, 12, 14). t-SNE performs pairwise comparisons of all events and maps them in a low dimensional space, optimally arranging similar events nearby and dissimilar events farther away. One major limitation of t-SNE and other nonlinear dimensionality reduction methods is that the values on the axes of the plots are arbitrary and have no intrinsic meaning. In particular, the function that t-SNE minimizes is invariant under rotations of the low-dimensional map, which implies that t-SNE visualizations can be arbitrarily rotated (15). Furthermore, even arbitrary directions in the visualizations have no meaning in the sense that they do not consistently indicate the same change in the underlying parameters. Thus, although the relative placement of cells by t-SNE is meaningful in that nearby events are phenotypically similar, understanding the relationships between the cellular arrangement and the underlying parameters can be tedious and labor-intensive.

A major goal of high-dimensional analysis of cells is to understand the relationships among various conceptual aspects of cellular biology. One lasting paradigm of human T cell immunology is based on experiments identifying the critical relationships between surface marker expression (e.g., CD45RA and CCR7) and functional properties (e.g., cytokines and cytotoxicity) of cells. These experiments created commonly used working definitions of naive, memory, effector, and terminally differentiated subtypes of T cells (16, 17), along with the addition of other differentiation markers such as CD127 and KLRG-1(18–20). With the advent of

<sup>\*</sup>Singapore Immunology Network, Agency of Science, Technology and Research, Singapore 138648; and <sup>†</sup>Pattern Recognition Laboratory, Delft University of Technology, 2628 CD Delft, the Netherlands

ORCIDs: 0000-0002-5906-7694 (Y.C.); 0000-0002-2889-243X (E.W.N.).

Received for publication August 31, 2015. Accepted for publication November 5, 2015.

Address correspondence and reprint requests to Dr. Evan W. Newell, Singapore Immunology Network, Agency for Science, Technology and Research, 8A Biomedical Grove, Immunos, 03-06, Singapore 138648. E-mail address: evan\_newell@immunol.a-star. edu.sg

Abbreviations used in this article: BABE, bromoacetamidobenzyl-EDTA; CRAN, Comprehensive R Archive Network; cRPMI, complete RPMI media; CyTOF, cytometry by time-of-flight; MAIT, mucosal-associated invariant T; One-SENSE, onedimensional soli-expression by nonlinear stochastic embedding; PCA, principal component analysis; Treg, regulatory T cell; *t*-SNE, *t*-distributed stochastic neighbor embedding.

mass cytometry, many more markers can be incorporated to better describe the differentiation state of cells, along with numerous indicators of T cell function (11), trafficking profiles (21), or

# Materials and Methods *PBMCs*

regulatory capacity (22).

Human PBMCs were obtained from three healthy donors within the Singapore Immunology Network, under the Institution Review Board (IRB) regulations. Whole blood was drawn and PBMCs were isolated using Ficolldensity gradient (Ficoll-Paque PLUS; GE Healthcare). Cells were cryopreserved in 90% FBS + 10% DMSO before use in experiments.

#### Cell stimulation, staining, and mass cytometry

PBMCs were thawed, washed, and rested in complete RPMI media (cRPMI, 10% FBS, 1% penicillin/streptomycin/L-glutamine, 1% 1M HEPES, and 0.1% 2-ME) overnight at 37°C in 24-well tissue culture plates (BD Falcon). On the next day, cells were washed with cRPMI and plated at  $15 \times$ 10<sup>6</sup> cells/ml in 96-well round-bottom plates (BD Falcon) in the presence of 150 ng/ml PMA, 1  $\mu$ M ionomycin, brefeldin A (eBioscience), monensin (eBioscience), and 0.5  $\mu$ g/ml anti-CD107a for 6 h at 37°C.

Cells were washed once with cRPMI and once with CyFACS buffer (2% FBS + 2 mM EDTA + 0.05% sodium azide in PBS) after stimulation. Cells were incubated with 200  $\mu$ M cisplatin (Sigma) on ice for 5 min to measure viability. For functional assessment of CD8<sup>+</sup> T cells, cells were washed once with CyFACS, then stained with a primary Ab mixture on ice for 30 min (Supplemental Table I). After incubation, cells then were washed once with CyFACS, once with PBS, and fixed with 2% PFA (paraformaldehyde; Electron Microscopy Sciences) overnight at 4°C. On the second day, cells were incubated in permeabilization buffer (Biolegend) at room temperature for 10 mins and then stained with an intracellular Ab mixture at room temperature for another 45 min for subsequent cell barcoding.

For examination of CD4<sup>+</sup> "Tregness," cells were resuspended with FOXP3 fixation/permeabilization buffer (eBioscience) on ice for 30 min. After cells were fixed and permed, cells were washed once and stained with a biotinylated anti-FOXP3 Ab in perm buffer on ice for 30 mins. Cells were washed with perm buffer, and an intracellular Ab mixture was added into the cell suspension and cultured on ice for another 30 min. After the staining, cell were washed with PBS and fixed with 2% PFA overnight at 4°C, and permed again using perm buffer (Biolegend) on the next day.

For cell barcoding, 2 mM bromoacetamidobenzyl-EDTA (BABE; Dojindo) with 0.5 mM PbCl<sub>2</sub> was dissolved in HEPES buffer. After the permeabilization, each sample was given a unique dual combination of metal-barcodes (BABE-Pd-104, BABE-Pd-106, BABE-Pd-108, BABE-Pd-110) for 30 min on ice (Supplemental Table I). After barcoding, cells were washed with perm buffer and then cultured in CyFACS on ice for 5 min. Cells were washed once and labeled by Rhodium-103 DNA interchelator (Fluidigm) in 2% PFA at room temperature for 20 min. Cells were further washed once with CyFACS and twice with MilliQ water. Upon cytometry by time-of-flight (CyTOF) acquisition, all samples were mixed together and prepared at  $5 \times 10^5$  cells/ml in water. Two percent of EQ beads (Fluidigm) were mixed with cell suspension for quality control. All calibration and acquisition settings were performed according to Fluidigm's instructions.

#### Data analysis

After CyTOF acquisition, any zero values were randomized using a uniform distribution of values between zero and minus-one using a simple R script (as was the default operation of previous CyTOF software). Note that all other integer values obtained by the CyTOF are also randomized by default. The signal of each parameter was normalized based on the EQ beads (Fluidigm) as previously described (23). T cells were gated and debarcoded using Boolean gating in FlowJo (Tree Star) similar to that described previously (12). After gating on populations of interest (Supplemental Fig. 1A), each sample was exported for further dimensionality reduction analysis, including t-SNE and one-dimensional soliexpression by stochastic embedding (One-SENSE) using custom R scripts based on the "flowCore" and "Rtsne" [using Comprehensive R Archive Network (CRAN) R packages]. In R, all data were transformed using the "logicleTransform" function and w = 0.25, t = 16409, m = 4.5, a = 0 as input parameters to roughly match scaling historically used in FlowJo. For each set of analyses, each cell subset (CD4<sup>+</sup> and CD8<sup>+</sup> T cells) from each donor was randomly downsampled to an equal number of cell events to normalize the contribution between donors: total CD8<sup>+</sup> T cells, 34,000 events; functional CD8<sup>+</sup> T cells, 22,000 events; and functional CD4<sup>+</sup> regulatory T cells (Tregs), 6000 events. Functional CD8<sup>+</sup> T cells were defined by Boolean gating containing cells expressing any of the 15 functional markers (Supplemental Fig. 1B). Functional CD4<sup>+</sup> Treg-like cells were Boolean-gated on cells positive for either one of six Treg markers (T cells that expressed only CTLA-4<sup>+</sup> were excluded to separate effector cells) (Supplemental Fig. 1C).

The parameters that were used in t-SNE or One-SENSE analysis are indicated in Supplemental Table I. For One-SENSE analysis, manually selected categories of markers (Supplemental Table I) were used as input for the function and R package "Rtsne" set to map data into a single dimension for each category. These values were appended to fcs files for subsequent analysis using R or FlowJo. Two- or three-dimensional plots were then constructed based on values obtained for each category. To describe the meaning of each of these one-dimensional maps (and each axis of the resulting plots), we constructed histograms and/or heat plots based on cells residing in bins corresponding to small ranges of values for each axis. FlowJo software was used to construct histograms of markerpositive cells across the entire range of One-SENSE axes values using default binning. Positive gate of each marker was manually defined, and markers within the same category were combined to present a histogram plot with the assigned categorical dimension on x-axis. It represents the distribution of marker-positive cells in percentage on each "bin" of One-SENSE axis. Heat plots were constructed using R by calculating marker frequency (or median intensities) of cells in each 250 bins for each axis. These values were normalized using manually set upper and lower limits to account for differing levels of background staining of each marker.

#### One-SENSE analysis and workflow

After cell acquisition using CyTOFII (Fluidigm DVS), metal signals of measured parameters were randomized and normalized. Cells of interest (e.g.,  $CD8^+$  T cells) were gated using FlowJo (Supplemental Fig. 1A) and exported as a single FCS file for each sample. The files were further loaded into an R environment for downsampling and dimensionality reduction analysis, including *t*-SNE and One-SENSE. The source code of *t*-SNE was obtained (http://lvdmaaten.github.io/), with necessary adjustment of "logicleTransform" function (12). The technical procedures of One-SENSE workflow and required R packages are described as follows and in Fig. 1.

Note 1: We used R (http://www.r-project.org), a functional computing language and interactive environment, to implant the function ("downsampling" and "logicleTransform") and algorithms (*t*-SNE and One-SENSE), and generate data visualization. One-SENSE was developed using RStudio (http://www.rstudio.com), a user-friendly interface for R. Another platform that has similar function could be also applied for One-SENSE analysis, such as MATLAB.

Note 2: Downsampling function was conjoined and run before the dimensionality reduction algorithm, including *t*-SNE and One-SENSE. The numbers of random downsampling between samples were described as detailed earlier.

Note 3: The "flowCore" package for flow cytometric data structure was installed in R from Bioconductor (http://www.bioconductor.org). "logicleTransform" function was applied to match the scaling historically used in FlowJo, with input parameters of w = 0.25, t = 16409, m = 4.5, a = 0, as previously described (12). All other CRAN R packages, including "Rtsne," "vegan," "permute," and "lattice," were downloaded from CRAN (https://cran.r-project.org/) using RStudio.

Note 4: One-SENSE and *t*-SNE were run in parallel for comparison of their utility in T cell analysis. The methodology and mathematical description of *t*-SNE can be found (9) and downloaded (http://lvdmaaten.github.io/).

Note 5: Compared with *t*-SNE that takes all measured markers to calculate the pairwise distance of cells to embed cells in a two- or threedimensional space, One-SENSE performs one-dimensional *t*-SNE for each predefined category of T cell markers. The assignment of T cell markers was written into a csv file and further loaded onto One-SENSE script in R (see Supplemental Table I).

Note 6: Each T cell category was independently analyzed by onedimensional t-SNE, and a range of 250 bins was used to arrange cells in single dimension. The alignment of 250 bins was optimized arbitrarily for better visualization.

Note 7: One-SENSE subsequently generated the heat plots for each onedimensional *t*-SNE of predefined category by using function heatmap.2 of R package "gplots." The heat plot is composed of the assigned markers and represents their categorical expression. The numbers of bins for each heat plot correspond to its categorical one-dimensional *t*-SNE. Importantly, the presentation of categorical expression is purpose-driven and user-defined, which can be visualized in heat plot



FIGURE 1. One-SENSE analysis workflow. The workflow describes the step-by-step details for analyzing T cell categorical relationship using One-SENSE. See *Materials and Methods*.

or histogram. The histograms can be plotted by overlapping categorical markers using FlowJo.

Note 8: One-SENSE uses these "binned" one-dimensional *t*-SNE to project the categorical expressions of cells from each axis onto a twoor three-dimensional plot, which depends on how many categories were predefined. Consequently, exported heat plot images were aligned on each axis of One-SENSE plots by manually adjusting logicleTransform parameters to match the scaling used by FlowJo, to annotate the cellular markers and categorical relationship.

Note 9: The axes (or categorical dimensions) of One-SENSE can be used interchangeably for investigating T cells in a FlowJo platform.

Note 10: A three-dimensional One-SENSE image (Fig. 5, Supplementary Video 1) is created by using an R-written script loaded with "rgl" and "scatterplot3d" packages with three categorical dimensions.

Note 11: After the analysis, One-SENSE generates and exports the data as FCS file, providing the accessibility for further analysis using FlowJo or other FCS-compatible analytical software.

#### Neighborhood preservation analysis

A random sample of 10,000 total CD8<sup>+</sup> T cells (as plotted in Fig. 1) and all 37 relevant parameters (Supplemental Table I) were used for nearest-neighbor analysis. Nearest neighbors for each cell were identified using the "knnx. index" function part of the "FNN" R package, with k set to 100. This function was applied to each cell using raw 37-dimensional data and to each dimensionality-reduced dataset generated using one- to seven-dimensional *t*-SNE versus PCA. For each cell and in each dimensionality-reduced dataset, the neighborhood preservation ratio was determined by calculating the percentage of k nearest neighbors matching those identified in 37 dimensions as in the dimensionality-reduced dataset. The average neighborhood preservation ratio was also calculated for each dimensionality-reduced set by averaging the neighborhood preservations ratio over the entire data set.

#### Three-dimensional video of CD8<sup>+</sup> T cells in One-SENSE view

CD8<sup>+</sup> T cells analyzed by One-SENSE with three dimensions (differentiation, function and trafficking) were exported and loaded into an R environment. The R-written script with rgl package that supports threedimensional visualization was used to construct numerous threedimensional images using these three One-SENSE dimensions. The continuous image sequences were further combined by Sequimago (AppleScript) to generate three-dimensional video.

#### Results

#### The rationale of One-SENSE

In this article, we propose and demonstrate an approach that facilitates a type of categorical analysis that we called One-SENSE). "Soli" comes from the musical term referring to the entire section of an ensemble performing together as opposed to solo. One-SENSE measures cellular parameters assigned to manually predefined categories, and a one-dimensional map is constructed for each category using t-SNE (see Fig. 1; see Materials and Methods). An advantage of this approach is that each dimension (axis) is informative and can be annotated through the use of heat plots aligned in parallel to each axis, allowing for simultaneous visualization of two categories across a two-dimensional plot. The cellular occupancy of the resulting plots allows for direct assessment of the relationships between the categories. Although t-SNE is becoming widely used for its ability to map cells into two-dimensional space with high resolution (4, 5, 12), t-SNE also performs remarkably well even when mapping is restricted to a single dimension (15). To address this and quantitatively compare the performance of t-SNE in mapping high-dimensional mass cytometry data into lower dimensional projections, we performed a neighborhood preservation analysis (Fig. 2). In this analysis, the k-nearest neighbors were identified for each cell using 37-parameter mass cytometry data (using k = 100, which represents 1% of the 10,000 human



**FIGURE 2.** Neighborhood preservation analysis of *t*-SNE versus PCA used at varying degrees of dimensionality reduction. To quantitatively assess the performance of *t*-SNE compared with linear PCA at various mapping dimensionalities, we used a neighborhood preservation analysis strategy. CD8<sup>+</sup> T cells (10,000 random events sampled from 3 different donors as in Fig. 3) were used and 37 parameters for each cell (all parameters used for Figs. 3 and 4). For each cell, the nearest 100 neighbors were identified in 37 dimensions. The same cells were mapped into one to seven dimensional space using *t*-SNE or PCA. Based on these results the 100 nearest neighbors were also identified, and the ratio of these cells matching those identified using all 37 dimensions was calculated to assign a neighborhood preservation ratio. (**A**) Histograms of the neighborhood preservation ratio for this 10,000-cell dataset are plotted for the one- to four-dimensional *t*-SNE and PCA mappings. (**B**) Average neighborhood preservation ratios are plotted for one- to seven-dimensional *t*-SNE outperforms three-dimensional PCA on this dataset (in terms of neighborhood preservation), and that gains in performance coming from the use of ≥2 dimensions of *t*-SNE are modest compared with that achieved by just a single dimension. Specifically, *t*-SNE performance in one dimension is already at >70% of the maximal performance possible for higher dimensional *t*-SNE mapping of these data. Note though that *t*-SNE performance in >3-dimensional mappings can improve by increasing the number of degrees of freedom of the Student *t* kernel (15).

CD8<sup>+</sup> T cells used for this analysis, in Fig. 3). After *t*-SNE analysis (and mapping to various numbers of dimensions), the k-nearest neighbors were identified for the same cells, and the fraction matching those identified for the 37-dimensional dataset was calculated as the neighborhood preservation ratio for each cell. This analysis shows that one-dimensional *t*-SNE is already more effective than three-dimensional PCA and performs to a level that is ~78% of maximum for *t*-SNE at any number of dimensions (Fig. 2B; see *Materials and Methods*).

To illustrate the utility of One-SENSE for analysis of single-cell data, we applied a mass cytometry panel of Abs to probe 44 proteins expressed by human CD8<sup>+</sup> T cells (Fig. 3, Supplemental Table I) from three healthy donors to assess the relationships among differentiation state, functional capacity, and trafficking receptor profiles. After using several markers to identify CD8<sup>+</sup> T cells (Supplemental Fig. 1A, 1B), the remaining markers were allocated to three different categories. Markers such as CD45RA, CCR7, CD127, KLRG-1, CD28, and CD62L were linked to T cell differentiation; cytokine secretion and cytotoxic activity, including IFN- $\gamma$ , TNF- $\alpha$ , IL-2, and granzyme B, were grouped as function; chemokine receptors like CXCR5, CXCR3, and CCR5 involved in T cell homing were classified as trafficking (Supplemental Fig. 1, Supplemental Table I). Consequently, One-SENSE generates three dimensions for each category composed of assigned protein markers (Figs. 1, 3). To make One-SENSE method more accessible, we provide a detailed description and

workflow of One-SENSE analysis in Fig. 1 and *Materials and Methods*.

Using these categories, we analyzed the One-SENSE representation of total human CD8<sup>+</sup> T cells in terms of differentiation versus function to investigate alterations in T cell functionality that would be hypothesized to occur as cells differentiate (Fig. 3A). This approach was compared with traditional biaxial gating and with a two-dimensional t-SNE plot. From this analysis, a large number of T cell subsets with distinct combinations of differentiation marker expression versus functional profiles could be readily seen. The identities of each subset could be inferred without subsequent analysis by interpreting and annotating the information from each axis using aligned heat plots (Fig. 3A). It is noteworthy that the expression level of markers on each bin using heat plot is user-defined, which can be presented as frequency of marker-positive cell, average marker intensity (data not shown), or using histogram to indicate the distribution of marker-positive cells. Naive-like cells that expressed relatively high levels of CD45RA, CCR7, CD62L, and CD28 occupied one large section on the plot. As expected, these naive cells were limited in their ability to produce cytokines, except for a small subset of IL-8-producing cells. In contrast, cells with a memory/effector profile displayed a large range of functional diversity, and minor populations with distinct combinations of differentiation and functional characteristics were promptly observed, such as a helper-like (24, 25) population



**FIGURE 3.** Rationale and representation of One-SENSE. Human PBMCs isolated from three healthy donors were stained and acquired by mass cytometry for dimensionality reduction analysis. Total  $CD8^+$  T cells were exported and analyzed by *t*-SNE and One-SENSE in parallel. (**A**) Plots show comparison of protein marker visualization between *t*-SNE (Tutti-expression) and One-SENSE (Soli-expression) in different dimensions. Protein markers were assigned to various immunological categories based on previous studies. Shown is a One-SENSE view of  $CD8^+$  T cells displaying differentiation against function. Cells are aligned to a range of 250 bins (column of cells) on the categorical dimensions (axes) that are composed of cellular protein expression. Each categorical dimension is a one-dimensional *t*-SNE analysis (Fig. 1). Heat plots indicate the frequency of marker-positive cells in each bin. One-SENSE shows nonfunctional/ naive T cells are mainly located in the upper left corner of the map, whereas multifunctional effector and memory T cells have diverse combination clusters on the opposite area. Cellular features of clusters can be described by the coordination of *cis* and *trans* coexpression from each axis. (**B**) Comparison of One-SENSE and *t*-SNE on MAIT cells is shown (*top*). Histograms (*bottom*) of the indicated markers are shown for the gated populations of MAIT cells. (**C**) The frequency and ratio of MAIT 1 and MAIT 2 population identified between donors using One-SENSE.

coexpressing CD40L, IL-2, and ICOS. It is noteworthy that this approach provided immunologically meaningful segregation of cells with functional differences within the same phenotypic subsets. For example, coexpression of CD161<sup>+</sup>V $\alpha$ 7.2 <sup>+</sup> is used to identify mucosal-associated invariant T (MAIT) cells (identified by a distinct profile of differentiation-state markers). One-SENSE analysis showed a clear separation of MAIT cells harboring two distinct functional subsets, whereas there was limited separation on its *t*-SNE counterpart (Fig. 3B). Notably, this approach also allows for comparisons between individuals. In this case, we observed heterogeneity in the frequencies of total MAIT cells and a consistent ratio of the two functionally distinct subsets identified in this study (Fig. 3C).

# *Use of One-SENSE to assess the heterogeneity of human peripheral CD8<sup>+</sup> T cells*

Unlike the clear functional and transcriptional classification of CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells are often designated by their differentiation status instead of functionality. This may ignore the functional diversity of CD8<sup>+</sup> T cells. Thus, we focused our One-SENSE analysis on how the functional attributes of CD8<sup>+</sup> T cells might be associated with cell differentiation or trafficking. Subsequently, functional CD8<sup>+</sup> T cells were exported and analyzed by One-SENSE (see *Materials and Methods*). We observed numerous multifunctional T cell subsets residing on different columns of the function dimension, indicating the differential composition of these 15 functional markers between subsets (Fig. 4A, Supplemental Fig. 2A). Notably, the nomenclatures used in this study are only for the purpose of labeling observed subsets, but do not necessarily correspond to previously published T cell subsets. We named the most abundant CD8 subset as Tc1 (IFN- $\gamma^{+}$ TNF- $\alpha^{+}$ CD107a<sup>hi</sup>GrzB<sup>+</sup>), which expressed markers typical of CD8<sup>+</sup> T cells. Cells with the highest cytotoxicity (Perforin<sup>hi</sup>GrzA<sup>hi</sup>GrzB<sup>+</sup>CD38<sup>+</sup>) (Supplemental Fig. 2A, 2B) expressed the chemokine receptor CXCR1 (26), but low levels of IFN- $\gamma$  and TNF- $\alpha$ , which we labeled as specialized killer Tc2 cells. Tc3 cells displayed a similar profile as Tc1, but expressed low levels of CD107a and increased MIP-1 $\beta$ . Exclusive production of GM-CSF, along with augmented chemokine expression (MIP-1 $\alpha$ , MIP-1 $\beta$ , and IL-8), were deemed as Tc4 cells that we hypothesized to have an important role in immune cell recruitment (Fig. 4A, Supplemental Fig. 2A).

One-SENSE also highlights the presence of several rare subsets, including IL-2<sup>hi</sup>CD40L<sup>hi</sup>CCR4<sup>+</sup> helper-like (24, 25) and CD39<sup>+</sup> CTLA-4<sup>+</sup> Treg-like CD8<sup>+</sup> T cells (Fig. 4A, Supplemental Fig. 2A). These subsets were observed consistently between individuals (Supplemental Fig. 2C). To further demonstrate the advantages of One-SENSE, we used the functional subset Tc1 as an example. Although this subset was clustered on the same column of the function dimension, they were separated with the differentiation dimension based on differential expression of CD57 and CD45RO (Fig. 4B). In addition, each subset of Tc1 could be further expanded to five subpopulations by their localization profile using the trafficking dimension, suggesting differential states of T cell memory and mobility (21) (Fig. 4B, 4C). Notably, Tc1 cells differing in trafficking receptor profiles were poorly segregated by two- (Fig. 4C) or three-dimensional (Fig. 5) t-SNE analysis (Supplementary Videos 1, 2). Thus, this analysis demonstrates the utility of One-SENSE as a simple and effective way to explore the remarkable versatility of CD8<sup>+</sup> T cells (summarized in Fig. 4D).



**FIGURE 4.** Versatile differentiation and trafficking of functional peripheral Tc1-profiled CD8<sup>+</sup> T cells. Functional CD8<sup>+</sup> T cells were defined by cells expressing at least one functional marker (Supplemental Fig. 1, Supplemental Table I) and combined by Boolean gating before One-SENSE analysis. (**A**) Cells were clustered using three independent categorical dimensions (differentiation, function, and trafficking). The functional heterogeneity of CD8<sup>+</sup> T cells is observed by the function dimension on the One-SENSE map. Diverse patterns of differentiation and trafficking of functional CD8<sup>+</sup> T cells are shown by changing the *y* dimension. Numerous functional subsets are indicated using this method as shown. (**B**) Differentiation and trafficking status of CD8<sup>+</sup> T cells that have the same functional profile are further analyzed using One-SENSE. Three Tc1 subsets are observed on the differentiation dimension, whereas the trafficking dimension further segregates each of them into the other five subpopulations. (**C**) *t*-SNE view shows very minimum separation of these subpopulations. (**D**) Summarized One-SENSE plot represents the complex heterogeneity of the Tc1 subset.



**FIGURE 5.** Three-dimensional images of trafficking characteristics of Tc1-profiled CD8<sup>+</sup> T cells subsets. Comparison of three-dimensional One-SENSE and three-dimensional *t*-SNE analysis. Functional CD8<sup>+</sup> T cells were gated (*Materials and Methods*), and Tc1\_1 CD8<sup>+</sup> T cells were defined as shown in Fig. 4. Three-dimensional One-SENSE view of functional CD8<sup>+</sup> T cells reveals unprecedented degree of heterogeneity, whereas three-dimensional *t*-SNE poorly segregates these minor subpopulations. Cells in three-dimensional image of One-SNESE were projected from three different axes of categorical dimensions (Fig. 1). Five subpopulations of Tc1 subset 1 (Tc1\_1) are colored as shown here and in Supplementary Videos 1 and 2, for three-dimensional video of One-SESNE and *t*-SNE, respectively.

#### Diverse profiles of human Tregs and regulatory-like T cells

Tregs are critical regulators of the adaptive immune system and play especially important roles in autoimmune disease and tumor immunology. However, prior studies have used inconsistent definitions to identify these cells with suppressive activity, and the extent of phenotypic and functional heterogeneity is not clear (22, 27). Therefore, by focusing on cells that express one or more markers associated with suppressive activity (markers of "Tregness"; Supplemental Fig. 1C), we examined the relationships between these markers and markers of differentiation, function, and trafficking using One-SENSE (Fig. 6). Our data illustrate a diverse expression pattern of regulatory markers among seven apparent Treg-like functional subsets (Fig. 6A, Supplemental Fig. 3A). FOXP3 is predominant in three subsets (labeled Treg-like 1, 2, and 3), with various expression of CD25, CD39, and CTLA-4 (Fig. 6B) (28). Although they are distinct, each of these subsets appears to be quite homogenous with respect to all other markers probed (on the "All Others" axis). These three subsets fit best with canonical descriptions of Treg populations previously reported (22).

In contrast, the subsets labeled Treg-like 4 and 5 have minimal FOXP3 expression, and are both heterogeneous in their effector and trafficking marker expression (Fig. 6B, Supplemental Fig. 3B,

3C), which is hardly observed by *t*-SNE (Supplemental Fig. 3D). Lastly, cells with a FOXP3<sup>-</sup> (29) regulatory profile, which we labeled Treg-like 6 and 7, display elevated expression levels of IL-10 and LAG-3 (29), respectively (Fig. 6B). In summary, this analysis of Treg-like cells demonstrated how One-SENSE could be used to highlight and quickly describe the heterogeneity of cells expressing markers associated with suppressive activity. We anticipate that this analysis approach would be well suited for identifying populations of cells associated with immunological dysfunction, such as in the context of autoimmunity or cancer.

#### Discussion

Using example datasets, we demonstrate the utility of One-SENSE in uncovering the depth of T cell heterogeneity. One-SENSE uniquely provides users with the ability to assign multiple parameters to predefined categories, while preserving the essence of the t-SNE algorithm. Our data demonstrate how this approach can be used to intuitively visualize relationships within high-dimensional data and to test hypotheses regarding the existence of these relationships.

One of the major limitations when using dimensionality reduction analysis on mass cytometry data, including SPADE-, PCA-, and *t*-SNE-based algorithms, is the annotation of cell clusters. Because visualization of protein markers one by one on a *t*-SNE map is not ideal, describing the coexpression of two or more markers is even more difficult. Researchers have to subjectively anticipate the possible combinations of markers, which could lead to potential bias when considering unknown cell subsets and the heterogeneity of cells. One-SENSE provides an objective and effective systemic overview of marker annotation (including protein coexpression). It allows direct assessment between cellular properties and the observation of subtle differences within common cell subsets, as we demonstrated in this article using MAIT cells as an example.

Descriptions of human CD8<sup>+</sup> T cell subsets have mostly relied on markers associated with cell differentiation (e.g., CD45RA and CCR7) (16). However, cellular profiles of human CD8<sup>+</sup> T cells based on either cell differentiation markers or functional capacities are each highly complex using our unsupervised One-SENSE analysis, suggesting that the traditional definitions of human CD8<sup>+</sup> T cell subsets based exclusively on a few differentiation markers may not be sufficient. In contrast, coexpression of IFN- $\gamma$ , TNF- $\alpha$ , and IL-2 are cytokines often used to designate polyfunctional CD8<sup>+</sup> T cells, which have been widely known as Tc1 cells (30). Previous studies have also described other CD8<sup>+</sup> T cell functional subsets, such as IL-4-producing CD8<sup>+</sup> T cells (31, 32), CD8<sup>+</sup> Tregs (33, 34), and helper-like CD8<sup>+</sup> T cells (24, 25). However, the functional heterogeneity of CD8<sup>+</sup> T cells has not been systemically examined. This is likely limited by traditional experimental and analytical approaches, where coexpression of functional proteins is difficult to identify objectively. Using One-SENSE, we demonstrate the functional versatility of CD8<sup>+</sup> T cells by examining 15 different functional markers and their possible coexpression combinations with an unsupervised analytical approach. This is poorly brought out using traditional differentiationbased classification. From this, we observed at least six different functional CD8<sup>+</sup> T cell subsets, which is consistent across donors. Although we did not perform subsequent experiments to characterize observed functional CD8<sup>+</sup> T cell subsets, the existence of these subsets has been implied. For instance, Feau et al. and Frentsch et al. have described the helper function of IL-2- (24) and CD40L (25)-expressing CD8<sup>+</sup> T cells, respectively. Interestingly, data from Frentsch et al. showed a population of IL-2<sup>hi</sup> CD40L<sup>hi</sup>CD8<sup>+</sup> T cells, which is in line with our observation of



**FIGURE 6.** Diverse subpopulations of human peripheral Tregs.  $CD4^+$  T cells with Treg-related function (FOXP3, CD25, CD39, CTLA-4, IL-10, and LAG-3) were exported and combined by Boolean gating. (**A**) Cells were analyzed by One-SENSE using the "Tregness" dimension against all other markers. Several functional Treg-like subsets are revealed by One-SENSE and separated by *cis* coexpression of different functional Treg markers. (**B**) Two Treg-like subsets (4 and 5) can be further dissected into multiple subpopulations by other functional or localization markers.

IL- $2^{hi}$ CD40L<sup>hi</sup>CCR4<sup>+</sup>CD8<sup>+</sup> T cells with additional production of TNF- $\alpha$  and GM-CSF (Fig. 4A, Supplemental Fig. 2A), and can be readily detected using One-SENSE (Figs. 3A, 4A).

Tregs are widely known to possess remarkable plasticity and functional diversity (22, 35, 36), and recent focus on the balance of Tregs and pathogenic Th17 cells in autoimmune disease has drawn significant attention. One major difficulty in isolating this unique subset is the paucity and inconsistency of human Treg markers (35). The master transcription factor FOXP3 (27) has been used as the most reliable marker of bona-fide Tregs. However, some effector T cells transiently express FOXP3<sup>hi</sup> upon activation in an inflammatory environment (37, 38). It has also been shown that a fraction of Tregs may lose FOXP3 expression while acquiring expression of the suppressive cytokine IL-10 (29). One-SENSE revealed heterogeneous Treg subsets based on the categorical expression of differential cellular and Treg properties. For instance, Treg-like subset 4 expressed FOXP3<sup>lo</sup>CD127<sup>lo</sup>CD39<sup>+</sup>CTLA-4<sup>+</sup>, with four subpopulations separated by their various expressions of IL-2, TNF- $\alpha$ , and CLA. This heterogeneity within a single Treg subset is again difficult to distinguish in a traditional t-SNE map. Taken together, we demonstrate that One-SENSE could be useful for studying T cell heterogeneity, which can be applied to other Th cell subsets; for instance, the discrepant trafficking and memory profile of human peripheral follicular Th cells, as recently described in the context of HIV (39) and influenza infection (40, 41).

The aim of this study is to provide detailed and thorough description of One-SENSE, and to demonstrate its utility for the analysis of immune cell heterogeneity. In the process, a number of interesting cellular subsets can be identified. We showed that these subsets can also be quantified for each sample and then these frequencies can be compared between samples run together. We also demonstrated this by comparing the frequencies of various cell subsets across three different donors analyzed. Although we have used only three donors in this report, the distinct T cell subsets (functional  $CD8^+T$  cell and  $CD4^+T$ reg) identified by One-SENSE were consistently observed across donors, in both frequency and cellular protein composition. In the same manner, we anticipate that One-SENSE would also be useful to investigate the categorical relationship of T cells in comparison of patient cohorts.

Compared with two-dimensional t-SNE, each axis in One-SENSE has a specific meaning that can be conveniently annotated on the plot. Importantly, this analysis approach is customizable, driven by the purpose or hypothesis of the experiment. The main benefit of One-SENSE is its utility for directly testing hypotheses about the relationships between different categories of cellular diversity. We anticipate that One-SENSE would be well suited for the analysis of any high-dimensional single-cell data (6) for the purposes of identifying novel biomarkers associated with immunological dysfunction, such as in the context of autoimmunity or cancer. One-SENSE would also be useful for defining new or rare immune cell subsets and their heterogeneity (e.g., function and differentiation), especially when comparing disease and healthy individuals. We expect that the robustness of One-SENSE will largely improve the analysis of high-dimensional mass cytometry data and should be broadly applicable for the analysis of other high-dimensional data, such as for single-cell RNA sequencing (4, 5) or pre-existing large statistical datasets.

#### Acknowledgments

We thank the Singapore Immunology Network community and all members of the Newell laboratory for helpful discussion and technical support. We also thank Michael Kuhns and members of the Kuhns laboratory for helpful comments. The KLRG-1 Ab was generously provided by Hanspeter Pircher.

#### Disclosures

The authors have no financial conflicts of interest.

#### References

- Lanier, L. L., E. G. Engleman, P. Gatenby, G. F. Babcock, N. L. Warner, and L. A. Herzenberg. 1983. Correlation of functional properties of human lymphoid cell subsets and surface marker phenotypes using multiparameter analysis and flow cytometry. *Immunol. Rev.* 74: 143–160.
- Chattopadhyay, P. K., T. M. Gierahn, M. Roederer, and J. C. Love. 2014. Singlecell technologies for monitoring immune systems. *Nat. Immunol.* 15: 128–135.
- Bandura, D. R., V. I. Baranov, O. I. Ornatsky, A. Antonov, R. Kinach, X. Lou, S. Pavlov, S. Vorobiev, J. E. Dick, and S. D. Tanner. 2009. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal. Chem.* 81: 6813–6822.
- Macosko, E. Z., A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, et al. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161: 1202–1214.
- Klein, A. M., L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, and M. W. Kirschner. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161: 1187– 1201.
- Stegle, O., S. A. Teichmann, and J. C. Marioni. 2015. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16: 133–145.
- Newell, E. W., N. Sigal, N. Nair, B. A. Kidd, H. B. Greenberg, and M. M. Davis. 2013. Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat. Biotechnol.* 31: 623–629.
- Bendall, S. C., G. P. Nolan, M. Roederer, and P. K. Chattopadhyay. 2012. A deep profiler's guide to cytometry. *Trends Immunol.* 33: 323–332.
- van der Maaten, L., and G. Hinton. 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9: 2579–2605.
- Amir E. D., K. L. Davis, M. D. Tadmor, E. F. Simonds, J. H. Levine, S. C. Bendall, D. K. Shenfeld, S. Krishnaswamy, G. P. Nolan, and D. Pe'er. 2013. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* 31: 545–552.
- Newell, E. W., N. Sigal, S. C. Bendall, G. P. Nolan, and M. M. Davis. 2012. Cytometry by time-of-flight shows combinatorial cytokine expression and virusspecific cell niches within a continuum of CD8+ T cell phenotypes. *Immunity* 36: 142–152.
- Becher, B., A. Schlitzer, J. Chen, F. Mair, H. R. Sumatoh, K. W. Teng, D. Low, C. Ruedl, P. Riccardi-Castagnoli, M. Poidinger, et al. 2014. High-dimensional analysis of the murine myeloid cell system. *Nat. Immunol.* 15: 1181–1189.
- Bendall, S. C., E. F. Simonds, P. Qiu, A. D. Amir, P. O. Krutzik, R. Finck, R. V. Bruggner, R. Melamed, A. Trejo, O. I. Ornatsky, et al. 2011. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 332: 687–696.
- Shekhar, K., P. Brodin, M. M. Davis, and A. K. Chakraborty. 2014. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). Proc. Natl. Acad. Sci. USA 111: 202–207.
- van der Maaten, L. 2009. Learning a parametric embedding by preserving local structure. In JMLR Workshop and Conference Proceedings, Vol. 5: Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, April 16–18. D. van Dyk and M. Welling, eds. Clearwater Beach, FL, p. 384– 391.
- Sallusto, F., D. Lenig, R. Förster, M. Lipp, and A. Lanzavecchia. 1999. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* 401: 708–712.
- Sallusto, F., J. Geginat, and A. Lanzavecchia. 2004. Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu. Rev. Immunol.* 22: 745–763.
- Arens, R., and S. P. Schoenberger. 2010. Plasticity in programming of effector and memory CD8 T-cell formation. *Immunol. Rev.* 235: 190–205.
- Kaech, S. M., and W. Cui. 2012. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat. Rev. Immunol.* 12: 749–761.

- Harty, J. T., and V. P. Badovinac. 2008. Shaping and reshaping CD8+ T-cell memory. Nat. Rev. Immunol. 8: 107–119.
- Griffith, J. W., C. L. Sokol, and A. D. Luster. 2014. Chemokines and chemokine receptors: positioning cells for host defense and immunity. *Annu. Rev. Immunol.* 32: 659–702.
- Liston, A., and D. H. Gray. 2014. Homeostatic control of regulatory T cell diversity. *Nat. Rev. Immunol.* 14: 154–165.
- Finck, R., E. F. Simonds, A. Jager, S. Krishnaswamy, K. Sachs, W. Fantl, D. Pe'er, G. P. Nolan, and S. C. Bendall. 2013. Normalization of mass cytometry data with bead standards. *Cytometry A* 83: 483–494.
- Feau, S., R. Arens, S. Togher, and S. P. Schoenberger. 2011. Autocrine IL-2 is required for secondary population expansion of CD8(+) memory T cells. *Nat. Immunol.* 12: 908–913.
- Frentsch, M., R. Stark, N. Matzmohr, S. Meier, S. Durlanik, A. R. Schulz, U. Stervbo, K. Jürchott, F. Gebhardt, G. Heine, et al. 2013. CD40L expression permits CD8+ T cells to execute immunologic helper functions. *Blood* 122: 405– 412.
- Hess, C., T. K. Means, P. Autissier, T. Woodberry, M. Altfeld, M. M. Addo, N. Frahm, C. Brander, B. D. Walker, and A. D. Luster. 2004. IL-8 responsiveness defines a subset of CD8 T cells poised to kill. *Blood* 104: 3463–3471.
- Tang, Q., and J. A. Bluestone. 2008. The Foxp3+ regulatory T cell: a jack of all trades, master of regulation. *Nat. Immunol.* 9: 239–244.
- Kleinewietfeld, M., and D. A. Hafler. 2013. The plasticity of human Treg and Th17 cells and its role in autoimmunity. *Semin. Immunol.* 25: 305–312.
- Gagliani, N., C. F. Magnani, S. Huber, M. E. Gianolini, M. Pala, P. Licona-Limon, B. Guo, D. R. Herbert, A. Bulfone, F. Trentini, et al. 2013. Coexpression of CD49b and LAG-3 identifies human and mouse T regulatory type 1 cells. *Nat. Med.* 19: 739–746.
- Makedonas, G., and M. R. Betts. 2011. Living in a house of cards: re-evaluating CD8+ T-cell immune correlates against HIV. *Immunol. Rev.* 239: 109–124.
- Sad, S., R. Marcotte, and T. R. Mosmann. 1995. Cytokine-induced differentiation of precursor mouse CD8+ T cells into cytotoxic CD8+ T cells secreting Th1 or Th2 cytokines. *Immunity* 2: 271–279.
- Vukmanovic-Stejic, M., B. Vyas, P. Gorak-Stolinska, A. Noble, and D. M. Kemeny. 2000. Human Tc1 and Tc2/Tc0 CD8 T-cell clones display distinct cell surface and functional phenotypes. *Blood* 95: 231–240.
- Boer, M. C., K. E. van Meijgaarden, J. Bastid, T. H. Ottenhoff, and S. A. Joosten. 2013. CD39 is involved in mediating suppression by *Mycobacterium* bovis BCGactivated human CD8(+) CD39(+) regulatory T cells. *Eur. J. Immunol.* 43: 1925–1932.
- Leavenworth, J. W., X. Tang, H. J. Kim, X. Wang, and H. Cantor. 2013. Amelioration of arthritis through mobilization of peptide-specific CD8+ regulatory T cells. J. Clin. Invest. 123: 1382–1389.
- Bin Dhuban, K., M. Kornete, E. S Mason, and C. A. Piccirillo. 2014. Functional dynamics of Foxp3<sup>+</sup> regulatory T cells in mice and humans. *Immunol. Rev.* 259: 140–158.
- Josefowicz, S. Z., L. F. Lu, and A. Y. Rudensky. 2012. Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* 30: 531–564.
- d'Hennezel, E., E. Yurchenko, E. Sgouroudis, V. Hay, and C. A. Piccirillo. 2011. Single-cell analysis of the human T regulatory population uncovers functional heterogeneity and instability within FOXP3+ cells. J. Immunol. 186: 6788–6797.
- Tran, D. Q., H. Ramsey, and E. M. Shevach. 2007. Induction of FOXP3 expression in naive human CD4+FOXP3 T cells by T-cell receptor stimulation is transforming growth factor-beta dependent but does not confer a regulatory phenotype. *Blood* 110: 2983–2990.
- 39. Locci, M., C. Havenar-Daughton, E. Landais, J. Wu, M. A. Kroenke, C. L. Arlehamn, L. F. Su, R. Cubas, M. M. Davis, A. Sette, et al; International AIDS Vaccine Initiative Protocol C Principal Investigators. 2013. Human circulating PD-1+CXCR3-CXCR5+ memory Tfh cells are highly functional and correlate with broadly neutralizing HIV antibody responses. *Immunity* 39: 758– 769.
- Bentebibel, S. E., S. Lopez, G. Obermoser, N. Schmitt, C. Mueller, C. Harrod, E. Flano, A. Mejias, R. A. Albrecht, D. Blankenship, et al. 2013. Induction of ICOS+CXCR3+CXCR5+ TH cells correlates with antibody responses to influenza vaccination. *Sci. Transl. Med.* 5: 176ra32.
- Schmitt, N., and H. Ueno. 2015. Regulation of human helper T cell subset differentiation by cytokines. *Curr. Opin. Immunol.* 34: 130–136.