

A Beginner's Guide to Analyzing and Visualizing Mass Cytometry Data

Abigail K. Kimball,* Lauren M. Oko,[†] Bonnie L. Bullock,[‡] Raphael A. Nemenoff,[‡]
Linda F. van Dyk,[†] and Eric T. Clambey*

Mass cytometry has revolutionized the study of cellular and phenotypic diversity, significantly expanding the number of phenotypic and functional characteristics that can be measured at the single-cell level. This high-dimensional analysis platform has necessitated the development of new data analysis approaches. Many of these algorithms circumvent traditional approaches used in flow cytometric analysis, fundamentally changing the way these data are analyzed and interpreted. For the beginner, however, the large number of algorithms that have been developed, as well as the lack of consensus on best practices for analyzing these data, raise multiple questions: Which algorithm is the best for analyzing a dataset? How do different algorithms compare? How can one move beyond data visualization to gain new biological insights? In this article, we describe our experiences as recent adopters of mass cytometry. By analyzing a single dataset using five cytometry by time-of-flight analysis platforms (viSNE, SPADE, X-shift, PhenoGraph, and Citrus), we identify important considerations and challenges that users should be aware of when using these different methods and common and unique insights that can be revealed by these different methods. By providing annotated workflow and figures, these analyses present a practical guide for investigators analyzing high-dimensional datasets. In total, these analyses emphasize the benefits of integrating multiple cytometry by time-of-flight analysis algorithms to gain complementary insights into these high-dimensional datasets. *The Journal of Immunology*, 2018, 200: 3–22.

Since its inception, mass cytometry, or cytometry by time-of-flight (CyTOF), has allowed researchers to gain deep insights into cellular phenotype and function

(1–3). The technology allows simultaneous quantification of >30 cellular parameters and when integrated with high-dimensional analysis algorithms, it has the potential to reveal extraordinary cellular diversity and heterogeneity (2, 4, 5). Many algorithms and software kits have been developed to facilitate analysis of CyTOF datasets, including, but not limited to SPADE (6), viSNE (7), Wanderlust (8), FlowSOM (9), PhenoGraph (10), Citrus (11), Scaffold (12), X-shift (13), and DensVM (14). These tools are typically developed by computational biologists or by laboratories that are leaders in the field of mass cytometry, using a variety of languages (e.g., R, Matlab, Java, Python), clustering methods (e.g., parametric, nonparametric), and dimensionality reduction approaches (15–17). CyTOF data visualization and quantitation continues to be a rapidly evolving field (e.g., 18, 19).

Despite the potential of mass cytometry, there remain multiple challenges to its widespread implementation, from instrument and reagent costs to determining optimal ways to visualize and quantify these high-dimensional data. For researchers with little to no computational background, entry into these data can represent a significant challenge. Even for laboratories well-versed in multiparameter flow cytometry, analyzing mass cytometry data requires a major shift in how to approach these data, moving away from user-defined Boolean gating strategies to automated identification of cell clusters and phenotypes. Although biological knowledge remains essential to interpret these high-dimensional data, understanding which tools to use when addressing specific biological questions remains a major challenge.

We present a practical guide for CyTOF data analysis, a process that we have defined empirically as recent adopters of these methods. First, we provide a detailed commentary on how to implement and interpret data using five established and widely used CyTOF analysis platforms (viSNE, SPADE,

*Department of Anesthesiology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045; [†]Department of Immunology and Microbiology, University of Colorado Anschutz Medical Campus, Aurora, CO 80045; and [‡]Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045

ORCIDs: 0000-0002-1580-0292 (B.L.B.); 0000-0003-2662-5554 (L.F.v.D.); 0000-0002-7972-9544 (E.T.C.).

Received for publication October 30, 2017. Accepted for publication October 30, 2017.

This work was supported by National Institutes of Health Grants R01 AI121300 and R01 CA168558 (to L.F.v.D.) and R01 CA162226 and P50 CA058187 (to R.A.N.), an American Heart Association National Scientist Development grant, the Crohn's and Colitis Foundation of America, and a Career Enhancement Award from the University of Colorado Lung Cancer Specialized Program of Research Excellence (all to E.T.C.).

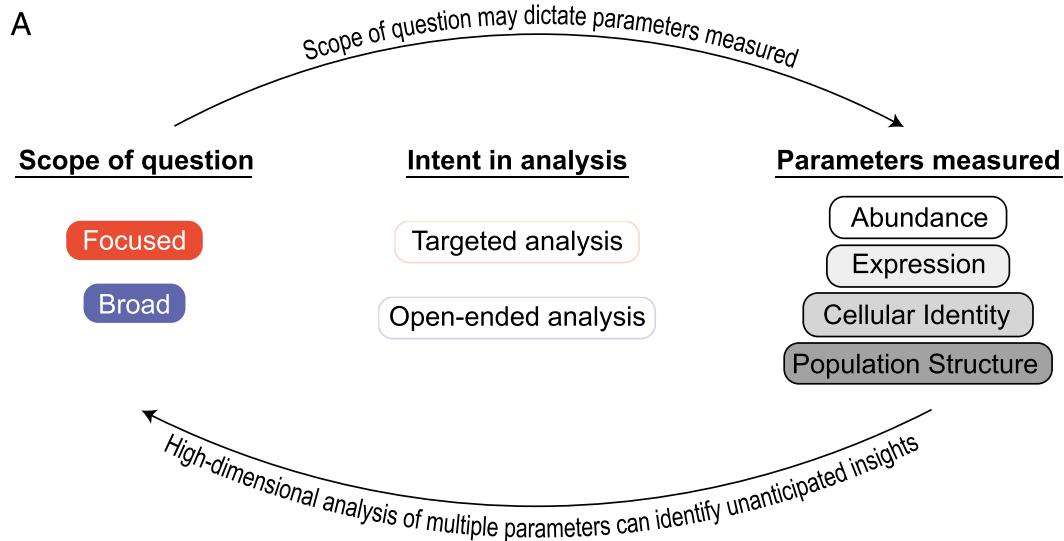
The Flow Cytometry Shared Resource receives direct funding support from the National Cancer Institute through Cancer Center Support Grant P30CA046934.

Address correspondence and reprint requests to Dr. Eric T. Clambey, Department of Anesthesiology, School of Medicine, University of Colorado Anschutz Medical Campus, 12700 East 19th Avenue, Box 112, Aurora, CO 80045. E-mail address: Eric.Clambey@ucdenver.edu

The online version of this article contains supplemental material.

Abbreviations used in this article: B6, C57BL/6; CyTOF, cytometry by time-of-flight; γHV68, gammaherpesvirus 68; IL-10KO, IL-10 knockout; LLC, Lewis lung carcinoma; tSNE, t-distributed stochastic neighbor embedding.

Copyright © 2017 by The American Association of Immunologists, Inc. 0022-1767/17/\$35.00



Examples:

- Focused Cellular Identity Are distinct CD8 T cell phenotypes found during virus infection?
- Broad Population Structure Does the distribution of cell populations differ between conditions?

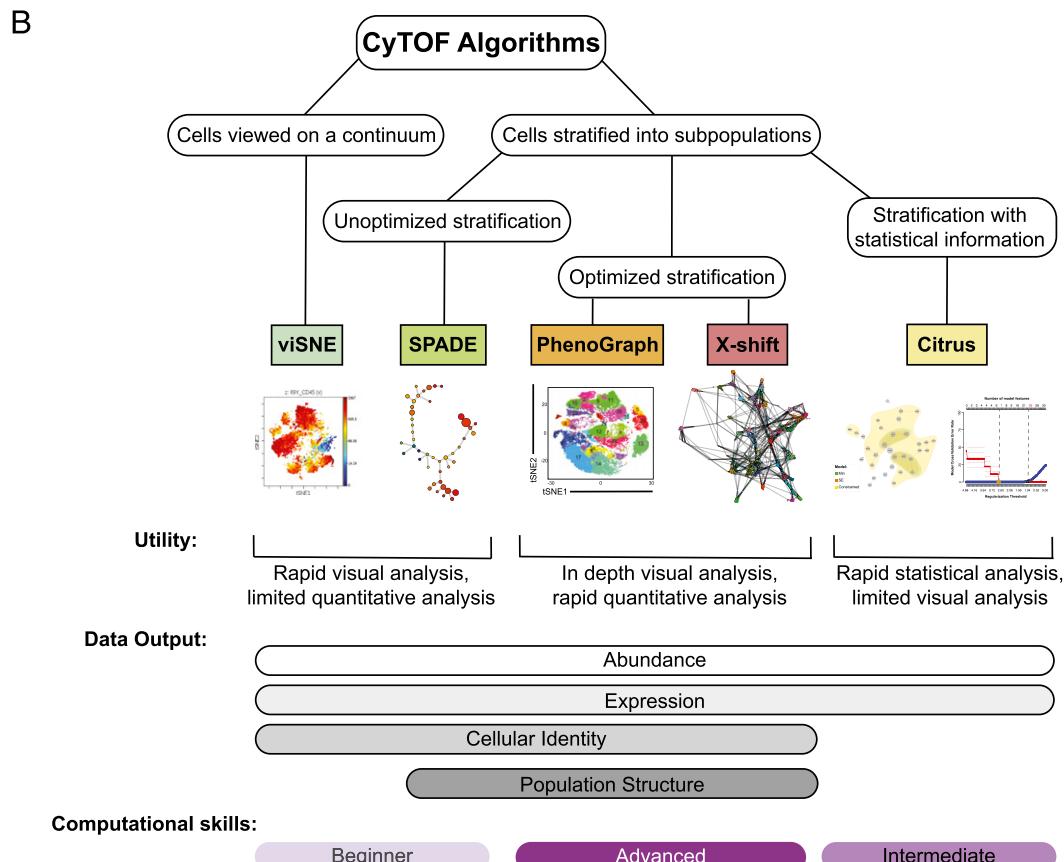


FIGURE 1. Important considerations in CyTOF experimental design and algorithm implementation. Examples of the type of research questions that can be answered by CyTOF (**A**) and considerations for the use of different CyTOF algorithms (**B**).

X-shift, Citrus, and PhenoGraph). Second, we supply annotated examples, illustrating how to use these different algorithms to gain complementary insights into a single dataset.

These analyses provide a conceptual framework and a resource for investigators interested in using CyTOF to gain high-dimensional insights into biological questions.

viSNE Settings (Cytobank)

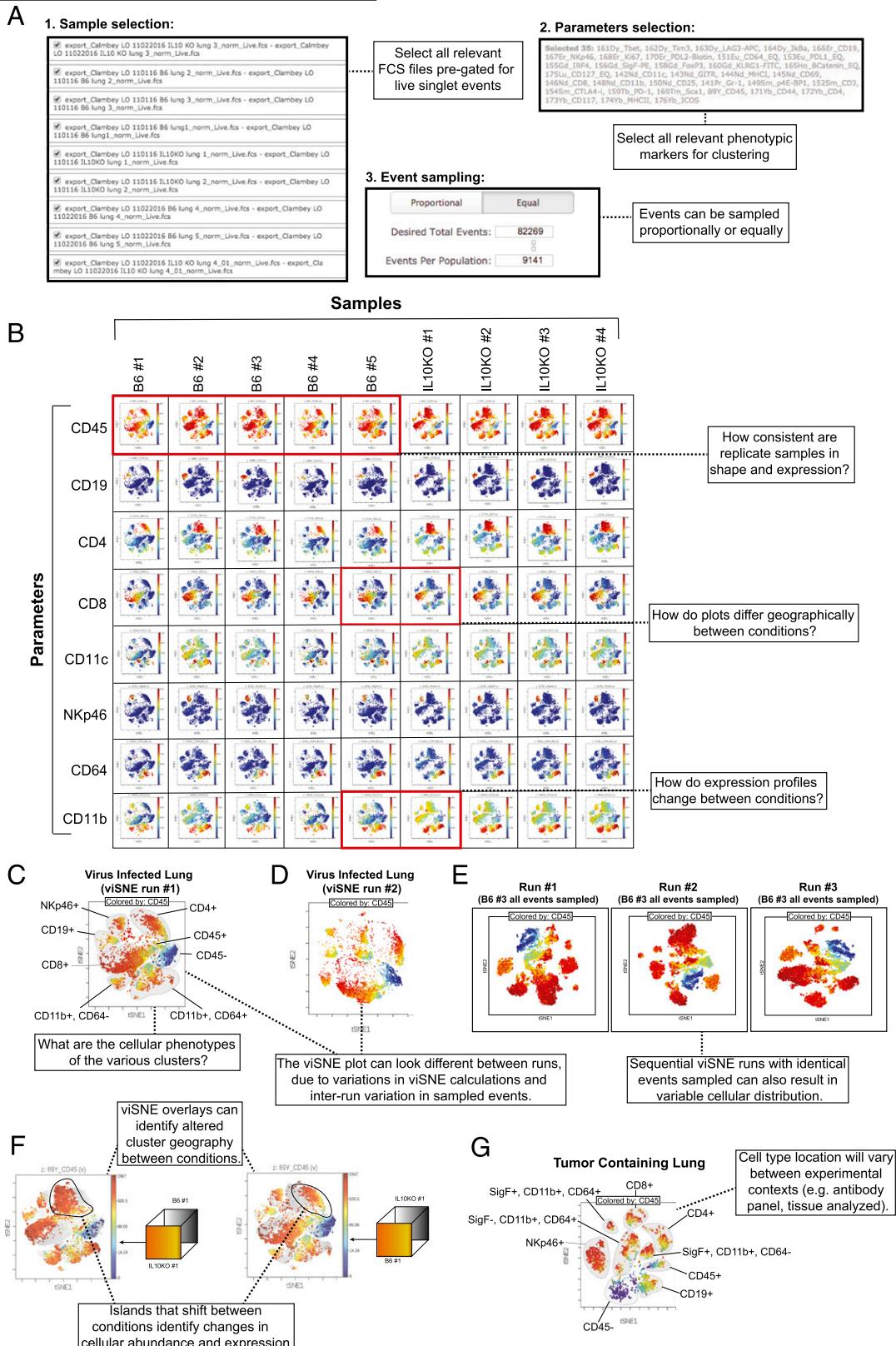


FIGURE 2. Basic considerations for viSNE analysis. Input settings (**A**) and graphical representation of viSNE analysis using the Cytobank platform, with data representing CyTOF analysis of γ HV68-infected lungs from B6 or IL-10KO individuals at 9 d postinfection (**B–F**) or B6 mice orthotopically implanted with the Lewis lung carcinoma (LLC) tumor cell line (**G**). Data show all viable single cells, subjected to the tSNE algorithm, which provides each cell with a unique coordinate according to its expression of the 35 measured parameters, displayed on a two-dimensional plot (tSNE1 versus tSNE2). (**A**) Input settings to run the viSNE algorithm in Cytobank. (**B**) Visualization grid of viSNE plots, with plots arranged according to marker expression (rows) relative to individuals (columns). (**C**) Identification of cellular populations identified by viSNE for individual B6 #1, with cell populations defined based on basic phenotypic markers (see *Materials and Methods*). (**D**) An additional viSNE plot produced using identical settings for individual B6 #1 and colored by CD45, demonstrating variable output of (*Figure legend continues*)

Materials and Methods

Experimental samples

Samples were obtained from two sources. First, 13-wk-old female C57BL/6J mice (B6, $n = 5$) or IL-10-knockout (IL-10KO) mice ($n = 4$; B6.129P2-I l 10^{tm1Cgn}/J; The Jackson Laboratory) were infected with murine gamma-herpesvirus 68 (γ HV68) by intranasal infection using 4×10^5 PFU wild-type virus containing a gene 73- β lactamase fusion protein, as described (20). Mice were euthanized, and lungs were harvested at 9 d postinfection, followed by perfusion with PBS. Second, UBI.GFP [C57BL/6-Tg(UBC-GFP)30Scha/J] mice were orthotopically injected in the left lobe of the lung with 1×10^5 cells suspended in HBSS containing 1.35 mg/ml Matrigel (product number 354234; Corning). Mice received firefly luciferase-expressing CMT167 cells (21) or luciferase-expressing Lewis lung carcinoma (LLC) cells (Caliper Life Sciences). Both cell lines were routinely tested for mycoplasma and were confirmed negative before orthotopic injection. All surgeries were performed under inhaled isoflurane anesthesia. A 4–5-mm incision was made in the skin along the left shoulder and s.c. fat was removed to completely visualize the left lung, as previously described (22). Following orthotopic injection, incisions were closed using veterinary-grade skin adhesive. Mice were euthanized at 2.5 (LLC cells) or 3.5 wk (CMT167 cells) postinjection, followed by perfusion of the circulation with PBS/heparin (20 U/ml) and collection of the tumor-containing left lobe for analysis. Pools of two mice were used for LLC tumors, pools of three to five mice were used for CMT167 tumors, and left and right lung lobes were used for uninjected controls. All procedures were performed under protocols approved by the Institutional Animal Care and Use Committee at the University of Colorado Anschutz Medical Campus.

Cell processing and Ab staining

Lungs were minced and enzymatically digested using collagenase D (from *Clostridium histolyticum*; Roche) for 1 h at 37°C (for virus-infected lungs) or were incubated at 37°C for 30 min in HBSS containing collagenase type 4 (8480 U/ml), elastase (7.5 mg/ml), and soybean trypsin inhibitor (2 mg/ml; all from Worthington Industries) (for tumor-containing lungs). Lungs were mechanically disrupted to single-cell suspensions, subjected to RBC lysis, washed, and resuspended for staining. Cell suspensions were stained with cisplatin (Fluidigm), incubated with Fc receptor blocking Ab (2.4G2; Tonbo Biosciences) for 20 min, following by the addition of primary surface Abs, and incubated for 15 min at 37°C and 15 min at 22°C. Secondary surface stains were done for 20 min, with intracellular Ag staining done using eBioscience Foxp3 Fix/Perm buffer and a 2-h stain at 4°C. Following cell staining, cells were washed and resuspended in Intercalator (Cell-ID Intercalator-Ir; Fluidigm). Abs are listed in Supplemental Tables I and II and were from Fluidigm, unless noted otherwise.

CyTOF run and sample normalization

Samples were collected on a Helios mass cytometer (Fluidigm), with samples resuspended with equilibration beads spiked into each sample to allow for signal normalization. Samples were normalized using NormalizerR2013b_MacOSX, downloaded from the Nolan laboratory GitHub page (<https://github.com/nolanlab/vortex/releases/tag/24-Dec-2016>), as well as the most recent version of Java (Version 8, Update 121, <https://java.com/en/download/>).

viSNE. Algorithm settings. Manually gated singlet ($^{191}\text{Ir}^+ \ ^{193}\text{Ir}^+$) viable ($^{195}\text{Pt}^-$) events from γ HV68-infected or tumor-containing mice were imported into Cytobank (<https://www.cytobank.org>) and then subjected to viSNE analysis. viSNE clustering analysis was performed on 35 of 61 possible parameters, focused on Abs (Supplemental Table I). Equal event sampling was selected, using 9,141 events per individual (the lowest common denominator across all samples), for a total of 82,269 events across all nine virus-infected samples.

Algorithm optimization, interrogation, and visualization. viSNE plots for each individual across all 35 parameters were downloaded from Cytobank, with plots arranged in a grid format in Adobe Illustrator CC 2017. Cellular phenotypes were assigned to the viSNE plot based on distribution and expression characteristics using phenotypic markers, with viSNE overlays manually generated in Adobe Illustrator.

Investigating cellular abundance. CD4 $^+$ T cells were manually gated in viSNE plots within Cytobank, to quantify the percentage of total events for all individuals.

Investigating cellular expression. CD4 $^+$ T cells were manually gated in viSNE plots, analyzed for expression across all 35 parameters, and downloaded from Cytobank. Expression profiles were visually inspected between conditions, with differences in expression between the conditions manually selected for further analysis using FlowJo.

PhenoGraph. Software downloads. The 1.0.136 version of R studio was downloaded from the official R Web site (<http://www.r-project.org/>). Release 3.4 of the cytofkit package was downloaded from Bioconductor (<https://www.bioconductor.org/packages/release/bioc/html/cytofkit.html>) and opened in R.

Algorithm settings. Manually gated singlet ($^{191}\text{Ir}^+ \ ^{193}\text{Ir}^+$) viable ($^{195}\text{Pt}^-$) events from γ HV68-infected mice were imported into cytofkit, subjected to PhenoGraph analysis, and clustered based on 35 of 61 possible parameters (Abs in Supplemental Table I), with additional settings: merge method: minimum, transformation: cytofAsinh, cluster method: Rphenograph, visualization method: tSNE, and cellular progression: NULL.

Algorithm optimization, interrogation, and visualization. All events were uploaded into PhenoGraph, which defined 29 subpopulations of cells or clusters. Clusters were displayed on tSNE plots within the R package “Shiny” to visualize individuals and experimental groups. Cluster identification number labels, dot size, and cluster color were customized, to indicate phenotype or to highlight specific clusters. Additional plots were colored according to the expression of cellular markers. Multiple .csv files were produced by the PhenoGraph analysis, including “cluster median data” and “cluster cell percentage,” which were used to determine cluster phenotype, distribution between conditions, and statistical significance between groups.

Investigating cellular abundance. Clusters were initially visually inspected to identify potential differences in cellular abundances between conditions. Subsequently, all clusters were analyzed for statistical significance between experimental groups.

Investigating cellular expression. CD4 $^+$ T cell clusters were visually inspected for differences in expression between experimental groups.

X-shift. Software downloads. The 12/24/2016 version of Vortex was downloaded (from the Nolan laboratory GitHub page [<https://github.com/nolanlab/vortex/releases/tag/24-Dec-2016>]), as well as the most recent version of Java (Version 8, Update 121, <https://java.com/en/download/>).

Algorithm settings. Manually gated singlet ($^{191}\text{Ir}^+ \ ^{193}\text{Ir}^+$) viable ($^{195}\text{Pt}^-$) events from virus-infected or tumor-containing lungs were uploaded into the Vortex clustering environment. The following importation settings were used: minimal Euclidean length of the profile: 1.0, import maximum: 10,000 events, and merge all files into one dataset. These settings were either automatic or recommended by the GitHub page. Thirty-five parameters were selected for clustering (Supplemental Table I). After the data were imported, before clustering, we validated the size and dimensions of the anticipated dataset. The following clustering settings were used: numerical transformation: arcsinh (x/f), $f = 5.0$, noise threshold: apply noise threshold of 1.0, feature rescaling: none, normalization: none, distance measure: angular distance, clustering algorithm: X-shift (gradient assignment), density estimate: N nearest neighbors (fast), number of neighbors for density estimate (K): from 150 to 5, with 30 steps, and number of neighbors for mode finding (N): determine automatically. These clustering settings were either automatic or recommended by the GitHub page. Following clustering, all 30 sets were selected and the K value of the switch point, or the elbow point, between the linear and exponential phase was calculated (in this case, $K = 20$, which corresponded to 45 X-shift-defined clusters within the virus-infected dataset).

Algorithm optimization, interrogation, and visualization. All 45 clusters in the $K = 20$ cluster set were selected, and a force-directed layout was created with the following settings: maximum number of events sampled from each cluster: 20, sample proportionally to the power of the cluster size: unchecked, distance measurement: angular distance, number of nearest neighbors: 10, edge settings: uncheck the options to “Vary the number of node connections by node density” and “Limit connections to the events within parameter

viSNE across independent runs, potentially reflecting variable viSNE calculation and events sampled. (E) Comparison of three sequential viSNE runs, in which the same 9141 cells were subjected to viSNE, demonstrates variable cellular distribution (for individual B6 #3). (F) Reciprocal viSNE overlays comparing the topography of viSNE plots from B6 and IL-10KO mice. (G) viSNE analysis of cellular populations from a tumor-containing lung, with cell populations defined based on basic phenotypic markers. Data from virus-infected lungs (B6, $n = 5$; IL-10KO, $n = 4$ mice), naive lung (right and left lobes of lung pooled together from $n = 2$ mice), and LLC tumor-containing lung (left lobe of lung pooled from $n = 2$ mice) per condition.

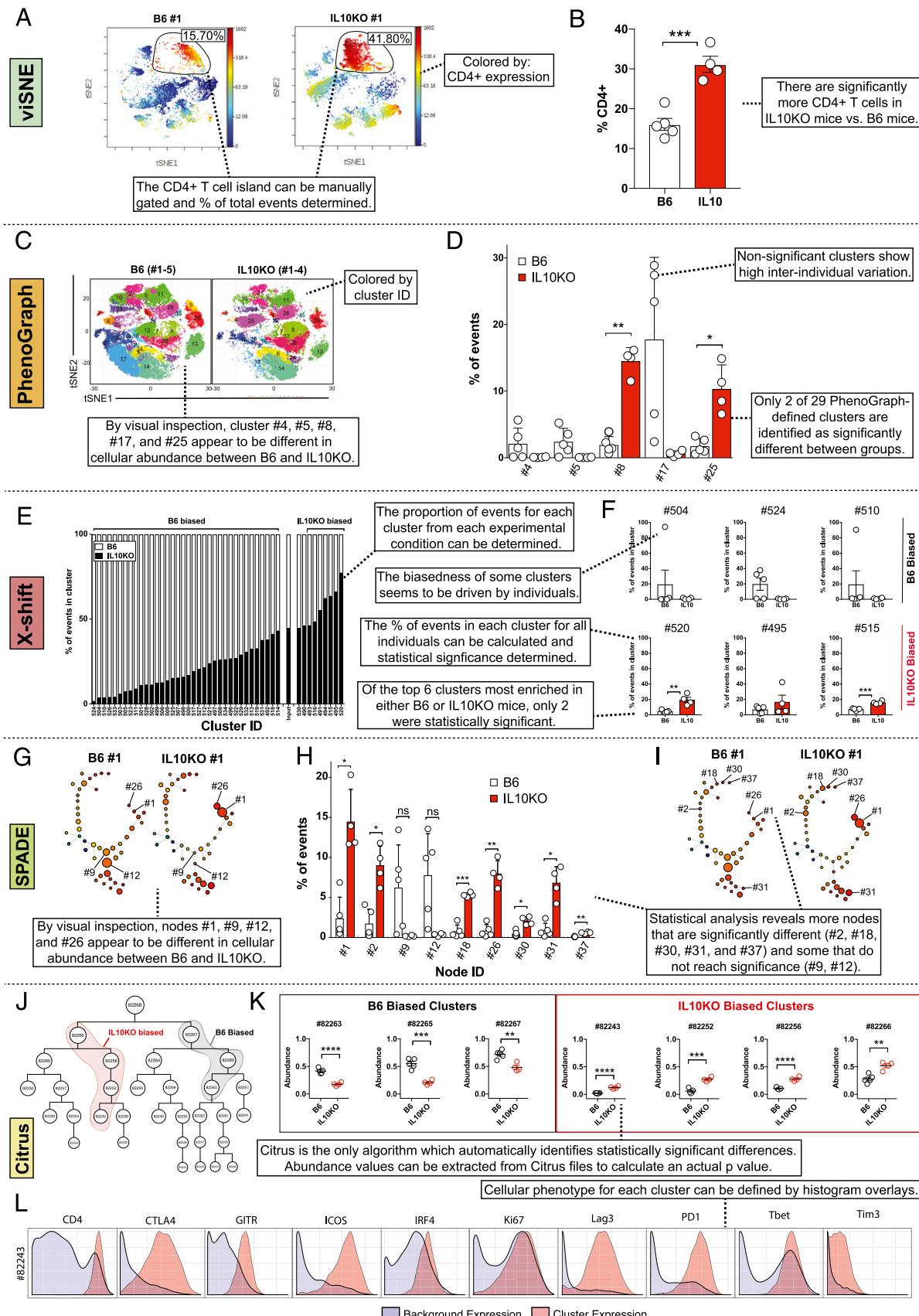


FIGURE 3. Investigating cellular abundance by viSNE, PhenoGraph, X-shift, SPADE, and Citrus algorithms. CyTOF analysis of γ HV68-infected lungs from B6 or IL-10KO individuals harvested at 9 d postinfection. Data show all viable single cells, subjected to the various algorithms, and include examples demonstrating insights obtained across algorithms. (A and B) viSNE analysis showing viSNE plots for individuals B6 #1 and IL-10KO #1 colored according to CD4 expression. (B) The CD4⁺ T cell island was visually identified as one notable change between B6 and IL-10KO mice, with this population manually gated within Cytobank, to define the frequency of CD4 T cells across all individuals. (C and D) PhenoGraph analysis of B6 and IL-10KO mice identified (Figure legend continues)

range,” and sample selection: select the option “Include cells from certain annotation groups” and select all nine samples. This layout was downloaded as a graphml file and opened in Gephi v 0.9.1 where color and node size were optimized. This file was then saved as a .pdf and opened in Adobe Illustrator CC 2017. The clusters were also colored according to phenotype designation in Gephi.

Investigating cellular abundance. All events from each cluster were downloaded from VortexX, and the frequency of events from different experimental groups (e.g., B6 versus IL-10KO) was determined in Excel, followed by manually adding these plots to the force-directed layout. These data were further interrogated to quantify the frequency of events derived from each individual mouse that contributed to the cluster.

Investigating cellular expression. We visualized cellular expression across all parameters by phenotypic barcodes within VortexX, to analyze average expression across a cluster and expression across individual events. This feature, combined with line graphs of median expression, allowed definition of a core cellular phenotype for all clusters (e.g., CD4⁺ or CD8⁺ T cell), with accessory phenotypes present in only a subset of cells. Core and accessory phenotypes were depicted using manually-generated infographics.

SPADE. Algorithm settings. Manually gated singlet (¹⁹¹Ir⁺ ¹⁹³Ir⁺) viable (¹⁹⁵Pt⁻) events from γ HV68-infected mice were imported into Cytobank and then subjected to SPADE analysis using the following settings: target number of nodes = 200 and percentage downsampling = 10% (both Cytobank default values). Single viable cells were selected for analysis, using 35 of 61 possible parameters (Abs in Supplemental Table I); no fold change calculations were made for this data set.

Algorithm optimization, interrogation, and visualization. A SPADE tree was manually rearranged online at <https://www.cytobank.org>. Additional SPADE analyses were run with various alterations: the target number of nodes was adjusted to an X-shift informed 45 or 89 nodes, and the number of clustering channels was decreased to include only 10 lineage markers. After these analyses, the SPADE tree informed by VortexX with all 35 parameters was chosen as most appropriate for this analysis. SPADE trees for each individual across all 35 parameters were downloaded from Cytobank, and the most relevant markers were arranged in a table using Adobe Illustrator. Basic cellular phenotypes were assigned to the SPADE trees, based on node expression and location.

Investigating cellular abundance. SPADE trees were initially visually investigated to identify nodes that were potentially different in cellular abundance between conditions. Data accompanying each SPADE tree were downloaded from Cytobank to calculate the frequency of events in each node, followed by testing for statistical significance. The identification of significant nodes on the SPADE tree was determined online at <https://www.cytobank.org> by hovering over the desired node.

Investigating cellular expression. Nodes of interest were colored by cellular markers that had been identified by previous algorithms to change in expression between experimental groups.

A change in color was visually assessed by the user, and nodes 1 and 26 were taken from their original downloaded SPADE trees. Node 1 was selected for each individual on Cytobank, and the median expression value for five markers was recorded and tested for statistical significance.

Citrus. Algorithm settings. Manually gated singlet (¹⁹¹Ir⁺ ¹⁹³Ir⁺) viable (¹⁹⁵Pt⁻) events from γ HV68-infected mice were imported into Cytobank and then subjected to Citrus analysis using the following settings: singlet viable cells were chosen as input, 35 of 61 possible parameters were selected (Abs outlined in Supplemental Table I), the files were assigned to their appropriate experimental groups (i.e., B6 versus IL-10KO), the Nearest Shrunken Centroid association model, cluster characterization of abundance, equal event sampling, events sampled per file: 9141 events, minimum cluster size: 5% (Cytobank default), cross-validation folds: 5 (Cytobank default), and false discovery rate: 1% (Cytobank default).

Algorithm optimization, interrogation, and visualization. To test the sensitivity of Citrus to different variables, additional Citrus analyses were run to test the impact of random assignment of samples to different groups, input cell number, and reducing the minimum cluster size. Analysis primarily focused on the resulting “Model Error Rate” models. The Citrus run with 0% cross-validation error rate was selected as the most appropriate predictive model, and results were downloaded from Cytobank for further visual and statistical analysis. The vertical hierarchical tree was manually generated based on the radial hierarchical tree in Adobe Illustrator.

Investigating cellular abundance. Daughters identified as differing in abundance between B6 and IL-10KO mice by the minimum and SE statistical models were analyzed further. The direction of this change in abundance was determined via the results downloaded from Citrus, with specific *p* values determined.

Investigating cellular expression. To query changes in cellular expression, another Citrus analysis was performed, with identical settings (see above), to examine changes in median expression rather than abundance.

Comparing algorithms. Plots/force-directed layouts/trees from each algorithm are colored according to the eight lineage markers used to identify cellular phenotypes. The percentage of events in all clusters/nodes from four algorithms was identified and combined according to cellular phenotypes. The number of unique clusters/nodes within each phenotypic group was determined for four algorithms. The median expression values for particular daughters/nodes/clusters, which appear to contain similar cellular events, have been plotted as a line graph to determine phenotypic similarity and dissimilarity. A network was created in Adobe Illustrator in which markers identified as positive within each algorithm were plotted. For this network map, positive expression for a parameter was defined as any value greater than the average expression across daughters/nodes/clusters within an algorithm.

Investigating population structure. Population structure was analyzed by the stratifying algorithms SPADE, X-shift, and PhenoGraph. Population structure was visualized by downloading the frequencies of all nodes/clusters across experimental conditions. These data were then visualized using a vertical hierarchy plot, in which node frequency was arranged in ascending order for each experimental group, identifying nodes that significantly changed in abundance between conditions, or they were visualized using the frequency and subset distribution of different cellular phenotypes across experimental groups. Lastly, clustering relationships were visualized in cytokit by dendograms, depicting the relatedness between clusters and changing relationships between conditions. The following settings were used for dendrogram generation within the Shiny app after selection of the “Expression Heat Map” tab: heatmap dendrogram: row, color palette: spectral2, heatmap type: median, and scale data: none. The dendrogram was downloaded as a .pdf file and opened and edited in Adobe Illustrator.

Software used and statistical analysis. Software for data analysis included R studio (Version 1.0.136), downloaded from the official R Web site (<https://www.r-project.org/>); the cytokit package (Release 3.4), downloaded from Bioconductor (<https://www.bioconductor.org/packages/release/bioc/html/cytokit.html>); VortexX (Version 12/24/2016), downloaded from the Nolan laboratory GitHub page (<https://github.com/nolanlab/vortex/releases/tag/24-Dec-2016>), Java (Version 8.121, <https://java.com/en/download/>), Gephi v 0.9.1, Excel 15.13.14, FlowJo 10.2, GraphPad Prism 7, and Adobe Illustrator CC 2017. We used two different versions of VortexX [12/24/2016 (Fig. 6) and 4/21/2017 (Fig. 9)]. Note that these identified different numbers of clusters despite using identical settings. Statistical significance was tested in GraphPad Prism

29 clusters, colored by cluster identification number and plotted according to tSNE1, tSNE2, identified multiple clusters with apparent changes between groups. (D) Statistical analysis of all PhenoGraph-defined clusters identified 2 of 29 clusters that were statistically significantly different between B6 and IL-10KO mice. (E) X-shift analysis of B6 and IL-10KO mice identified 45 clusters of cells, with each cluster composed of different frequencies of cells from B6 or IL-10KO individuals. (F) The proportion of events contributed by individual mice within the top six most enriched clusters in B6 or IL-10KO mice identified statistically significant clusters and clusters prominently driven by a single individual. (G) SPADE analysis, including a comparison of SPADE trees for B6 #1 and IL-10KO #1 mice, identified nodes that appear visually discrepant between these groups. Analysis of node frequency across individuals identified seven nodes that were statistically significantly increased in IL-10KO mice (H), with statistically significant nodes identified on SPADE trees for B6 #1 and IL-10KO #1 (I). Citrus analysis of cluster abundance identified multiple clusters that are B6 or IL-10KO biased (J), with analysis of cluster frequency across individuals (K). (L) The cellular phenotype of cluster 82243, defined by analysis of selected Citrus-generated line graph overlays. Individual symbols on all plots identify values from individual mice. All data are from optimized algorithm settings in Figs. 2 and 5–8. **p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001, unpaired *t* test, with statistical analyses subjected to multiple testing correction. ns, not significant.

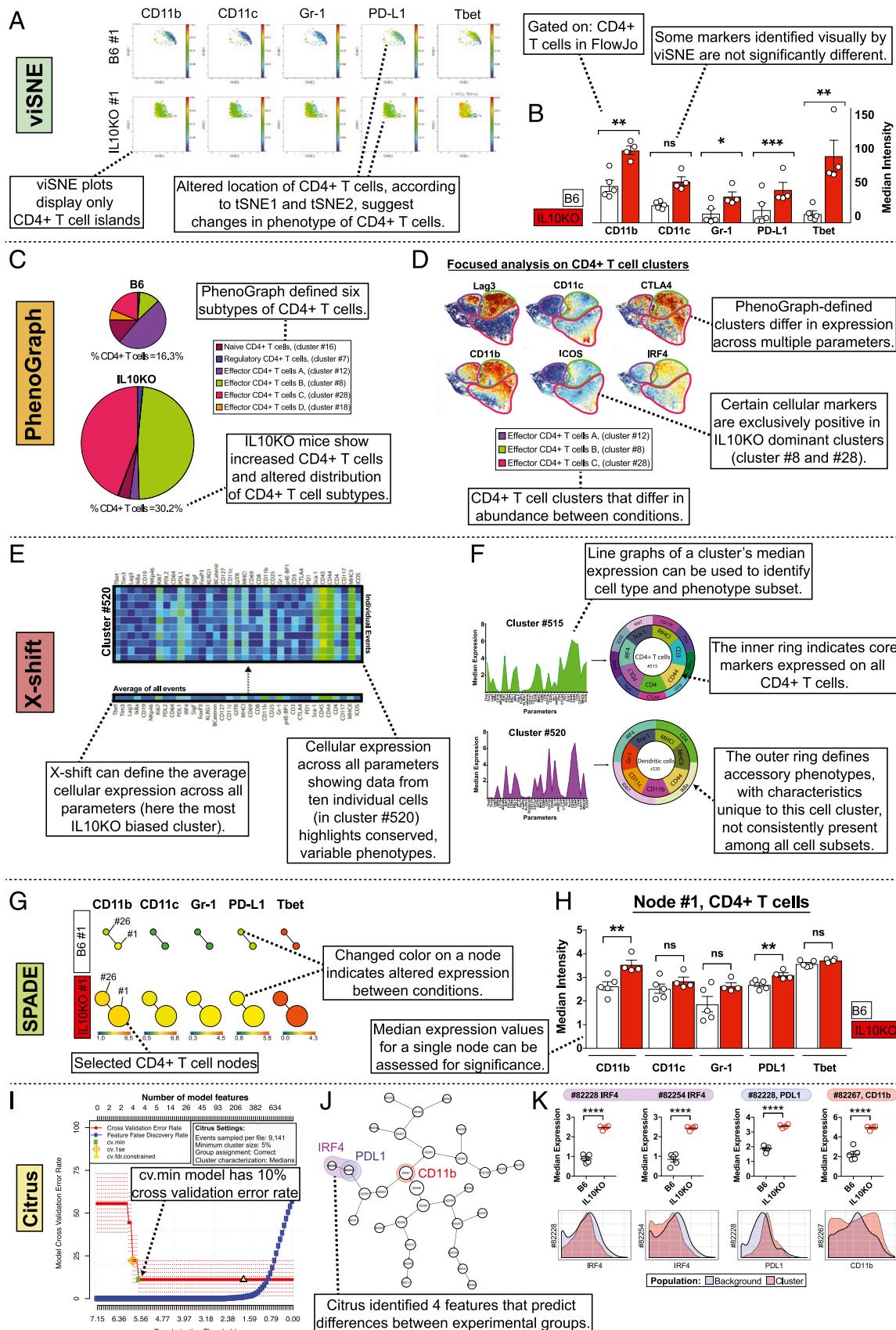


FIGURE 4. Investigating changes in cellular expression by viSNE, PhenoGraph, X-shift, SPADE, and Citrus algorithms. CyTOF analysis of γHV68-infected lungs from B6 or IL-10KO individuals harvested at 9 d postinfection. Data show all viable single cells, subjected to the various algorithms, and include examples demonstrating insights obtained across algorithms. **(A)** viSNE-driven analysis of CD4+ T cell phenotypes, with viSNE plots displaying gated CD4+ T cells characterized for expression across five markers that visually appeared to be different between B6 and IL-10KO mice. **(B)** The raw median intensity values for five cellular markers in CD4+ T cells, quantified by FlowJo, and plotted for all individuals with statistical significance, as indicated. *(Figure legend continues)*

using an unpaired *t* test comparing B6 and IL-10KO mice, with statistical significance as identified. For situations in which we tested statistical significance for all identified nodes/clusters, analysis was corrected for multiple comparisons by multiplying the individual *p* values for each comparison by the number of statistical tests performed.

Results

Conceptual framework for CyTOF analysis

CyTOF studies can interrogate a wide range of scientific questions, from focused deep phenotypic analysis of a specific cell subset to broad studies of cellular heterogeneity in complex cell populations (Fig. 1A). These datasets contain a wealth of data: cellular abundance and expression profiles (akin to flow cytometry studies), identification and quantification of cellular diversity, and changes in population structure, defined by changes in the relative frequencies of cellular populations and phenotypic subsets across experimental conditions (Fig. 1A). Although conventional flow cytometric algorithms and Boolean gating can give insights into some of these data (e.g., cellular abundance and expression), deeper data analysis quickly becomes inaccessible with >30 dimensions. In this article, we consider five CyTOF analysis algorithms that use computationally distinct methods, require different levels of computational skills, and vary significantly in data visualization (Fig. 1B).

One major distinction between the CyTOF algorithms considered in this article is whether cells are displayed as a continuum of phenotypes (e.g., viSNE) or stratified into subpopulations, to quantify population structure (e.g., PhenoGraph, X-shift) (Fig. 1B). These different types of data visualization and output significantly impact the data that can be retrieved. For example, changes in population structure are most readily obtained by algorithms that stratify all events into subpopulations (SPADE, PhenoGraph, X-shift) (Fig. 1B). Additional considerations of these algorithms are presented in Fig. 1B, with subsequent figures containing practical notes on how to use and to interpret each of these algorithms. These analyses primarily focus on a CyTOF dataset examining the immune response to γ HV68, a small animal model of herpesvirus infection (23). After sample collection, data were normalized (relative to equilibration beads), and viable single cells were subjected to data analysis using the analysis platforms viSNE, SPADE, X-shift, PhenoGraph, and Citrus. For each method, we define basic considerations in the application and interpretation of data and provide an example of how these algorithms can provide insight.

viSNE visualization and insights

viSNE is a well-established CyTOF analysis tool that uses the t-distributed stochastic neighbor embedding (tSNE) algorithm to analyze and display high-dimensional data on a two-dimensional map (7, 24). Practically, the resulting image shows a continuum of cellular phenotypes, distributed by the parameters tSNE1 and tSNE2, with cells colored according to expression of a chosen parameter (e.g., CD45). We used viSNE within Cytobank (<https://www.cytobank.org>.

org), a cloud-based computational platform, that requires three user-defined inputs: sample selection (i.e., which samples are to be analyzed), parameter selection (i.e., which phenotypic markers are to be used for the clustering analysis), and event sampling (which can be done by sampling proportionally or equally) (Fig. 2A). The resulting viSNE plot is a two-dimensional figure with the axes tSNE1 and tSNE2, with cells plotted on a continuum of expression with phenotypically related cells clustered together, often manifesting as phenotypic “islands.” Events can be colored according to any user-defined parameter (e.g., CD45) to identify the cellular identity of the island.

One of the first goals in interpreting a viSNE plot is to define where different cell types and phenotypes are located. This is defined empirically by the user and heavily relies on the user’s biological knowledge. For a complex mixture of cells (e.g., virally infected lung tissue), we find that assembling a grid of viSNE plots, organized by individual (in columns) and colored by different parameters (in rows), is particularly helpful (Fig. 2B). This grid allows rapid identification of cellular phenotypes across different islands, by visualizing where lineage markers are clustered (e.g., a B cell island, defined as CD45⁺ CD19⁺); definition of interindividual variation, with all events plotted according to a common axis of tSNE1 and tSNE2; and insights into populations and cellular phenotypes that change between experimental groups (e.g., wild-type B6 versus IL-10KO mice, Fig. 2B). Although viSNE plots can be annotated to identify the location of cell phenotypes, two independent viSNE runs on the same dataset will give two different plots that vary in terms of island location yet have relatively comparable island abundance and expression (Fig. 2C, 2D). Note that this variation between sequential viSNE runs in Cytobank is not simply a factor of variable events chosen during downsampling; it occurs even when the exact same cells are subjected to three sequential viSNE runs (Fig. 2E). Given interrun variability, it is only possible to directly cross-compare between experimental groups when all samples are subjected to the same viSNE analysis run. This approach ensures that all events are plotted according to a common scale of tSNE1 and tSNE2, with potential differences between conditions identified by manually overlaying viSNE plots for two experimental conditions. In this case, a gap in island distribution between two conditions can indicate an altered cell type or phenotypes (Fig. 2F). It is important to realize that there is no absolute location for any cell type in viSNE, as illustrated by a comparison of cellular distribution in a virus-infected lung versus a tumor-containing lung (Fig. 2G).

Investigating cellular abundance by viSNE. viSNE overlays comparing γ HV68-infected B6 and IL-10KO mice revealed multiple changes in the distribution of phenotypic islands, including changes in CD4⁺ T cells (Fig. 2F). Based on this visual difference, we next interrogated differences in the cellular abundance of CD4 T cells between γ HV68-infected B6 and IL-10KO mice. First, we analyzed the frequency of CD3⁺ CD4⁺ events between B6 and IL-10KO infected mice by manually gating the CD4⁺ T cell island on a viSNE plot (Fig. 3A). This analysis revealed that IL-10KO mice had a reproducible

(C) PhenoGraph analysis of CD4 T cell subsets identified six subsets of CD4 T cells with varied frequencies between B6 and IL-10KO mice. (D) Focused analysis of three PhenoGraph-defined clusters, obtained from a portion of the PhenoGraph-defined tSNE map (Fig. 5). Cluster boundaries are based on PhenoGraph, colored according to expression of six user-identified markers. (E and F) X-shift analysis of cluster 520, an IL-10KO-biased cluster, characterized by phenotypic barcodes to define the average expression and expression across the first 10 events in this cluster. (F) Line graphs of median expression values for two clusters revealed to be significantly increased in IL-10KO mice. Manually generated phenotype infographics reveal core and unique accessory phenotypes of each cluster. (G and H) SPADE analysis of two CD4⁺ T cell nodes shown to increase in IL-10KO mice, characterizing changes in expression across five user-defined parameters [parallel to (A)]. (H) Quantitation of median intensity value across multiple parameters, characterizing expression in SPADE node 1 for each individual. (I) Citrus analysis of medians identified a cv.min predictive model with 10% cross-validation error, characterized by four model features whose expression increased in IL-10KO mice. Proteins with altered expression were manually identified on a radial Citrus hierarchy tree (J), with median expression between groups quantified and visualized by Citrus-generated line graph overlays (K). Individual symbols on all plots identify values from individual mice. Data are from optimized algorithm settings in Figs. 2 and 5–8. **p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001, unpaired *t* test, with statistical analyses subjected to multiple testing correction. ns, not significant.

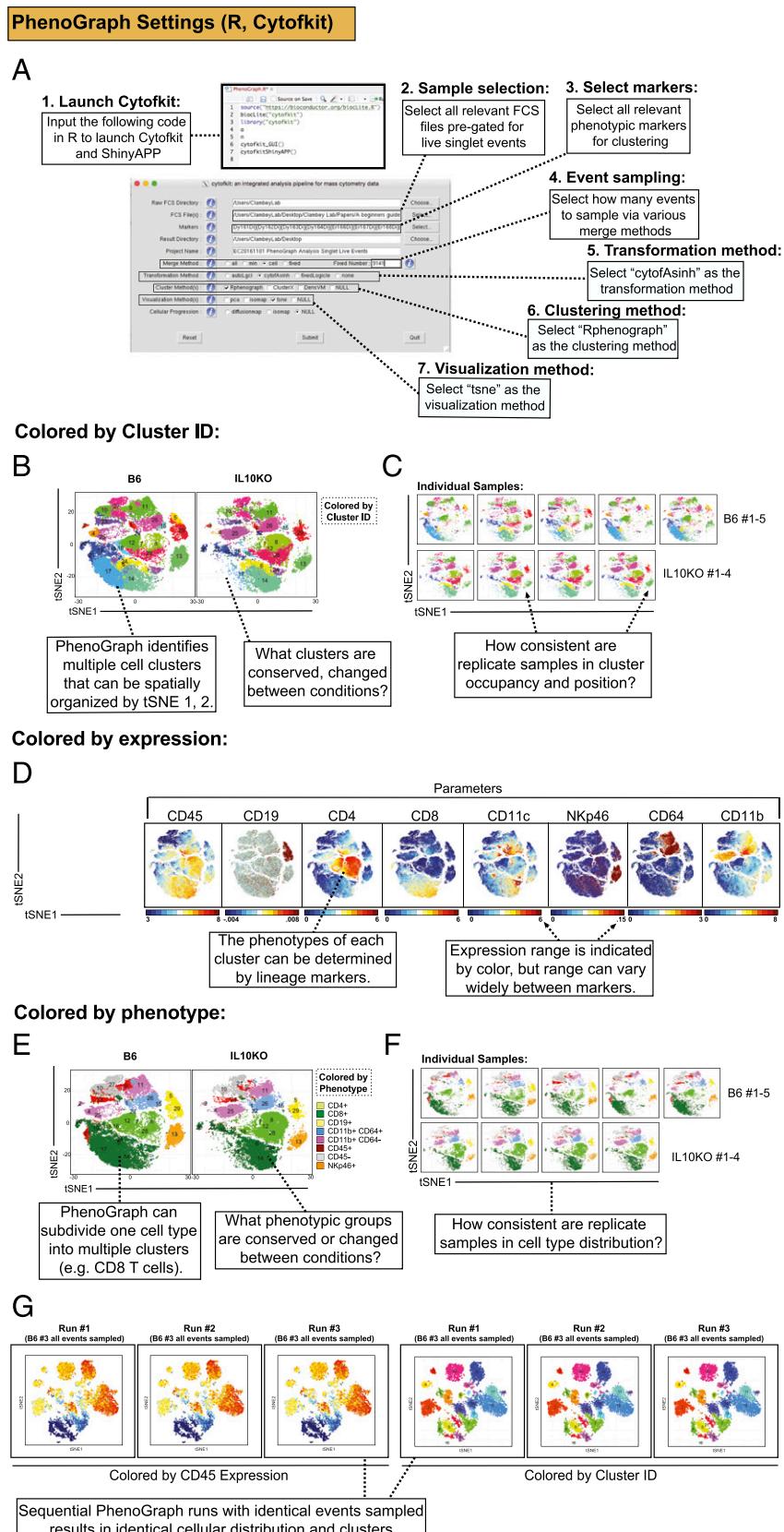


FIGURE 5. Basic considerations for PhenoGraph analysis. Input settings (A) and PhenoGraph data visualization (B–G), focused on CyTOF analysis of γ HV68-infected lungs from B6 or IL-10KO individuals at 9 d postinfection. Data show all viable single cells, subjected to PhenoGraph in cytofkit, which calculates the optimal amount of clusters, with data plotted on a tSNE plot. (A) Input settings to run the PhenoGraph algorithm in cytofkit. (B and C) PhenoGraph-defined cellular distribution and clustering, as defined by tSNE1 and tSNE2, colored by cluster for compiled B6 or IL-10KO samples (B) or for individual mice (C). (D) PhenoGraph-based visualization on a tSNE plot, colored according to expression of lineage markers, demonstrates cell clustering and varied scaling. (E and F) PhenoGraph visualization with clusters colored by phenotype in compiled B6 or IL-10KO samples (E) or for individual mice (F), with cell populations defined based on basic phenotypic markers according to the key. (G) Comparison of three sequential PhenoGraph runs, in which the same 9141 cells (from individual B6 #3) were subjected to PhenoGraph, visualized by CD45 (left three plots) or by cluster identification number (right three plots). Numbers identify the physical location of PhenoGraph-defined clusters. Data from virus-infected lungs (B6, $n = 5$; IL-10KO, $n = 4$ mice).

statistically significant increase in the percentage of CD4 T cells relative to infected B6 mice (Fig. 3B).

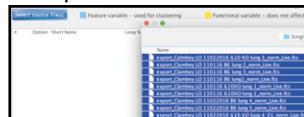
Investigating cellular expression by viSNE. Another notable feature defined by viSNE was the altered spatial placement of CD4 T cells

(according to tSNE1 and tSNE2) between these two groups, which suggested phenotypic differences between B6 and IL-10KO CD4 T cells (Fig. 4A). By analyzing viSNE plots across multiple parameters, and looking for altered expression levels defined by varying intensities, IL-10KO CD4 T cells appeared to have higher

X-shift Settings (VorteX)

A Dataset Importation:

1. Sample selection:



Select all relevant FCS files pre-gated for live singlet events
Determine how many events from each FCS file will be imported

2. Importation settings:



4. Successful dataset importation:



B Clustering settings:

1. Transformation method:

Select "arcsinh (x/f)" with $f = 5.0$

2. Noise threshold:

Apply a noise threshold of 1.0

3. Feature rescaling:

Select "none"

4. Normalization:

Select "none"

5. Distance Measure:

Select "Angular Distance"

6. Clustering Algorithm:

Select the "X-shift algorithm"

7. Density Estimate:

Select the density estimation method "N nearest neighbors"

8. Define K values:

Determine what K values will be calculated, (e.g. K values from 5 to 150 at 30 step intervals)

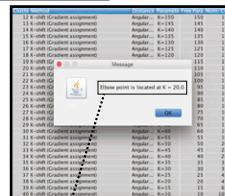
9. Number of neighbors:

Select "determine automatically"

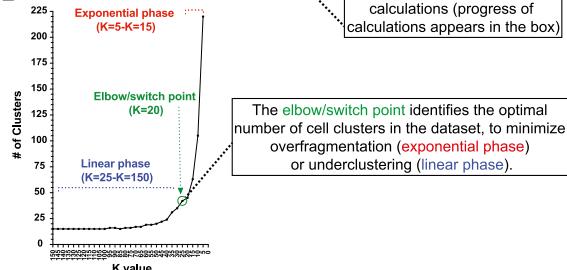
10. Begin clustering:

Press 'Go' to begin clustering calculations (progress of calculations appears in the box)

C



D



E Force directed layout settings:

1. Events sampled:

The number of events from each cluster can be selected by the user

2. Proportional sampling:

The events can also be sampled proportionally to the cluster size

3. Distance measurement:

Choose "angular distance"

4. Num. nearest neighbors:

Use the automatic value here of 10

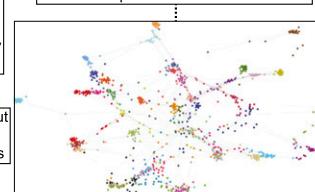
5. Edge settings:

The connections between nodes can be determined based on node density or by a specific parameter

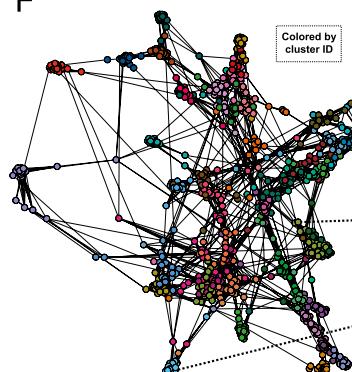
6. Sample selection:

A force directed layout can be created from one or more samples

The force directed layout generated in VorteX will look like this and can be colored by cluster ID or expression and can be visualized across individuals or experimental conditions.



F



Colored by cluster ID
The force directed layout can be opened in the Gephi app to customize figure.
Force directed layout shows different cell clusters (phenotypes) spatially distributed based on phenotypic similarity.

Clusters can be identified by cluster ID or phenotype.

G

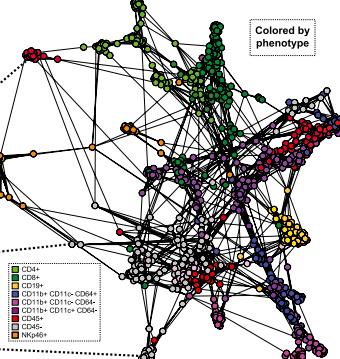
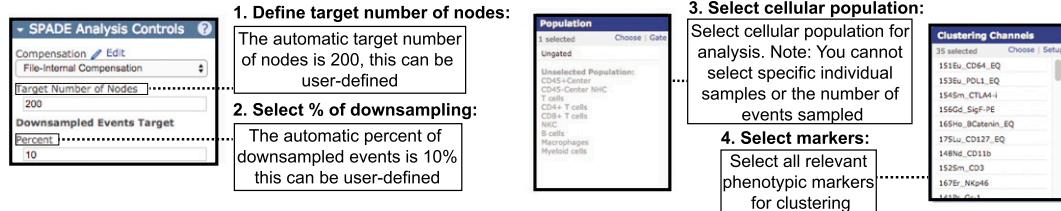
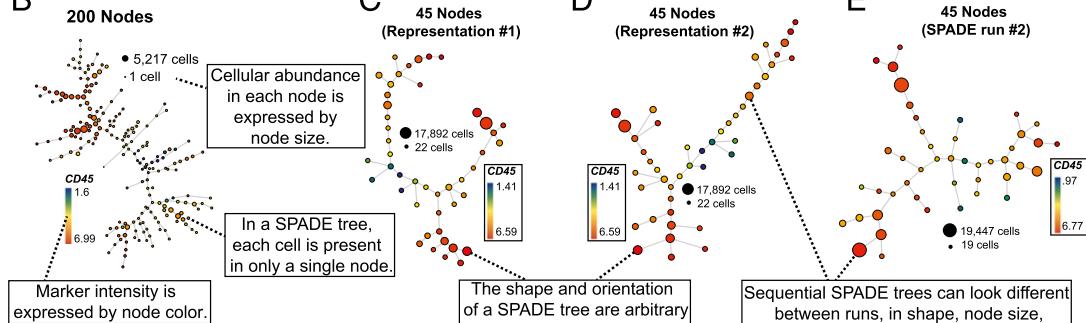
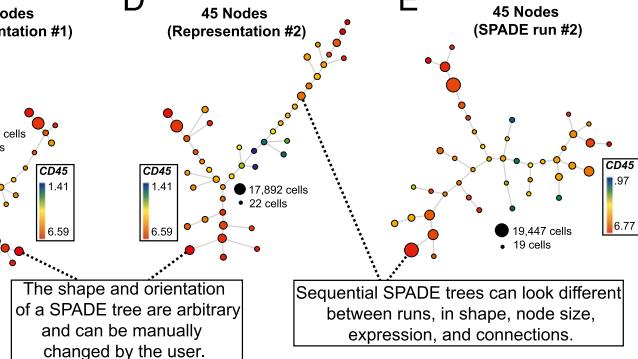
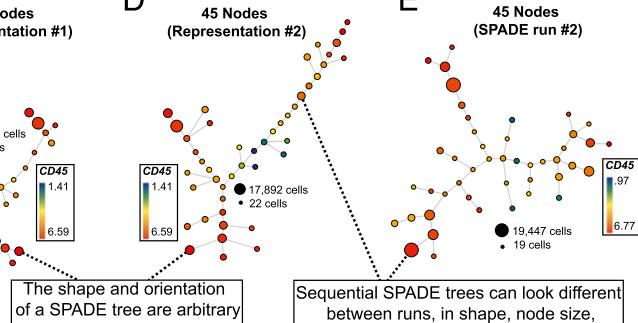
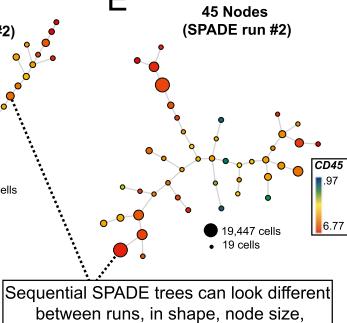
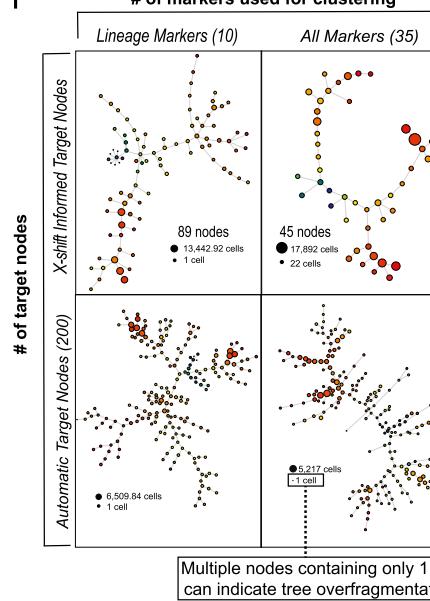
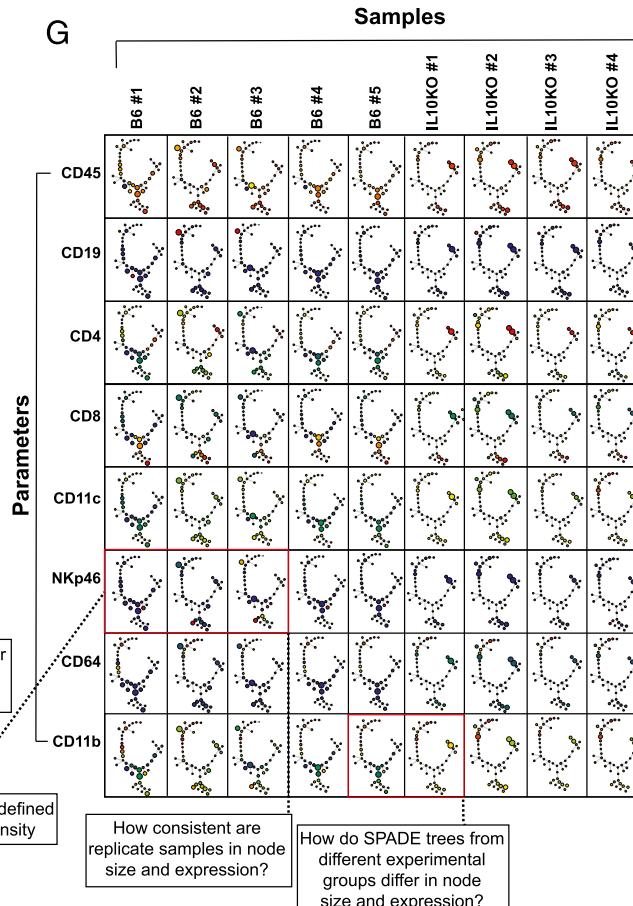
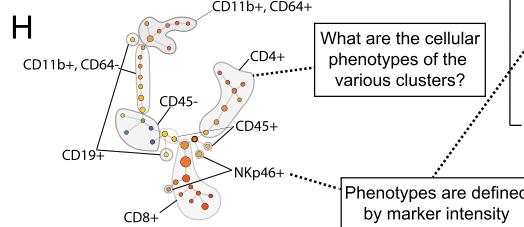


FIGURE 6. Basic considerations for X-shift analysis. Input settings (A–E) and X-shift visualization (F and G), focused on CyTOF analysis of γ HV68-infected lungs from B6 or IL-10KO individuals at 9 d postinfection. Data show all viable single cells, subjected to X-shift in the VorteX graphical environment, which calculates the optimal amount of clusters, with data plotted on a force-directed layout. Input settings to run X-shift for dataset import (A) and clustering settings (B). X-shift-defined clustering is depicted as a function of the number (K) of nearest neighbors tested, which can be used to calculate elbow point (C and D). The boundaries of the linear phase, switch point, and exponential phase are indicated. The force-directed layout curated by VorteX (E), and modified in Gephi (F and G), shows all 45 unique clusters identified at the $K = 20$ switch point, colored by cluster identification number (F) or by phenotype (G). Data from virus-infected lungs (B6, $n = 5$; IL-10KO, $n = 4$ mice).

SPADE Settings (Cytobank)

A**B****C****D****E****F****G****H**

How consistent are replicate samples in node size and expression?

How do SPADE trees from different experimental groups differ in node size and expression?

FIGURE 7. Basic considerations for SPADE analysis. Input settings (**A**) and SPADE visualization (**B–G**), focused on CyTOF analysis of γ HV68-infected lungs from B6 or IL-10KO individuals at 9 d postinfection. Data show all viable single cells, subjected to the SPADE algorithm, which clusters cells with similar protein expression levels into a customizable hierarchy. (**A**) Input settings to run SPADE within Cytobank. All of the events organized into a SPADE tree, colored by CD45 expression, using a target of 200 nodes (default setting) (**B**) or 45 nodes (X-shift informed) (**C–E**). (**C** and **D**) SPADE tree that has been manually modified by the user in Cytobank. (**E**) An independent SPADE analysis created from the same dataset using identical settings as in (**C** and **D**). (**F**) SPADE trees colored by CD45 generated based on clustering using 10 lineage markers (left) or 35 markers (right), comparing SPADE trees with an X-shift-defined optimal number of nodes (89 nodes for 10 lineage markers, 45 nodes for 35 marker clustering; upper panels) with a 200 target node tree (lower panels); (*Figure legend continues*)

expression of multiple proteins, including CD11b, CD11c, Gr-1, PD-L1, and Tbet (Fig. 4A). These findings were further corroborated by calculating median fluorescent intensities across individuals in FlowJo, a conventional flow cytometry analysis software (Fig. 4B). These data suggest heightened CD4 T cell activation in γ HV68-infected IL-10-deficient mice. This further exemplifies how integration of the tSNE analysis platform with conventional cytometric analyses can power new insights.

PhenoGraph visualization and insights

PhenoGraph is an algorithm that relies on graphs/networks of the recorded events and the connections between them to determine the accurate clustering of events into phenotypic categories (10). In contrast to tSNE, which portrays cells as a continuum of phenotypes, PhenoGraph stratifies all events into subpopulations (i.e., clusters), providing the user with a rapid quantification of population structure, defined as the diversity of cell types and phenotypes among all cells.

We ran the PhenoGraph algorithm in the R-based cytofkit package, a software package that allows users to run a variety of algorithms and data visualizations (25). We then used the Shiny application in R to examine, customize, and interrogate the results. Cytofkit requires six user-defined parameters, for which we selected tSNE as the visualization method, to allow cross-comparison with tSNE results (Fig. 5A). The results provide a two-dimensional figure plotted according to tSNE1 and tSNE2, like tSNE, with one notable addition: PhenoGraph stratifies all events into cell clusters, colored and numbered according to cluster identification number (Fig. 5B). For example, PhenoGraph identified 29 cell clusters that are present in various frequencies between γ HV68-infected B6 and IL-10KO cohorts, plotted according to tSNE1 and tSNE2 (Fig. 5B). This visualization can examine a composite of all animals in a condition or interrogate individuals, allowing rapid cross-comparison between individuals (Fig. 5B, 5C). Similar to tSNE, the resulting plots can be colored according to expression levels for various markers (Fig. 5D), to assist in the identification of cell types and subtypes (Fig. 5E, 5F). However, in contrast to tSNE run in Cytobank, we found that sequential PhenoGraph runs on the exact same cells gave reproducible cellular distribution and cell cluster identification between runs (compare Figs. 2E and 5G).

Investigating cellular abundance by PhenoGraph. When we visually compared the frequencies of PhenoGraph-identified clusters between B6 and IL-10KO individuals, clusters 4, 5, 8, 17, and 25 appeared to be dramatically different in cellular abundance (Fig. 3C). However, when we analyzed the frequencies of all clusters for statistical significance, only cluster 8 (CD4 T cells) and cluster 25 (myeloid cells) were found to be significantly increased in IL-10KO mice (Fig. 3D). Clusters that did not reach statistical significance had high interindividual variation, particularly in B6 individuals.

Investigating cellular expression by PhenoGraph. To better understand CD4 T cell populations identified by PhenoGraph, we analyzed the frequency, distribution, and phenotype of CD4 $^{+}$ T cell clusters between B6 and IL-10KO mice. PhenoGraph defined six subsets of CD4 $^{+}$ T cells in virally infected lungs (Fig. 4C). IL-10KO mice had an increased frequency of CD4 T cells, with a pronounced shift in the frequencies of four CD44 $^{\text{high}}$ effector T cell subsets (arbitrarily defined as effector types A–D, Fig. 4C). To determine how these CD4 $^{+}$ T cell effector subtypes differ phenotypically, and perhaps functionally, we visualized markers that appeared to have the greatest variation between CD4 T cell subsets (Fig. 4D). Although B6

individuals had a high frequency of type A (ICOS $^{+}$ CTLA4 $^{+/-}$ Lag3 $^{+/-}$ IRF4 $^{\text{mid}}$) effector CD4 T cells, IL-10KO infected mice had higher frequencies of type B (CD11b $^{\text{mid}}$ CD11c $^{\text{mid}}$ ICOS $^{+}$ CTLA4 $^{+}$ Lag3 $^{+}$ IRF4 $^{\text{hi}}$) and type C (CD11b $^{\text{mid}}$ CD11c $^{\text{mid}}$ ICOS $^{+/-}$ CTLA4 $^{+/-}$ Lag3 $^{-}$ IRF4 $^{\text{mid}}$) effector CD4 T cells (Fig. 4D). These phenotypic subtypes of CD4 T cells were uniquely identifiable because of PhenoGraph-defined clustering and were not predicted based on current literature.

X-shift visualization and insights

The X-shift algorithm, run in the Vortex visual space, estimates the number of cell clusters in high-dimensional data by using weighted k -nearest-neighbor density estimation (13). To do this, X-shift defines the impact of different numbers of nearest neighbors on the number of cell clusters, with the optimal number of cell clusters defined to occur at the “switch point” between underclustering and overfragmenting data. Once optimized cell clustering is calculated for a dataset, these data can be visualized by a number of methods. For purposes of this illustration, we show a force-directed layout in which cell clusters are physically clustered or separated based on relative similarity.

We ran the X-shift algorithm within Vortex. This algorithm has a high-degree of user-defined inputs, from dataset importation to clustering settings, prior to the generation of a graph that allows identification of the elbow/switch point, which is the optimized estimate of cell clusters (Fig. 6A–C). When we applied the X-shift algorithm to our virus-infection data set, we calculated a K value of 20 for the switch point between the linear and exponential phase, which corresponded to 45 clusters (Fig. 6C, 6D). The force-directed layout allows further user customization that typically results in a relatively low-resolution force-directed layout (Fig. 6E). This force-directed layout can be saved as a graphml file, opened in the Gephi app, where the size and color of nodes and edges can be manipulated, and further edited in Adobe Illustrator. In this case, the force-directed layout was colored by cluster identification number or by phenotype (Fig. 6F, 6G). Despite the high degree of computational effort required for running X-shift, it is notable how much data can be extracted from this algorithm (discussed below).

Investigating cellular abundance by X-shift. One notable feature of X-shift, and the Vortex graphical interface, is that it is possible to identify the individual events (i.e., cells) that contribute to each cluster. This information is incredibly useful, because different clusters are made up of variable amounts of cells derived from different individuals and experimental groups. To visualize this, we defined the percentage of each cluster that was from B6 or IL-10KO mice and ranked cell clusters in virally infected B6 and IL-10KO mice, from most B6 biased to most IL-10KO biased (Fig. 3E). This analysis identified some clusters that were equally present in B6 and IL-10KO mice (e.g., 514, 530), some B6-biased clusters (e.g., 504, 524), and some IL-10KO-biased clusters (e.g., 495, 520). Notably, when we calculated the frequency of events contributed by each individual mouse for the six most-biased clusters, only two IL-10KO-biased clusters were significantly different (515, 520), with other clusters driven by a single individual (504, 510) (Fig. 3F). These data emphasize a limitation of many of the clustering algorithms: the challenge of identifying interindividual variation.

Despite the presence of biased clusters, no cell clusters were found exclusively in B6 or IL-10KO mice. However, one notable use of this X-shift-driven approach is that it can be applied to situations in which a cell type is either present or absent. To demonstrate this, we used X-shift to analyze cellular complexity in an orthotopic mouse

default Cytobank setting). (G) Visualization grid of SPADE trees, with data organized according to marker expression (rows) relative to individuals (columns). (H) Identification of cellular populations identified by the SPADE tree for individual B6 #1, with phenotype-based cell populations as identified. Data from virus-infected lungs (B6, $n = 5$; IL-10KO, $n = 4$ mice).

Citrus Settings (Cytobank)

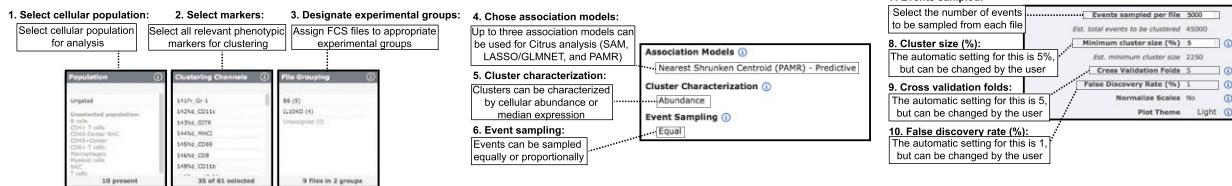
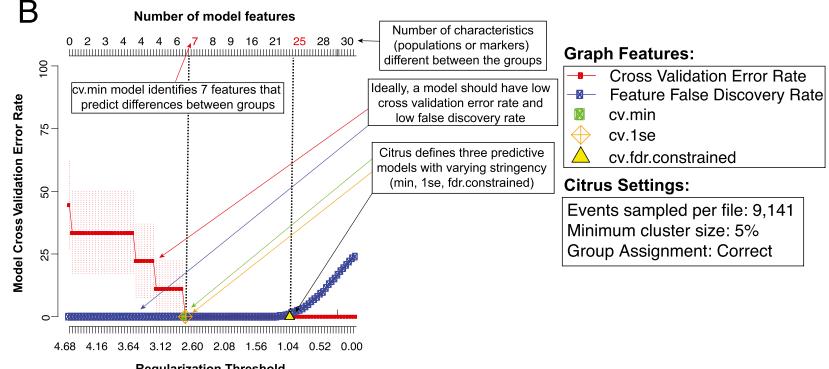
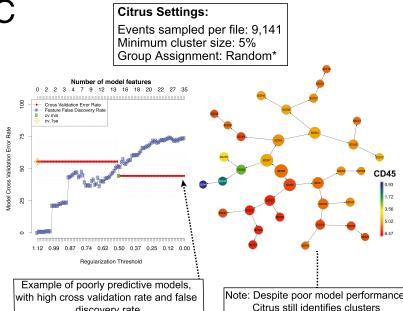
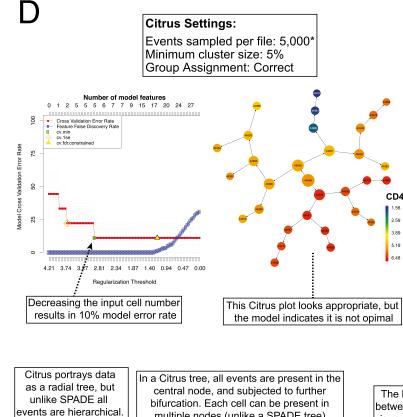
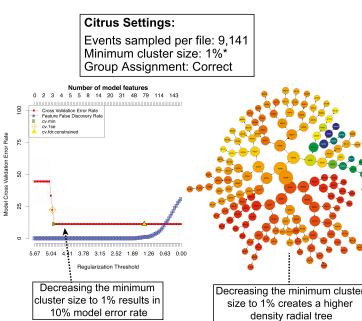
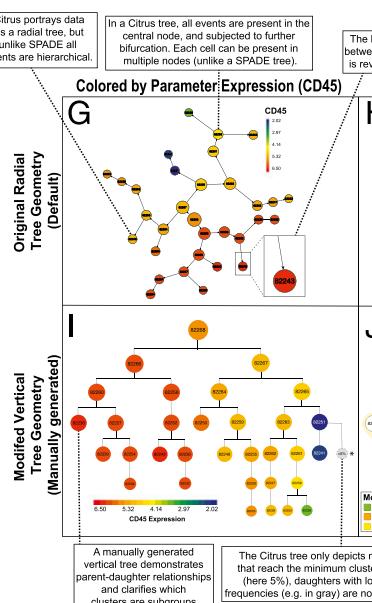
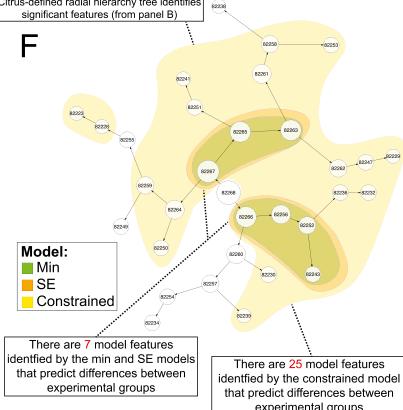
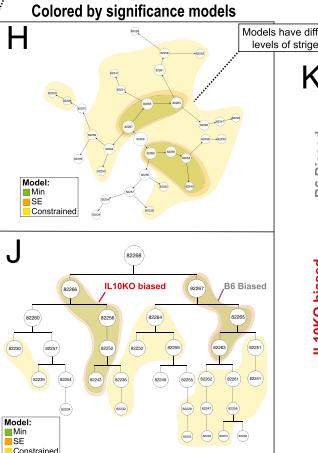
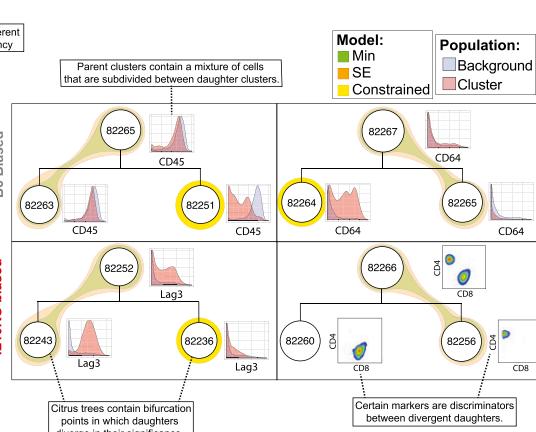
A**B****C****D****E****F****G****H**

FIGURE 8. Basic considerations for Citrus analysis. Input settings (**A**) and Citrus visualization (**B–K**), focused on CyTOF analysis of γ HV68-infected lungs from B6 or IL-10KO individuals at 9 d postinfection. Data show all viable single cells, subjected to the Citrus (cluster identification, characterization, and regression) algorithm that hierarchically clusters cells and identifies statistically significant biological differences between two or more parameters. (**A**) Input settings to run Citrus within Cytobank. (**B**) Citrus-generated model error rate plot, which defines cross-validation rate and feature false discovery rate for three models of statistical stringency (cv.min, cv.1se, cv.fdr.constrained). Vertical dotted lines were added to better illustrate how many model features were identified by each model. Model error rate plot (left panels) and radial hierarchy tree colored by CD45 expression (right panels), comparing different input settings (alternate setting identified by asterisk) for Citrus analysis, using randomized group assignment (**C**), reduced input cell number (5000 input) (*Figure legend continues*)

model of lung cancer. This approach identified cell clusters that were uniquely present in the tumor-containing lung (886, 887), suggesting that X-shift can identify tumor cells, even in the absence of a unique marker for the tumor cell (Supplemental Fig. 3).

Investigating cellular expression by X-shift. The VorteX graphical environment has many unique visualization features, including the phenotypic barcode, which shows expression levels across all parameters for a cluster or for individual cells (Fig. 4E). Analysis of expression in individual cells can be particularly informative; it identifies the basic cellular phenotype of the cluster, defines secondary characteristics that make the cluster unique, and allows investigation of phenotypic heterogeneity/homogeneity of events within clusters (Fig. 4E). VorteX also allows analysis of expression of all parameters within a cluster by a line graph (Fig. 4F). To distill this information further, we have found that infographics can be particularly useful, for example, depicting each cluster by its core phenotypes (e.g., features present among all CD4 T cells) and accessory phenotype (e.g., features present in only a subset of CD4 T cells) (e.g., cluster 515, Fig. 4F).

SPADE visualization and insights

SPADE was the original dimensionality reduction tool created for CyTOF and remains commonly used today (6, 26). The SPADE algorithm first downsamples data to capture rare populations then it hierarchically clusters phenotypically similar cells into “nodes,” sorts the remaining data into these nodes, and then represents these nodes and the relationships between them in a minimum spanning tree format (i.e., a SPADE tree).

We used SPADE within Cytobank, a platform that requires four user-defined inputs: target number of nodes (i.e., how many nodes will be present on the SPADE tree), percentage downsampling (i.e., what percentage of events will be considered for the analysis), population selection, and parameter selection (i.e., which phenotypic markers are to be used for the clustering analysis) (Fig. 7A). The resulting SPADE tree contains a series of interconnected nodes, with node size indicating cell number and node color quantifying the parameter of interest (e.g., CD45) (Fig. 7B). One important distinction about the SPADE tree is that the relative distance between different nodes does not present meaningful information (e.g., in contrast to the force-directed layout method used in X-shift). As such, users can manually modify the appearance of a SPADE tree within Cytobank (Fig. 7C, 7D). Note that as a result of downsampling, multiple runs with identical settings will produce slightly different SPADE trees (Fig. 7E).

One important consideration in using SPADE is that the user must decide how many target nodes should be present in a SPADE tree. For example, when we applied the SPADE algorithm to our data using a Cytobank default setting of 200 target nodes, the resulting SPADE tree contained many nodes that contained only a single event, suggesting data overfragmentation (Fig. 7F, lower right panel). Although the target number of nodes can be empirically modified, it is difficult to know what is truly optimal. To minimize overfragmentation, we integrated two approaches, first applying the X-shift algorithm (discussed above) to define the number of cellular phenotypes (in this case, 45 clusters), which we then used to generate a SPADE tree with

an informed number of nodes (Fig. 7F, upper left panel). Another important consideration when using SPADE, and all clustering algorithms, is how many parameters are included in clustering analysis. For example, we performed clustering on all 35 markers or clustering based only on lineage markers (CD3, CD4, CD8, CD19, CD45, CD64, CD117, NKp46, Sca1, SigF), plotted onto SPADE trees using a 200-node default or an X-shift–defined target node. Although changing the number of clustering parameters altered the SPADE tree structure, the most noticeable impact was that X-shift predicted 89 clusters based on 10 clustering parameters but only predicted 45 clusters based on 35 clustering parameters (Fig. 7F). Although each graph allows one to visualize cell populations, a major limitation of SPADE is that there are no clear quality-control metrics to identify optimal versus suboptimal analyses.

Once a SPADE tree is generated, one primary objective is to define which cells are present in which node. We used an optimized SPADE tree, informed by X-shift and clustered on all 35 parameters, to generate a visualization grid comparing SPADE trees between individuals (in columns) as a function of marker expression (in rows) (Fig. 7G). Because the branches of the SPADE tree are common across samples, and nodes vary in marker intensity and size, this approach allows rapid inspection across individuals and phenotypes. Much like viSNE, this allows rapid assessment of cellular identity of various nodes and interindividual reproducibility and variability (Fig. 7H).

Investigating cellular abundance by SPADE. The SPADE tree can be a powerful visualization tool, allowing rapid clues about potential differences between groups. For example, by visual inspection it appears that nodes 1, 9, 12, and 26 had the most dramatic change in size between B6 and IL-10KO mice (Fig. 3G). However, when we calculated the frequency of events across all nodes, it is clear that visual inspection alone is insufficient. For example, nodes 9 and 12 demonstrated high interindividual variation (Fig. 3H). Conversely, multiple nodes were statistically significant between B6 and IL-10KO mice (1, 2, 18, 26, 30, 31, and 37), including some changes that would have been easily overlooked by visual inspection (e.g., nodes 30 and 37) (Fig. 3I). These data emphasize the value in integrating quantitative and visual analyses to fully interrogate a SPADE analysis.

Investigating cellular expression by SPADE. SPADE also allows rapid analysis of expression across multiple parameters. For example, we analyzed markers previously characterized by viSNE (Fig. 4A). By eye, IL-10KO mice appeared to have increased expression of CD11b, CD11c, Gr-1, and PD-L1 across nodes 1 and 26 (Fig. 4G). However, when we quantified median expression of these markers for node 1 across all individuals, we found that only CD11b and PD-L1 had statistically significant increases in median expression (Fig. 4H). These studies further emphasize the necessity of integrating SPADE visualization with quantitative data to avoid misinterpretation.

Citrus visualization and insights

The Citrus algorithm was designed to identify statistically significant differences in high-dimensional datasets between different experimental groups (11), and it has been able to identify immune cell signatures of patients that correlate with divergent surgical outcomes

cells per sample) (D), or reduced minimum cluster size (1%) (E). (F) Citrus-defined radial hierarchical plot for optimized Citrus settings (B), shaded according to the statistical significance of three models. (G) Citrus-defined radial hierarchical plot colored by CD45 expression, with a magnified section of the tree to better illustrate node connections. (H) Citrus-defined radial hierarchical plot shaded by statistical significance according to different models [identical to (F), included for comparison]. (I) A Citrus-defined vertical hierarchical tree, colored by CD45 expression, generated manually from (G). The gray cluster, identified by an asterisk (*) and labeled <5% indicates a ghost daughter whose abundance is <5% and is, therefore, excluded from the hierarchical tree. Note that the asterisk in this case denotes manual inclusion of this population, not statistical significance. (J) A Citrus-defined vertical hierarchical tree, shaded by statistical significance according to three models. (K) Illustration of parent–daughter relationships in a Citrus tree, identifying markers whose expression bifurcates between daughters. Data from virus-infected lungs (B6, $n = 5$; IL-10KO, $n = 4$ mice).

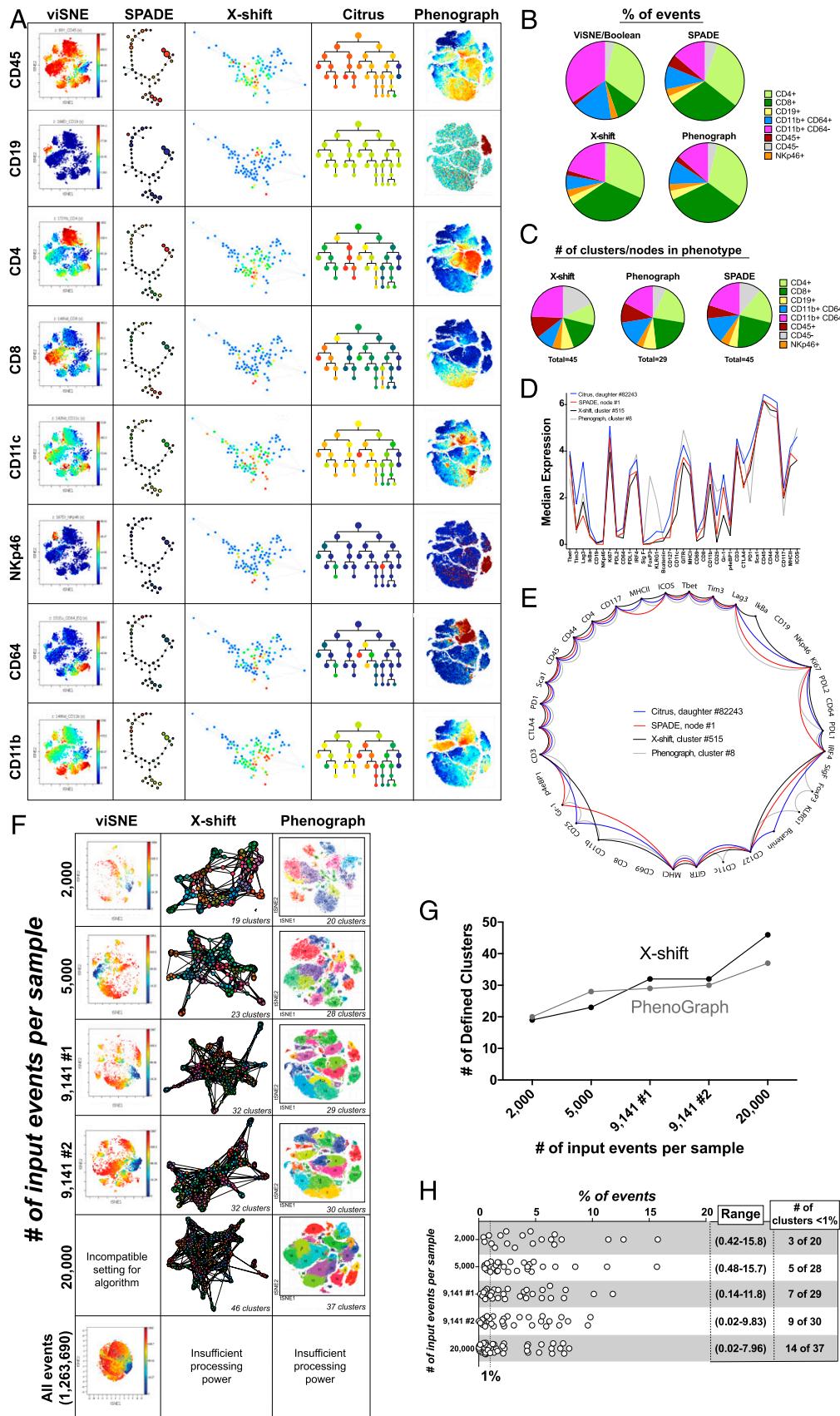


FIGURE 9. Comparison of data visualization, cellular identification, and reproducibility across CyTOF algorithms. CyTOF analysis of γHV68-infected lungs from B6 or IL-10KO individuals harvested at 9 d postinfection. **(A)** Direct cross-comparison of data visualization across multiple phenotypic markers (rows), comparing different algorithms (columns), including plots shown in previous figures. **(B)** Quantitation of the percentage of events identified as different cell phenotypes, comparing tSNE/Boolean gating, SPADE, X-shift, and PhenoGraph. **(C)** Comparison of the number of clusters/nodes identified according to each phenotype, comparing X-shift, PhenoGraph, and SPADE. Comparison of CD4 T cell clusters/nodes identified as significantly increased. (Figure legend continues)

(27). Citrus accomplishes this by compiling events across multiple samples, hierarchically clustering cells by similarity, and then interrogating whether these populations differ significantly between user-identified experimental conditions. Citrus tests for statistical changes in abundance or in median expression by using predictive or correlative linearized regression models, contingent on sufficient sample size and statistical power ($n \geq 3$, with higher n associated with increased power) (11).

We investigated Citrus within Cytobank, a platform that requires a series of 10 user-defined inputs, including sample and parameter identification, for a statistical model to be evaluated (11) and criteria to be used for the statistical analysis (e.g., false discovery rate and minimum cluster size) (outlined in Fig. 8A). For our purposes, we tested for predictive models that had statistical power to discriminate between groups.

Once a Citrus run is completed within Cytobank, multiple figures are returned. Based on our experience, it is critical to first analyze the Model Error Rate figure, which provides basic details about the validity of the statistical model. This figure plots the cross-validation rate (in red), quantifying how frequently the model correctly identifies whether a sample is in condition A or B, and the feature false discovery rate (in blue) (Fig. 8B). Superimposed on these two parameters, this graph further identifies three statistical models with differing stringencies: the cross-validation (cv).min model, which identifies the minimal number of features that predict the lowest cross-validation error rate between experimental groups; the cv.1se model, which identifies the fewest features with a cross-validation error rate 1 SE greater than the minimum; and the cv.fdr.constrained model, which identifies the maximum number of features that can be included in the model yet remain below the user-defined false discovery rate (<https://support.cytobank.org/hc/en-us/articles/226678087-How-to-Configure-and-Run-a-CITRUS-Analysis>).

Ideally, a Citrus-defined predictive model would achieve a cross-validation rate of 0 (i.e., the model can reliably predict whether a sample is in condition A or B) while having a low (<1%) false discovery rate (as in Fig. 8B). To demonstrate a Citrus analysis with no predictive power, we randomized experimental samples, intermingling B6 and IL-10KO samples into a hypothetical group A and B; the resulting models had no predictive power, with an ~50% cross-validation error rate (Fig. 8C). Note that Citrus-defined models can be significantly influenced by reducing the number of input cells (e.g., from 9141 events per sample to 5000 events per sample, Fig. 8D) or by altering the minimum cluster size (e.g., from 5 to 1%, Fig. 8E). In both cases, these altered settings resulted in a higher cross-validation error rate, indicating a reduced predictive power of the defined features. One critical challenge when using Citrus is that, even when it creates models with unacceptably high rates of error (i.e., no predictive power, as in Fig. 8C), the algorithm still identifies characteristics/features from the failing model. Thus, if a user does not verify the strength of the predictive model, it is easy to misinterpret features of the model. We strongly recommend including Citrus settings and the Model Error Rate plot when reporting Citrus results, to facilitate proper Citrus interpretation.

For data visualization, Citrus generates radial hierarchy trees (“featurePlots”), with all events originating from a central node (Fig. 8F). These plots can be visualized by parameter expression (e.g., CD45) (Fig. 8G) or according to statistical model, highlighting clusters that predict differences between the two experimental conditions (Fig. 8H).

Despite an apparent similarity to a SPADE tree, a Citrus radial tree portrays events in a hierarchical manner, with parent clusters giving rise to two daughters. This means that each daughter contains a subset of cells that was present in the parent cluster. As a result, a single cell is not confined to a single cluster in a Citrus tree (unlike in SPADE). To clarify these relationships, we created a vertical hierarchical tree to examine phenotype and statistical significance (Fig. 8I, 8J). This vertical hierarchical visualization approach avoids confusion with SPADE visualization, identifies branch points in which daughter populations diverge in their significance (i.e., where only one daughter is predictive of a difference, whereas the other daughter is not predictive of a difference), and identifies branch points with a single daughter, a context that occurs when the second, “ghost,” daughter is less than the minimum cluster size and is excluded from the hierarchical tree. A ghost daughter is illustrated by a gray dot labeled “<5%” (Fig. 8I).

The ultimate goal of Citrus is to identify cell clusters that are significantly different between groups. For example, the cv.min model (Fig. 8B) identified seven model features between groups, with these clusters visualized on the radial or vertical tree (Fig. 8H, 8J). However, inspection of the Citrus tree clearly shows that there are not seven distinct cell types that are significantly different, because many of the clusters are linked in direct parent–daughter relationships. Detailed analysis of cell phenotypes in these parent–daughter pairs clearly shows parent clusters contain a mixed population of cells (e.g., a cluster containing CD4 and CD8 T cells), subdivided between daughters (e.g., CD4 versus CD8 T cells) (Fig. 8K). This bifurcation of cell phenotypes can be particularly useful when the two daughter populations diverge in significance, with phenotypes associated with this bifurcation giving insight into the population of interest. Given the mixed phenotype of parent clusters, the most refined cell type designations are contained within the terminal branches of the Citrus tree.

Investigating cellular abundance. We used Citrus to identify clusters with significant differences in their abundance between wild-type and IL-10KO infected mice, using the minimum number of cells contained across all experimental samples (9141 events per sample), a minimum cluster size of 5%, and the nearest shrunken centroid predictive association model. Citrus identified three models (cv.min, cv.1se, cv.fdr.constrained) with varying numbers of features that had a high predictive value (0% false discovery rate, 0% cross-validation rate) (Fig. 8H, 8J). We focused on the cv.min model, which identified seven features that predict differences between groups (Fig. 3J). Although Citrus identified these clusters as significantly different, it does not automatically return statistical significance values. Therefore, we downloaded individual abundance values for each cluster to calculate p values (Fig. 3K). For each cluster, Citrus provides a series of graph overlays depicting expression across all parameters between the cluster and background. When we analyzed the most terminal clusters identified by the cv.min model, the B6-enriched cluster 82263 had a mixed phenotype (e.g., containing CD3⁺ and CD3⁻ events) (data not shown). In contrast, the IL-10KO-enriched cluster 82243 was uniformly CD4⁺ T cells, characterized by proliferation (Ki67⁺) and expression of multiple inhibitory receptors (CTLA4⁺ GITR⁺ PD1⁺ Tim3⁺). These studies emphasize the potential for Citrus to rapidly identify changes in cellular abundance, following careful interpretation of the results.

in IL-10KO mice, depicting median expression values (D) or a phenotype network identifying parameters that were positive (defined as higher than the average expression for all events) (E). Analysis of the impact of input cell number per sample on data visualization across algorithms, showing data visualization (F), cluster number (in X-shift and PhenoGraph) (G), and distribution of cluster frequencies (in PhenoGraph) (H). Data from virus-infected lungs (B6, $n = 5$; IL-10KO, $n = 4$ mice).

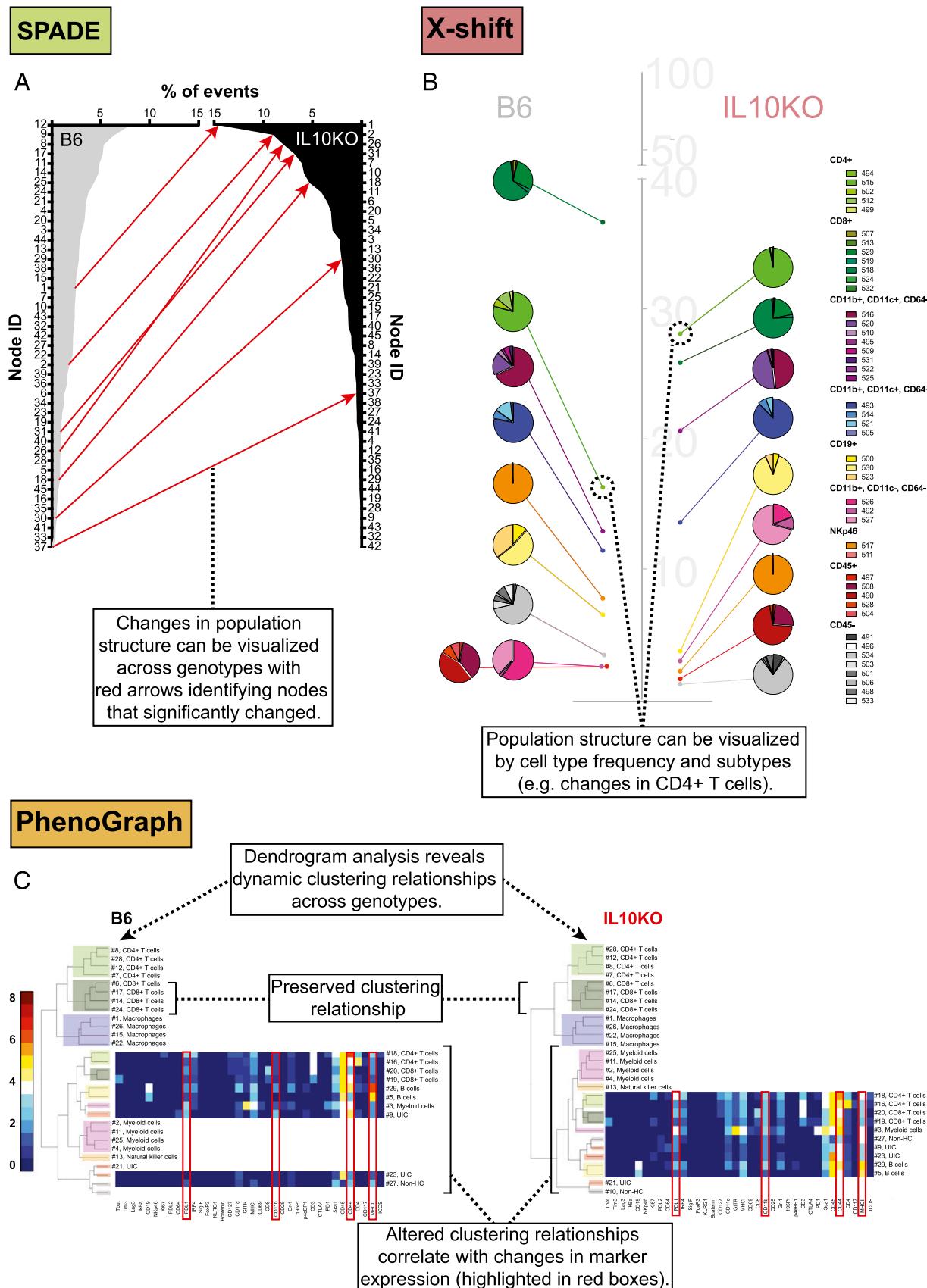


FIGURE 10. Population structure analysis using SPADE, X-shift, and PhenoGraph. CyTOF analysis of γ HV68-infected lungs from B6 or IL-10KO individuals harvested at 9 d postinfection. Data show all viable single cells, subjected to CyTOF analysis algorithms that stratify all events into discrete subsets. (A) SPADE definition of node frequency across all events defined altered frequencies and prominence between B6 and IL-10KO groups, with statistically significant changes identified in Fig. 3 annotated by red arrows. (B) X-shift-defined frequencies of cell types and subsets (identified by cluster number) revealed altered frequencies of cell types and changes in subset distribution within each cell type. (C) PhenoGraph-defined clustering relationships, visualized by (Figure legend continues)

Investigating cellular expression. In parallel, we used Citrus to test for changes in expression between experimental groups. This analysis returned a Model Error Rate with a low (~10%) cross-validation error rate, with cv.min and cv.fdr constrained models identifying multiple potential differences (Fig. 4I). Analysis of the cv.min model identified four features between groups, including increased expression of IRF4, CD11b, and PD-L1 in IL-10KO infected individuals (Fig. 4J). The significance of these differences in cellular expression was determined, and the expression of these markers relative to background expression was visualized by Citrus-generated graph overlays (Fig. 4K).

Comparing methods visualization and insights

Given the diverse CyTOF algorithms and visualization methods studied in this article, we directly compared data visualization from each of these algorithms side by side (Fig. 9A). This approach exemplifies the challenges of CyTOF data visualization and how a single method is insufficient to fully extract the robustness of these high-dimensional data (algorithm comparisons outlined in Supplemental Table III). Although viSNE and PhenoGraph emphasize the continuum of cellular phenotypes in two-dimensional space, as revealed by the tSNE algorithm, SPADE and Vortex reduce this complexity by placing multiple events into discrete bins. Citrus, in contrast, seeks to identify significant differences between groups and may not define the full cellular complexity across all events. For example, a parameter such as CD19, which does not vary between groups, can appear to be absent from a Citrus data-visualization tree, although it is clearly present by alternate algorithms (Fig. 9A). Moving beyond visualization, we quantified the frequency of events and the number of clusters/nodes assigned to various cellular phenotypes (Fig. 9B, 9C). Although all algorithms identified multiple major cell populations, there were also variations in the number of cells identified by different approaches. For example, CD8 T cells were disproportionately less frequent when analyzed by viSNE, with CD11b⁺ CD64⁻ myeloid cells less frequent when analyzed by SPADE and PhenoGraph (Fig. 9B). In addition, X-shift identified fewer CD4 T cell clusters, and more myeloid and nonhematopoietic clusters, than its counterparts (Fig. 9C).

A major goal in applying multiple algorithms to these data is to identify reproducible changes between experimental groups. To determine whether a particular cell type was identified independently across multiple methods, we selected the daughter/node/cluster that increased most significantly from B6 to IL-10KO individuals (identified in Figs. 3 and 4): daughter 82243 (Citrus), node 1 (SPADE), cluster 8 (PhenoGraph), and cluster 515 (X-shift). We then assessed marker expression across these methods, focusing on markers that were positive (i.e., higher than the average expression for all events) (Fig. 9D). Twenty-seven total parameters were positive among at least one of these four methods, and 17 parameters were identified as positive by all of the methods, visualized by a phenotype network (Fig. 9E). This cross-comparison shows conserved features among CD4 T cells identified across algorithms, as well as unanticipated variation.

How does cell number influence our interpretations? We examined the effect of varying input cell number on viSNE, X-shift, and PhenoGraph (Fig. 9F). For viSNE, low or high event counts generated suboptimal visualization that was either sparse or condensed (Fig. 9F). For X-shift and PhenoGraph, higher cell number increased visual compression and how many cell clusters were identified (Fig. 9F, 9G). Increased cluster identification appeared to be driven

by an increased identification of low-abundance populations (<1%) and an overall reduction in high-frequency clusters (Fig. 9H). These comparisons emphasize that input cell number can profoundly affect data output and visualization. It is also notable that these clustering algorithms are not well equipped to handle large datasets: viSNE (a cloud-based platform) took 11.5 h to analyze all events (1.26 million cells), whereas X-shift and PhenoGraph were unable to process all events on a standard desktop computer (Fig. 9F).

Investigating population structure

Beyond analysis of abundance and expression, CyTOF data can be used to define alterations in population structure, characterizing changes in cellular abundance and diversity between conditions. In our experience, shifts in population structure are most readily defined by stratifying algorithms (SPADE, X-shift, or PhenoGraph) that place cells into distinct nonoverlapping cell phenotypes. Changes in population structure can be visualized by multiple approaches. For example, by ordering node abundance between experimental conditions and then identifying SPADE-defined nodes whose abundance has significantly changed (red arrows), it is possible to see subdominant cell populations becoming dominant in IL-10KO mice (Fig. 10A). Clusters can be further plotted to reveal changes in population structure, focused on changes in cellular abundance and phenotypic subsets (Fig. 10B). This latter analysis reveals specific cell types that have undergone pronounced changes in abundance and phenotype (e.g., CD4 T cells), as well as populations that are relatively constant between conditions (e.g., CD19⁺ B cells). Finally, PhenoGraph in cytofkit can calculate clustering relationships by a dendrogram, to show the relatedness of cell clusters across experimental conditions. This is informative, because it reveals that clustering relationships are not always constant between experimental conditions (Fig. 10C). These insights into population relatedness, abundance, and phenotypic diversity are particularly accessible using stratifying CyTOF algorithms and have not been readily accessible using previous methodologies.

Discussion

The invention and application of mass cytometry transformed single-cell analysis, allowing deeper understanding of cellular heterogeneity among single cells at the molecular and phenotypic levels (3). Beyond the technical aspects of performing a mass cytometry experiment (recently reviewed in Ref. 28), there remain significant challenges in how to effectively analyze these complex data, particularly for the new user (17). The development of multiple algorithms, which use different data-analysis methods and visualization approaches, has enabled new insights, but it has also made it difficult to know where to begin for CyTOF analysis.

In this article, we present a practical guide for CyTOF data analysis using a common dataset to interrogate five established CyTOF analysis platforms. These algorithms varied widely in terms of ease of entry, computational skill required, data visualization, and extractability. To establish a resource for those interested in CyTOF data analysis, we have provided annotated figures to identify important steps in implementing and interpreting CyTOF data analysis, a direct cross-comparison of data visualization across these algorithms using a single dataset, and an illustration of insights that can be gained by using complementary algorithms.

On one end of the spectrum, algorithms run through the Cytobank platform (in this case, viSNE, SPADE, and Citrus) required minimal

dendrogram analysis, identified dynamic relationships between cell clusters in B6 (left panel) and IL-10KO (right panel) mice. Phenotypic markers that showed wide variation and appeared to correlate with changing clustering relationships are outlined by red boxes. Data from virus-infected lungs (B6, $n = 5$; IL-10KO, $n = 4$ mice).

computational background and gave complementary visualization approaches. Although viSNE demonstrates a continuum of phenotypes and can convey subtle variations, as well as rare populations, SPADE trees provide a highly simplified overview of the cellular phenotype and structure.

One notable consideration for these algorithms is the likelihood of misinterpreting or misapplying results. We found that SPADE and Citrus were more likely to give misleading results for an inexperienced user. For SPADE, an important and significant concern is how many nodes should be generated on the tree. The absence of clear guidelines, or the arbitrary use of a 200-node default, can easily lead to data overfragmentation. For Citrus, there are at least two major challenges: accurately interpreting statistical results for your Citrus run and understanding that a Citrus “tree” portrays parent–daughter relationships, such that a single cell can be present in multiple nodes of a Citrus tree (unlike SPADE, which places events in mutually exclusive nodes). A common limitation of each of these Cytobank-driven platforms is the lack of power to accurately predict total cluster number (i.e., approximately how many different types of cell clusters/phenotypes are present in your dataset). It is worth noting that, despite the caveats noted for Citrus, it remains the sole algorithm that provides a metric to quantify predictive power of the completed analysis.

We found that X-shift and PhenoGraph are particularly useful, given their ability to quantify the number of cell clusters in a population. This metric is particularly useful for understanding the diversity of cells in a CyTOF dataset. In addition, these algorithms have the potential to extract a large amount of data, from individual event expression (in X-shift/Vortex) to the interrelationship of cell phenotypes by dendrogram (in PhenoGraph within cytofkit). Despite these benefits, there are important considerations in using these algorithms. Both require increased computational background for the user, including familiarity with R. These algorithms are also very sensitive to input cell number, with an increasing number of cells correlating with more clusters. Given this caveat, any comparison between experimental conditions must be standardized based on cell number. Finally, these algorithms are not equipped with automated statistical analysis (unlike Citrus). Although these issues are important to consider, the ability of X-shift and PhenoGraph to transcend conventional user-defined populations can provide unprecedented insights into cellular complexity and phenotypic diversity and is particularly powerful in gaining new insights from CyTOF data.

One common challenge with the approaches analyzed in this article is that they rely on data display and visualization that is wholly different from conventional biaxial plots and Boolean gating, the gold standard by which cell populations have traditionally been analyzed by flow cytometry. Although flow cytometry typically defines populations through a series of exclusions and inclusions, clustering algorithms like PhenoGraph or X-shift can sometimes identify cell clusters of indeterminate phenotype, lacking expression of a definitive lineage marker (e.g., CD45⁺ cells that are CD3⁻ CD19⁻ CD64⁻ CD11c⁻). As such, assigning a unique and definitive identity to each cell cluster identified by these stratifying algorithms can be challenging. Based on this caveat, the Scaffold algorithm has been developed to allow a hybrid approach, combining Boolean gating to define major populations, which then serves as a scaffold on which to build a force-directed layout (12).

Although the primary focus of this article is on implementing and interpreting data from CyTOF algorithms, we did find populations that changed between γ HV68-infected B6 and IL-10KO mice. One consistent change defined across all algorithms was that γ HV68-infected IL-10-deficient mice showed an increased frequency of highly activated CD4 T cells in the lungs. Although each algorithm identified a common core of parameters expressed by these cells, including a

Ki67⁺ CTLA4⁺ GITR⁺ PD1⁺ Tim3⁺ phenotype, there were also discrepancies between the populations and phenotypes defined that will be the subject of future investigation. One may question whether CyTOF was required to identify an increase in effector CD4 T cells in IL-10-deficient mice. However, it is worth noting that the subsets and phenotypes of CD4 T cells identified by this approach would have been difficult to identify by conventional flow cytometry, because variations were not readily predictable based on precedent in the literature. Therefore, CyTOF afforded identification of unanticipated heterogeneity within effector CD4 T cell subsets, an area for future investigation. Whether this diversity of effector CD4 T cells is present in the context of lung cancer remains to be determined. These comparisons will require direct comparison of virus-infected and lung cancer samples, because it remains difficult to integrate and standardize data analysis across separate CyTOF runs.

In total, this study illustrates important practical considerations in the analysis of CyTOF datasets. Although established algorithms can allow critical insights into these complex data, we find that different algorithms provide complementary strengths, varying significantly in terms of the level of computation skill, the ability to extract additional data for further hypothesis testing, and the speed by which investigators can generate interpretable presentation-ready figures.

Based on our experiences, we would ultimately recommend integrating at least three algorithms/approaches to investigate CyTOF datasets, including a data-visualization method, to rapidly interrogate cellular phenotypes across high-dimensional data; a stratifying method (e.g., X-shift or PhenoGraph), to define the number and diversity of cellular clusters; and, in the case of experiments with at least three samples per condition, a method to define statistically significant differences in high-dimensional data (i.e., Citrus). The choice of data-visualization method will ultimately depend on the point you are trying to make, whether discussing a change in the populations (e.g., conveyed by SPADE or X-shift) or a change in the continuum of expression (e.g., conveyed by viSNE or tSNE-based algorithms). Given the changing landscape of CyTOF data analysis, we anticipate that this study will serve as an initial resource for those seeking to use CyTOF for their studies, as well as open the door to future discussion on best practices for analysis of these unique high-dimensional datasets.

Acknowledgments

We thank Amber Johnson, Jeff Kwak, and Melissa Ledezma for technical assistance; Kristina Terrell, Christine Childs, and Karen Helm for technical support for CyTOF studies; and Dr. Elena Hsieh for fruitful discussions regarding CyTOF analysis.

Disclosures

The authors have no financial conflicts of interest.

References

- 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.
- Bendall, S. C., E. F. Simonds, P. Qiu, el-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.
- Spitzer, M. H., and G. P. Nolan. 2016. Mass cytometry: single cells, many features. *Cell* 165: 780–791.
- 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 virus-specific cell niches within a continuum of CD8⁺ T cell phenotypes. [Published erratum appears in 2013 *Immunity* 38: 198–199.] *Immunity* 36: 142–152.
- Horowitz, A., D. M. Strauss-Albee, M. Leipold, J. Kubo, N. Nemat-Gorgani, O. C. Dogan, C. L. Dekker, S. Mackey, H. Maecker, G. E. Swan, et al. 2013. Genetic and environmental determinants of human NK cell diversity revealed by mass cytometry. *Sci. Transl. Med.* 5: 208ra145.

6. Qiu, P., E. F. Simonds, S. C. Bendall, K. D. Gibbs, Jr, R. V. Bruggner, M. D. Linderman, K. Sachs, G. P. Nolan, and S. K. Plevritis. 2011. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat. Biotechnol.* 29: 886–891. <https://doi.org/10.1038/nbt.1991>.
7. Amir, el-A.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.
8. Bendall, S. C., K. L. Davis, el-A. D. Amir, M. D. Tadmor, E. F. Simonds, T. J. Chen, D. K. Shenfeld, G. P. Nolan, and D. Pe'er. 2014. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157: 714–725.
9. Van Gassen, S., B. Callebaut, M. J. Van Helden, B. N. Lambrecht, P. Demeester, T. Dhaene, and Y. Saeys. 2015. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A* 87: 636–645.
10. Levine, J. H., E. F. Simonds, S. C. Bendall, K. L. Davis, el-A. D. Amir, M. D. Tadmor, O. Litvin, H. G. Fienberg, A. Jager, E. R. Zunder, et al. 2015. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* 162: 184–197.
11. Bruggner, R. V., B. Bodenmiller, D. L. Dill, R. J. Tibshirani, and G. P. Nolan. 2014. Automated identification of stratifying signatures in cellular subpopulations. *Proc. Natl. Acad. Sci. USA* 111: E2770–E2777.
12. Spitzer, M. H., P. F. Gherardini, G. K. Fragiadakis, N. Bhattacharya, R. T. Yuan, A. N. Hotson, R. Finck, Y. Carmi, E. R. Zunder, W. J. Fantl, et al. 2015. Immunology. An interactive reference framework for modeling a dynamic immune system. *Science* 349: 1259425.
13. Samusik, N., Z. Good, M. H. Spitzer, K. L. Davis, and G. P. Nolan. 2016. Automated mapping of phenotype space with single-cell data. *Nat. Methods* 13: 493–496.
14. Becher, B., A. Schlitzer, J. Chen, F. Mair, H. R. Sumatoh, K. W. Teng, D. Low, C. Ruehl, P. Riccardi-Castagnoli, M. Poidinger, et al. 2014. High-dimensional analysis of the murine myeloid cell system. *Nat. Immunol.* 15: 1181–1189.
15. Chester, C., and H. T. Maecker. 2015. Algorithmic tools for mining high-dimensional cytometry data. *J. Immunol.* 195: 773–779.
16. Mair, F., F. J. Hartmann, D. Mrdjen, V. Tosevski, C. Krieg, and B. Becher. 2016. The end of gating? An introduction to automated analysis of high dimensional cytometry data. *Eur. J. Immunol.* 46: 34–43.
17. Newell, E. W., and Y. Cheng. 2016. Mass cytometry: blessed with the curse of dimensionality. *Nat. Immunol.* 17: 890–895.
18. Diggins, K. E., A. R. Greenplate, N. Leelatian, C. E. Wogsland, and J. M. Irish. 2017. Characterizing cell subsets using marker enrichment modeling. *Nat. Methods* 14: 275–278.
19. Spitzer, M. H., Y. Carmi, N. E. Reticker-Flynn, S. S. Kwek, D. Madhireddy, M. M. Martins, P. F. Gherardini, T. R. Prestwood, J. Chabon, S. C. Bendall, et al. 2017. Systemic immunity is required for effective cancer immunotherapy. *Cell* 168: 487–502.e415.
20. Diebel, K. W., L. M. Oko, E. M. Medina, B. F. Niemeyer, C. J. Warren, D. J. Claypool, S. A. Tibbetts, C. D. Cool, E. T. Clambey, and L. F. van Dyk. 2015. Gammaherpesvirus small noncoding RNAs are bifunctional elements that regulate infection and contribute to virulence in vivo. *MBio* 6: e01670–e14.
21. Layton, M. G., and L. M. Franks. 1984. Heterogeneity in a spontaneous mouse lung carcinoma: selection and characterisation of stable metastatic variants. *Br. J. Cancer* 49: 415–421.
22. Li, H. Y., M. McSharry, B. Bullock, T. T. Nguyen, J. Kwak, J. M. Poczobut, T. R. Sippel, L. E. Heasley, M. C. Weiser-Evans, E. T. Clambey, and R. A. Nemenoff. 2017. The tumor microenvironment regulates sensitivity of murine lung tumors to PD-1/PD-L1 antibody blockade. *Cancer Immunol. Res.* 5: 767–777.
23. Barton, E., P. Mandal, and S. H. Speck. 2011. Pathogenesis and host control of gammaherpesviruses: lessons from the mouse. *Annu. Rev. Immunol.* 29: 351–397.
24. van der Maaten, L., and G. Hinton. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9: 2579–2605.
25. Chen, H., M. C. Lau, M. T. Wong, E. W. Newell, M. Poidinger, and J. Chen. 2016. Cytofkit: a bioconductor package for an integrated mass cytometry data analysis pipeline. *PLOS Comput. Biol.* 12: e1005112.
26. Anchang, B., T. D. Hart, S. C. Bendall, P. Qiu, Z. Bjornson, M. Linderman, G. P. Nolan, and S. K. Plevritis. 2016. Visualization and cellular hierarchy inference of single-cell data using SPADE. *Nat. Protoc.* 11: 1264–1279.
27. Gaudillière, B., G. K. Fragiadakis, R. V. Bruggner, M. Nicolau, R. Finck, M. Tingle, J. Silva, E. A. Ganio, C. G. Yeh, W. J. Maloney, et al. 2014. Clinical recovery from surgery correlates with single-cell immune signatures. *Sci. Transl. Med.* 6: 255ra131.
28. Brodie, T. M., and V. Tosevski. 2017. High-dimensional single-cell analysis with mass cytometry. *Curr. Protoc. Immunol.* 118: 5.11.1–5.11.25.