Chapter 2

High-Dimensional Modeling for Cytometry: Building Rock Solid Models Using GemStone™ and Verity Cen-se'™ High-Definition t-SNE Mapping

C. Bruce Bagwell

Abstract

This chapter outlines how to approach the complex tasks associated with designing models for high-dimensional cytometry data. Unlike gating approaches, modeling lends itself to automation and accounts for measurement overlap among cellular populations. Designing these models is now easier because of a new technique called high-definition t-SNE mapping. Nontrivial examples are provided that serve as a guide to create models that are consistent with data.

Key words Probability state model, t-SNE, Cen-se' mapping, High-dimensional modeling

1 Introduction

I can think of no analysis activity for cytometry that is as difficult yet fulfilling as high-dimensional modeling. It is very satisfying watching a well-designed model automatically analyzing hundreds of list mode files. In the past few years a number of models for various cytometry applications have been published [1–4], but many are proprietary and will never be published. In developing these models, a long list of practical steps, tricks, and tips has accumulated and it is the purpose of this chapter to organize and describe these hard-learned concepts in order that you do not need to waste time relearning them.

A new analysis tool called "t-SNE" was introduced to cytometry a few years ago. The t-SNE dimensionality reduction algorithm [5–7] is arguably the most exciting addition to the cytometry analysis arsenal in the last 5–10 years. A new high-resolution variant of the SNE method called Verity Cen-se'™ will be used to visualize high-dimensional data as simple dot-plot displays. This visualization process helps in the design of models that reflect the real biology of one or more populations in a sample.

Teresa S. Hawley and Robert G. Hawley (eds.), *Flow Cytometry Protocols*, Methods in Molecular Biology, vol. 1678, DOI 10.1007/978-1-4939-7346-0_2, © Springer Science+Business Media LLC 2018

This chapter will describe an approach for building rock solid models that can automate very complex cytometry applications. The approach involves the building and refining of probability state models [8] by leveraging published results and critically examining high-definition Cen-se' data patterns. Analyzing PBMC samples derived from the Helios^{tm} mass cytometer will serve as an example of how to approach building complex models that involve 35+ correlated measurements.

2 Modeling vs. Gating

2.1 Gating

Gating is a data partitioning methodology that separates one population from another via one or more measurement boundaries. Typically, gating involves drawing closed polygons around clusters of dots on two dimensional surfaces commonly called "dot-plots." Gating can also be performed with defined ranges over one-dimensional frequency distributions. Figure 1a shows two gating regions for one measurement, M1, which happens to have a bimodal distribution containing two populations labeled as "PopA" and "PopB." The ranges are placed to enumerate as much of their encompassed populations as possible without involving too many overlapping events. Because the underlying populations are not defined mathematically, the decisions for placing gates are often subjective and therefore difficult to reproduce and automate. Gates usually result in false positive and negative populations that can become confounding as gated event states cascade into other gates.

2.2 Frequency-Based and Probability-Based Modeling

Models on the other hand usually define the theoretical shape of one or more distributions and then "fit" the distributions to the data by optimizing a set of model parameters. As its name implies, a model attempts to describe one or more biological processes. Figure 1b shows a modeling approach to enumerating the "PopA" and "PopB" populations. A key attribute of a modeling approach is that it accounts for overlap between populations. Another key attribute of many models is that once it has fit data, it has the capability of synthesizing similar data. We will take advantage of that attribute in Subheading 5 when we discuss the synergistic relationships between modeling and Cen-se' mapping.

There are two fundamentally different approaches to modeling cytometry data. One approach represents data by fitting frequency on the *y*-axis and the measurement intensity on the *x*-axis as shown in panel b. Most early modeling theory and software packages were frequency-based [9] as depicted in panel b. The "best" parameters for the model are found by either minimizing or maximizing objective functions.

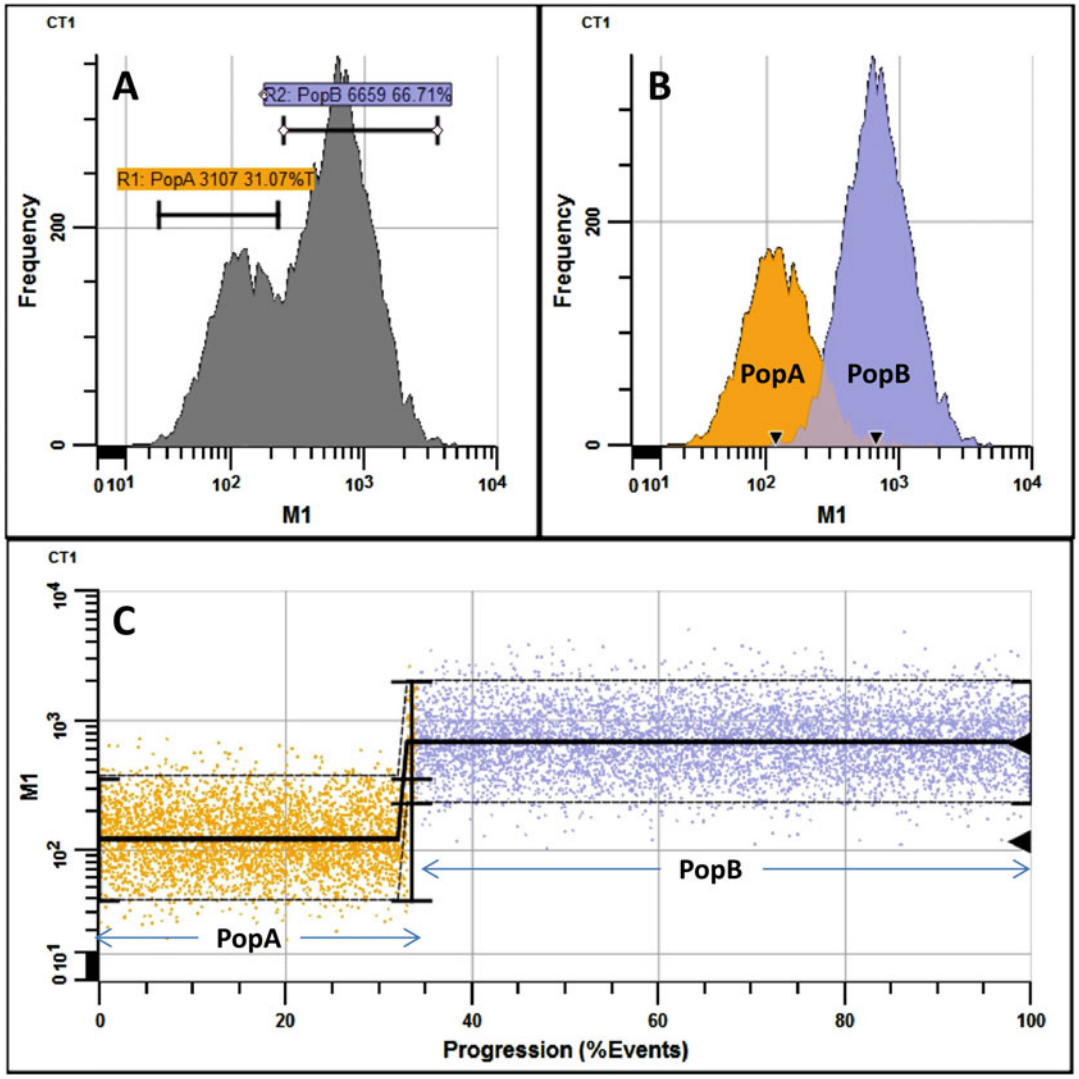

Fig. 1 Gating vs. modeling. (**a**) It shows two gating regions encompassing populations "PopA" and "PopB" for one measurement, *M1*. Ranges are placed to enumerate as much of their encompassed populations as possible without involving too many overlapping events. (**b**) It shows the two populations accounting for overlap. (**c**) It shows the same data in a probability state format where the *x*-axis is cumulative percent or "Progression" and the *y*-axis is *M1* intensity

Another type of data representation is shown in panel c. In this case, measurement intensity is the *y*-axis and cumulative percent is the *x*-axis. This type of configuration is used in probability state modeling (PSM, [4, 8, 10–12]). There are two major advantages for this type of data representation. The first is that process directions can be assigned to PSM models. For example, "PopA" in panel C can be interpreted as marking the beginning of a cellular progression and "PopB," the ending, since "PopA" appears first

and "PopB" appears last. The second advantage is that if cellular processes are defined by many different measurements, they all can be overlaid since they have the same cumulative percent *x*-axis (*see* Subheading 5.1). Since cumulative percent has many of the same characteristics as time, such as being nondecreasing, it can act as a surrogate for time and therefore represent cellular kinetic progressions defined by many measurements.

Whether models are frequency-based or probability-based, they can account for overlap between populations, which is a fundamental improvement over gating. Because models use nonsubjective objective functions to optimize their "parameters," they also tend to be able to automate complex cytometric analyses.

3 The Anatomy of Probability State Models

3.1 Cell Types

A model cell type roughly corresponds to a particular population of cells. Bone marrow $CD19^+$ B cells and $CD3^+CD4^+$ T cells are examples of different cell types. When defining cell types, one or more measurements are chosen to select for specific events. These selection measurements normally have relatively constant levels of measurement intensities. The selection can be based on positive or negative measurement intensity levels. T cells, for example, are positively selected with $CD3^+$, but negatively selected with $CD33^-$. Model cell types not only define selection measurements but also the order in which they are applied to the data. Cell types usually have unique signatures in high-dimensional space and appear well separated in high-dimensional Cen-se' maps (Subheading 5.2 and refs. 5–9).

Probability state modeling works out the mixture of cell types in a sample using probabilistic algorithms [8]. Some cell types such as B cells, T cells, NK cells, and Monocytes modulate key gene products as they differentiate [13, 14]. The appearance and disappearance of these gene products can stratify events into specific stages. For example as B cells develop in the bone marrow, proteins such as CD34, CD10, CD20, CD38, and CD45 modulate as they differentiate and form at least four stages [12]. A GemStone™ (Verity Software House, Topsham, ME) model can contain any number of cell types but typically the number of cell types ranges between 2 and 10.

3.2 Expression Profiles

An expression profile describes the general type of modulation that a measurement has in a cell type. It is important not to use cytometry slang that refers to measurements as parameters since the word, "parameter," has a very specific meaning in the context of modeling and statistics. Expression profiles are composed of a set of control definition points (CDPs) that describe the important

modulations of specific measurements. The CDPs also define the degree of heterogeneity in data as a function of a progression.

3.2.1 Constant

Constant expression profiles are typically used for cell type selections. The modeling notation used for these expression profiles mimics standard gating notations. For example, $CD3^{-}$, CD3d, $CD3^{+}$, and $CD3^{++}$ indicate the relative intensity of CD3 measurements (negative, dim, positive, bright). CDPs can be adjusted to combine different intensities. For example $CD3^{-}$, *d* indicates that the expression profile is designed to encompass both the negative and dim intensities. Inverse expression profiles are also available to the modeler for these types of expression profiles. For example, since live/dead measurements, L/D, usually have discrete peaks for the live cells, but complex set of peaks for dead cells; dead cells can be selected by using the inverse of the low L/D measurement, Inv (L/D^{-}).

3.2.2 Variable

Variable expression profiles describe how measurements change with progressions. In many cases, a measurement of a gene product can downregulate or upregulate, which is represented with step-down or step-up expression profiles. In other cases, expression profiles may be more complex and have three or more levels of expression. The job of a modeler is to choose the simplest set of expression profiles that are consistent with both the data and the literature. The notation used in this chapter to denote this type of modulation is to separate the intensities with the symbol, ">." For example, $CD10^{++} > CD10^{+} > CD10^{-}$ indicates that the CD10 expression profile has three decreasing intensity levels. The first level is bright for CD10, the second level decreases to a positive state, and the third level is negative or very low intensity. In some cases, progressions can branch. The notation for a branch will look similar to the following: $CD57^{-}$ > Br(+,−). This notation indicates that CD57 is initially negative and then some fraction of events upregulate to a positive state while others will remain negative.

3.2.3 Stages/Zones

If there are multiple measurements that have variable expression during a progression, many times it is advantageous to use the points of modulation to describe boundaries and labels for specific stages of progressions. For example, staging B cells in bone marrow [12] may be represented as, (B1) $CD34^{+}CD45^{+}CD10^{+} >$ (B2) $CD34^{-}CD45^{+}CD10^{+} >$ (B3)$CD34^{-}CD45^{++}CD10^{+} >$ (B4) $CD34^{-}CD45^{++}CD10^{-}$. The identification of these stages can be error prone with conventional dot-plot displays since phenotypic transitions are many times difficult to visualize. Modeling and HD Cen-se' maps have made it much easier to visualize these stages and often show that some published stages do not exist and other stages are present but have yet to be recognized as valid stages. When faced with a situation where the data is not consistent with the

literature, it is incumbent on the modeler to try to come up with a reasonable compromise where model stages are enumerated by the model as closely as possible to published stages.

3.2.4 Line-Spreads

An important aspect of expression profiles is the degree of heterogeneity in the data at each point along the expression profile. This heterogeneity is quantified as line-spreads or standard deviations (sds) associated with specific CDPs. Graphically, line-spreads are shown as 95% confidence limits (*see* Fig. 1c), which represent + and – 1.96 sds from measurement means. Line-spreads represent both biologic heterogeneity and measurement uncertainty.

3.2.5 Matching/Un-matching Expression Profiles

Each expression profile can be in either an "un-matched" or "matched" state. If it is un-matched, then the expression profile is not involved in the classification process. When building models, it is common to turn on and off the match status option to better understand how to approach modeling tasks.

3.3 TriCOMs

In some cases, it is important to perform a combinatory analysis of subpopulations within one or more stages within a specific cell type. Expression profiles naturally divide measurements into three partitions: 0: within some defined confidence limit (CL, e.g., 95%), 1: lower than CL, and 2: greater than CL. By applying these three states to each involved expression profile, specific combinations can form identifying trinary numbers. For example, the trinary number, 021, for CXCR3, CXCR5, and CCR6, represents events that are within the defined CL for CXCR3, greater than the CL for CXCR5, and less than the CL for CCR6. Expression profiles used in TriCOMs are set to not participate in the classification process. Cell types can contain any number of TriCOMs.

3.4 Model Documents

All cell types, reports, and database entries are stored as a model document. There are two important types of model documents. A template model generally contains everything in a model except for model data. Also, in the process of saving a template model, the program will automatically un-match or inactivate all expression profiles. This document format is designed to be loaded into GemStone™ and then processed by Auto Analysis. The other type of model document has all the modeled data and reflects the state of the system after modeling. Typically, batch systems are programmed to automatically save the model document in this format after an analysis has been performed.

4 HD t-SNE or Cen-se' Mapping

In 2002 at the Neural Information Processing Systems Conference, Geoffrey Hinton and Sam Roweis presented a novel set of

algorithms, called "SNE" that attempted to place events or objects defined in high-dimensional space into low-dimensional space that preserve much of their "neighborhood identity" [15]. Since then, there have been many published variants of their method where currently the most popular is t-SNE.

The t-SNE algorithms were first described in 2008 by Laurens van der Maaten [5] and later made more efficient with the Barnes Hut approximation [6, 7]. The algorithms use a *t*-distribution with one degree of freedom, and are *s*tochastic, involving *n*earest neighbor events that *e*mbed or map high-dimensional data to low-dimensions. It is a mapping process that attempts to preserve local and regional high-dimensional data relationships in low-dimensional space.

The original algorithms have been heavily modified and improved in GemStone™ to optimize resolution of mapped data, minimize random placement of populations, run at reasonably high speeds, and require far less memory. Because these changes result in higher resolution t-SNE maps than originally published, they will be referred to as "high-definition" or HD t-SNE maps or Cen-se'™ maps. Although a detailed description of how and why this algorithm works as well as it does for high-dimensional cytometry data it is outside the scope of this chapter, the major involved algorithms for SNE mapping are described below.

4.1 Normalization and Nearest Neighbor Distances

The SNE process begins with data normalization. This process attempts to eliminate any scaling differences among involved measurements while preserving their variance differences. Figure 2a shows a simple nine-event example. The events or points are defined in high dimensions and form three clusters of three points where two clusters are closer together than the other pairs of clusters. A high-speed vantage point tree [16] then finds nearest neighbor events for each event. The number of nearest neighbor events is an important free parameter of the system. Figure 2b depicts this neighborhood about event 0 with an ellipse encompassing six neighboring events. In the original t-SNE implementation, this value was determined from the parameter, perplexity, by multiplying it by three.

4.2 Similarities and Symmetrizing

Euclidean distances between events and their nearest neighbors are then converted to probabilities by employing a suitable probability density function (*see* Fig. 2c). These probabilities determine the "similarity" of the neighboring events to each event. Note that distant events outside the neighborhood end up having similarities of zero which is why the algorithm does not preserve distant data relationships. A symmetrizing algorithm then ensures that these similarities are symmetric for all participating events. The events and their nearest neighbor similarities are encoded in the matrix $\mathbf{p}_{i,j}$ (*see* lower-left corner of Fig. 2c). In the original implementation,

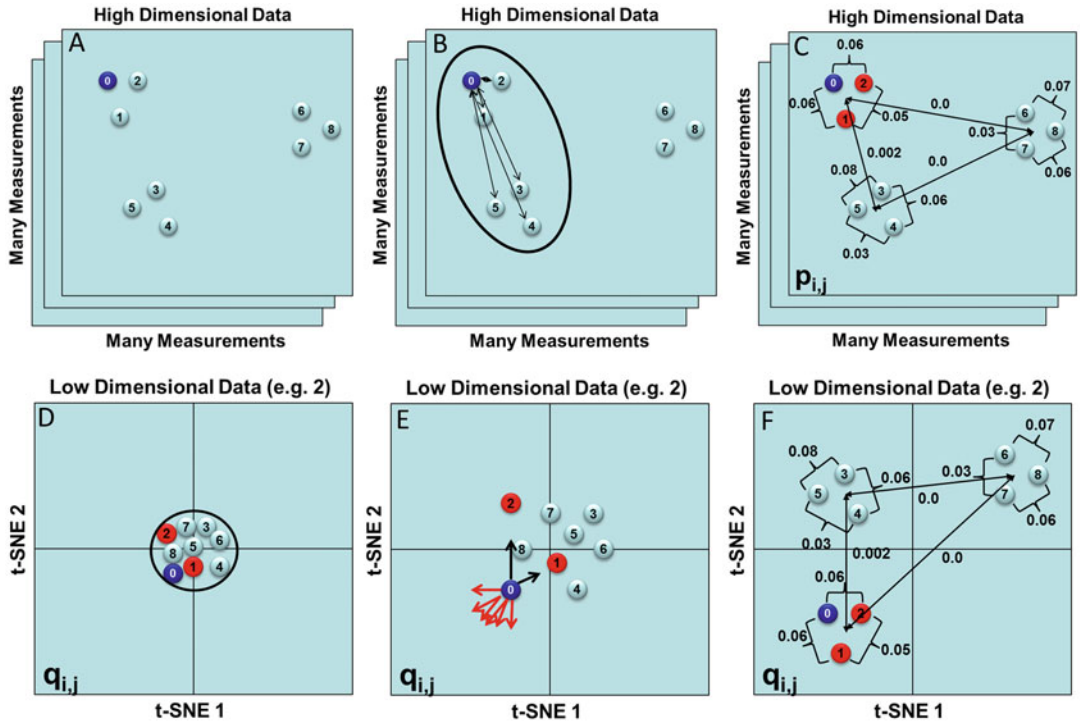

Fig. 2 HD t-SNE mapping (Cen-se'). (**a**) It schematizes a nine-point example of some high-dimensional data. (**b**) It shows a six-point neighborhood around event 0. The other events have similar nearest neighborhoods. The distance between each neighboring event is then converted to similarity probabilities by the use of a suitable probability density function. (**c**) These similarity probabilities define local structures (e.g., the three neighboring events) and regional structures (the two closest pairs of clusters) at the expense of more distant clusters. (**d**) Corresponding points are randomly placed in a small spot in t-SNE space. The distances are converted to similarity probabilities, q_{ij}, using the Cauchy probability distribution. (**e**) Positive and negative forces are computed on each event in t-SNE space. (**f**) As the events expand, the positive forces bring similar events together forming structures that preserve the local and regional structures defined in high-dimensional space

the corresponding mapped points are randomly placed in a very small spot in t-SNE space (*see* Fig. 2d). This random placement normally results in t-SNE maps with random population locations. GemStone™ Cen-se'™ mapping employs a more structured initialization that mitigates much of this randomness.

4.3 Force-Directed Solution

The t-SNE mapped points are converted to a similarity matrix much like the one described above for the high-dimensional points. A Cauchy distance distribution [17], which is equivalent to the Student's *t*-distribution with one degree of freedom, is used to convert the low-dimensional distances to probabilities, $\mathbf{q_{ij}}$. All the preceding steps have been to prepare the data for the final force-directed step in the mapping process. At this point, the system has a

high-dimensional nearest neighbor matrix of event similarities, $\mathbf{p_{i,j}}$, and a corresponding matrix of low-dimension t-SNE event similarities, $\mathbf{q_{i,j}}$. The purpose of this last stage is to move the t-SNE points in such a way that the difference between the two sets of probabilities is minimized. The Kullback-Leibler Divergence formula [18] quantifies differences between the two sets of probabilities with a single unit less number, D_{KL}.

$$D_{\mathrm{KL}} = \sum_i \sum p_{i,j} \cdot \ln\left(\frac{p_{i,j}}{q_{i,j}}\right), \quad i,j \in 1 \ldots n.$$

Lower D_{KL} values are associated with sets of probabilities that are more similar. Note that as $\mathbf{q_{i,j}}$ approaches $\mathbf{p_{i,j}}$, the ratio approaches unity. Since the log of unity is zero, D_{KL}, approaches zero as the two probabilities approach each other. Also note that distant pairs of points where $\mathbf{p_{i,j}}$ values are near zero will have little contribution to the D_{KL} value, thus making the method resistant to the effects of outlier populations.

By appropriately differentiating the D_{KL} function, gradient equations provide a means for calculating positive and negative forces on each event (*see* Fig.2e, red are the negative forces and black arrows are the positive forces on event 0). The positive and negative force equations are given as.

$$\mathrm{pos}F_{i,d} = \sum_{N_j} p_{i,j} q_{i,j} Z\left(\Upsilon_{i,d} - \Upsilon_{j,d}\right),\ i \in 1 \ldots n, j \in N_i,$$

$$\mathrm{neg}F_{i,d} = \sum_j {q_{i,j}}^2 Z\left(\Upsilon_{i,d} - \Upsilon_{j,d}\right), i,j \in 1 \ldots n,$$

where,

$$q_{i,j} = \frac{\left(1 + \sum_d \left(\Upsilon_{i,d} - \Upsilon_{j,d}\right)^2\right)^{-1}}{Z},$$

$$Z = \sum_k \sum_{l \neq k} \left(1 + \sum_d \left(\Upsilon_{k,d} - \Upsilon_{l,d}\right)^2\right)^{-1}.$$

Note that as $\mathbf{q_{i,j}}$ approaches $\mathbf{p_{i,j}}$ the positive and negative forces will cancel out for all event pairs within the *i*th event's neighborhood, N_i. What is fascinating about this system is that the negative forces are computed between all events using the Barnes Hut approximation [7, 19], not just the neighborhood events. As a consequence, the events will have a net repelling force that continually expands the collection of points on the t-SNE map. As the points expand, the positive or attractive forces between neighborhood events form domains of similar events. The patterns eventually stabilize as the two forces cancel and the result is a dot plot pattern that closely represents local and regional structures defined in high-dimensional space (*see* Fig. 2f).

4.4 Very High-Definition t-SNE Maps

For most applications, the ability of the HD t-SNE map to distinguish populations defined in high dimensions is more than adequate. However, if maximum resolution is desired, there are a few additional strategies that can be employed. The first strategy is to leverage the program's enriching option to only examine a single cell type. By not including all cell types, the t-SNE algorithms will optimally separate all the subpopulations within the enriched cell type. Also, the enriching process ensures that every selected cell type event in the file is available for the mapping process. Another important strategy is to not include measurements in the t-SNE that are not relevant for the cell type. Care should be taken with this strategy since the mapping process normally leverages noncanonical measurements for separating subpopulations. When defining staging expression profiles within a cell type, a useful strategy is to create a HD t-SNE map that involves only those measurements that are known to modulate at stage boundaries.

5 Synergistic Relationships Between Modeling and HD t-SNE Maps

A good way of initially evaluating the relevancy of HD t-SNE maps is to generate relatively simple data of known characteristics with probability state models and then render the associated t-SNE maps. A detailed comparison between the two is often illuminating on how best to leverage t-SNE maps for designing models.

5.1 Two Cell Type Model

Figure 3a shows an expression overlay for the first example cell type, CT1, and Fig. 3b shows an expression profile overlay for the second cell type, CT2. Both the cell types are defined with six hypothetical measurements: M1–M6.

The first cell type, CT1, is selected with $M1^{+}$ $M2^{+}$ $M3^{-}$. M4 downregulates ($M4^{+} > M4^{-}$) and M5 slightly upregulates and then downregulates ($M5^{+} > M5^{++} > M5^{-}$). M6 is initially negative and then some of the events upregulate while others do not ($M6^{-} >$ Br ($M6^{-}$,$M6^{+}$)). CT1 has three stages defined by M4 downregulating and M5 downregulating later.

The second cell type, CT2, is selected with $M1^{+}M2^{-}M3^{+}M6^{-}$. The progression begins with M4 downregulating, $M4^{+} > M4^{-}$, and ends with M5 downregulating, $M5^{+} > M5^{-}$. CT2 also has three stages defined by M4 downregulating and then M5 downregulating. Note that all stages have transitions with intermediate intensity events except for the first to second stage in CT1 where it has little or no transitional events.

5.2 Associated HD t-SNE Maps

Figure 3c shows the associated t-SNE map with dots colored by the associated stages and appropriately labeled from the model. Since CT1 and CT2 have different selection phenotypes, they appear well separated in the HD t-SNE map. Because of this separation t-SNE

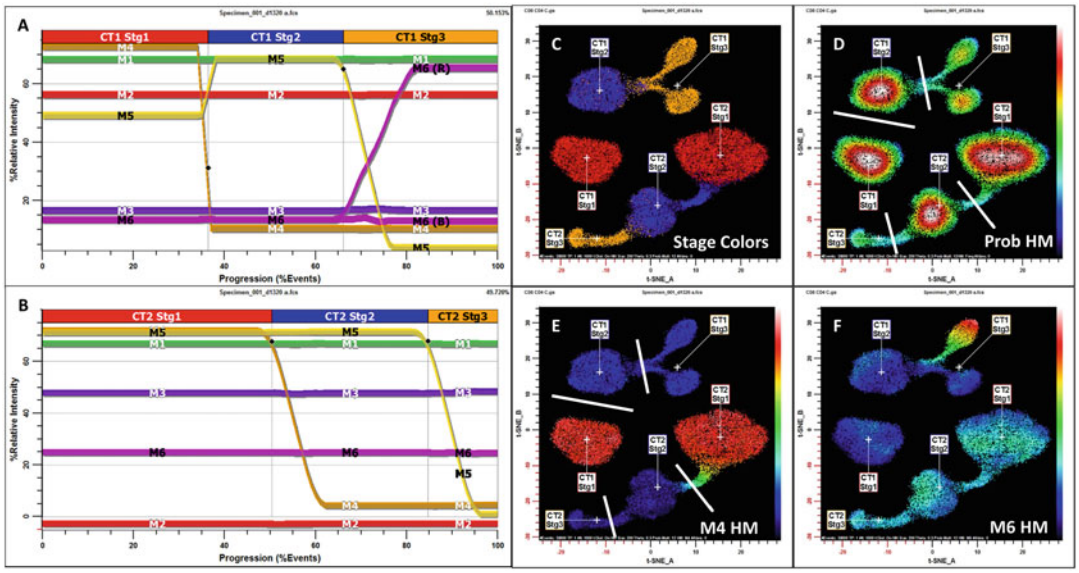

Fig. 3 Modeling and HD t-SNE. (**a**) It shows an expression overlay for the first example cell type, CT1. (**b**) It shows an expression profile overlay for the second cell type, CT2. Both cell types are defined with six hypothetical measurements: M1–M6. (**c**) It shows a HD t-SNE map of simulated data derived from the model. The *dot colors* are the same as the stage colors shown in (**a**) and (**b**). (**d**) It shows a probability heat map of the same data demonstrating that the cell types as well as the stages can be easily delineated (see *white lines*). (**e**) It shows a M4 heat map of the data demonstrating that the initial stage demarcations correspond to M4 expression levels. (**f**) It shows the variable intensity levels of the branched marker, *M6*

maps can help modelers create selection criteria that optimally define cell types. The transition from CT1 stage 1–2 is abrupt and as a consequence, a clear gap is present between these stages; whereas, the other stage transitions have a thin isthmus of transitional events connecting them. Not shown in this example is the tendency for stages to be separated by gaps with high-dimensional data. Examples of this characteristic will be demonstrated with the CD8 cell type (*see* Subheading 8).

Because of the nature of t-SNE mapping, progressions like the ones shown here will normally have sequential stages next to each other and nonsequential stages distal to each other. One exception to this rule occurs if the t-SNE map is inappropriately configured to only look at local structure. In the original implementation, this exception frequently occurred when the entropy and/or perplexity parameters were set too low (e.g., perplexity <= 50). Another exception to the adjacency rule occurs if a stage has no intermediate events as with CT1, stage 1. This stage can act like a different cell type and can be positioned anywhere since distant structures are not preserved in the mapping. Fortunately, this problem is very rare when examining high-dimensional data since there are usually intermediate events in other dimensions. Also, redoing the map

with random event starting positions is a valuable technique that can rule in or out this particular problem.

A good way of approaching the proper staging of a cell type is initially to draw lines through the gaps and valleys of the map without being biased by any additional information (*see* Fig. 3d for probability heat map). As demonstrated in this panel, the probability heat map is a good graphic option to use to better visualize peaks and valleys. Spot regions before and after the lines or heat maps will reveal which markers are modulating (*see* Fig. 3e for M4 heat map). Heat maps generally use heat-related colors to denote measurement intensity or event frequency/probability. A common sequence of colors is black- > blue (cold)- > Cyan- > Green- > Yellow > Red (hot) > White. Although heat maps are quite convenient, they can be tedious to interpret and they are qualitative in nature. The best way of examining marker transitions is with animated regions. Animated regions blink dots in all other graphics with a specified color.

Look at other samples to make sure the boundaries and marker transitions are consistent. If there are multiple measurements changing at the same time, choose the one where the transition is greatest with the least degree of variability. In order to assign a direction to the stages, it is necessary to know at least one part of a progression's directional information. For example, to know that CT1's progression moves left to right and CT2's moves right to left, the stage that contains high-intensity M4 needs to be known.

One major advantage of HD t-SNE maps is that branched progressions can be directly inferred from the dot patterns. A close examination of the CT1 stage 3 heat map of M6 in Fig. 3f shows transitional events forming a fork in the progression where some events upregulate M6 while others do not. This ability to detect branched progressions is difficult if not impossible using conventional dot-plot displays and is important in properly designing models.

If measurement expression profiles are branched as shown for M6, then they will not work well to identify stages of progression. One of the reasons for this is that other marker transitions will appear much more complicated than they really are. For example, if M6 were not known to be branched and were modeled as a non-branched expression profile, then M5 would inappropriately appear branched. This problem progressively worsens with higher numbers of correlated measurements. In the gating world, using branched measurements to define stages can result in subsets with overly complicated phenotypes.

Once the appropriate measurements are chosen for staging, they can be introduced into the model as step-up, step-down, or multiple level expression profiles. After ensuring that all the validation data are consistent with the model stages, the literature can be consulted to determine how best to label the stages. If a model

stage and the literature are consistent with each other, then use literature-determined label for the stage; otherwise, use simple labels such as "early," "int," "late," or just a sequence such as "B1," "B2," etc. Model and literature determined stages may be very different. If they are different, it is always possible to output data that is equivalent to literature-determined stage phenotypes using the TriCOM combinatory system. The mistake a modeler can make is to force literature-determined stages onto data that are not consistent with those stages.

5.3 Error Analysis

All analysis methods have some degree of error. It is important to attempt to evaluate this error for all designed models. There are a few strategies that can be employed to perform error analyses on your models. All these strategies require that you have some estimates that represent "truth."

5.3.1 Synthesized Data

Since models can synthesize data, it is possible to produce or synthesize data as shown in this section where all information is known and therefore can act as a "truth" set. For example, the number of CT1 events synthesized from the data shown in Fig. 3 was exactly 14,906 and the number of CT2 events was 15,094. After Auto Analysis, the actual numbers of events classified as CT1 and CT2 events were 14,892 and 14,916. The error of model classification is the percentage of the number of misclassified events to the number of "truth" events. For CT1, the error is $(14{,}906–14{,}892) \times 100/14{,}906 = 0.13\%$ and for CT2 it is 1.2%. Because probability state modeling defines cell type boundaries probabilistically, there will always be some unclassified events if the probability of exclusion property is set to some positive fraction. By default, it is set to 0.001 or 0.1%. For some applications such as rare event detection, the probability of exclusion can be set to 0 to force the system to categorize every event. The same %error can then be calculated when in this mode.

5.3.2 HD t-SNE Cell Types

Since HD t-SNE maps can separate cell types based on many measurements, a truth set can be approximated by enclosing the cell type events shown in the map with polygonal boundaries (not shown). For this example, the numbers of events reported by the regions are 15,071 and 14,929 for CT1 and CT2, which are quite close to the synthesized numbers. Keep in mind that regions will enumerate all the enclosed events including the unclassified events. To find the number of classified and unclassified events enclosed by a region use Information equations. For example, if region 4 is used to encompass CT1 events, the number of events classified as CT1 is given by GSUM(C1,R4) and the number of unclassified events is given by GSUM(C0,R4). For many applications such as the ones discussed in Subheading 8, this is a very practical method for assessing model classification error.

Another important calculation that is possible using this technique is to calculate the % specificity of a model's cell type. % Specificity is defined as the percent of cell type events that are appropriately classified. For example, the % specificity equation that returns the percent of CT1 events that are properly classified is GSUM(C1,R4) × 100/GSUM(C1).

5.3.3 Legacy Method

Another strategy for assessing the merits of a model is to compare it against some legacy method. A legacy method could be a gating method or it could be a previous version of a model. Since gating is highly inaccurate, the best that normally can be done is to perform correlations with the gated results. The degree of correlation is usually enumerated as a correlation coefficient. Keep in mind that a probabilistic model will normally capture more cell type events than gating.

5.3.4 Measurement Intensities

All the above strategies involve estimating numbers or percentages of events. Another type of error analysis is to examine the veracity of measurement intensities. For high-dimensional data sets, this kind of analysis can be quite tedious and rarely justified. If the application does require an accurate assessment of one or more measurement intensities, use medians and robust standard deviations to estimate central tendencies and variabilities. There are two largely equivalent methods for measuring intensity variability, rSD and qSD. If X is a set of values for measurement X, then the equations for rSD and qSD are

$$\begin{aligned}
&\text{median} = \text{Median}(X),\\
&\text{rSD} = \text{Median}(|X_i - \text{median}|) \times 1.4826,\\
&\text{IQR} = Q3(X) - Q1(X),\\
&q\text{SD} = \frac{\text{IQR}}{1.34898}.
\end{aligned}$$

6 Initial Preparations and General Considerations

6.1 Computer System

Before beginning to model, make sure the computer and display systems work well with the computationally intensive demands of high-dimensional modeling. Keep in mind that performance will be directly proportional to the computer's clock speed, the amount of RAM, and number of core processors. Since this type of analysis is especially graphics intensive, consider using two or more high-resolution color monitors.

6.2 Organize Folders and Files

Organizing your work before beginning a modeling project has a lot of personal preference elements, but there are some basics that should be followed. Create folders for analysis, batch system, files, images, models, and templates for your modeling project. Most of

these are self-explanatory; however, the analysis folder should use something like PowerPoint to document your progress as you develop models. As the template model develops, save it to the templates folder identified with different version numbers. It is very important to be able to go back to specific versions of the model as it is developed with different strategies.

As mentioned earlier, a template model differs from a regular model in that it does not contain any data and all the expression profiles are unmatched in order that Auto Analysis can enable them in a specific order. If you intend on eventually moving all the folders to different drives, consider using a virtual drive for your project folders. Since the paths for files, models, images, and other key documents contain the drive letter, you can save time moving the system if you use a virtual drive letter for the drive containing you project.

6.3 Test Feasibility and Create Test Data Sets

Gather an adequate number of validation files. Quickly determine the feasibility of a modeling project with a few example files. After feasibility is assured, then use 50 to 100 sample files for initial testing purposes. Many times, it is necessary for these files to be carefully analyzed by manual gating with key results databased. The databased results can later be compared with the automated modeling results. Another approach to validation is to create a set of contextual plots that can be inspected after modeling to determine if event classifications are reasonable. For fluorescence cytometers, all files need to be properly compensated.

6.4 Optimize Measurement Transforms

Each model measurement must have a valid transformation. Transformations mathematically convert raw linear measurements to scales that appropriately stabilize population variabilities [20]. Both t-SNE maps and modeling results are affected by transforms. Inappropriate transforms can render measurements useless or even worse, deceiving; therefore, it is very important to take time to evaluate each measurement's transformation using either dot-plot or histogram displays. Once a measurement transform has been optimized, rarely is it desirable to have the software modify the transform's parameters with subsequent sample files.

6.5 Start Simply and Document as You Go

There is a strong tendency to want to build the complete model in one step. The problem with this approach is that if something fails, it is hard to determine reasons the model failed. A better approach is to inspect models against validation files at key steps in their evolution. Once a cell type is designed, do not go to the next step in model building until it is tested for all validation files. It is important to remember that since models support automation, it is relatively easy to make small changes to them and retest against validation sample files. Use the batch system to automatically analyze files with specific model versions and generate useful sets of contextual images and databased results.

6.6 Organize a Study with the Batch System

The batch system contains not only pathways to FCS data files but also sets of actions to apply to these files. It also contains a complete database that is normally hidden from view. Each level of the batch can have a different set of actions assigned to it. Usually, the first level is left in its default state, which is to read a FCS file into the system. Higher batch levels are used to open a model template, read the associated FCS files, Auto Analyze the data with the template, save the analyzed model, and also save images. There is a rich set of other actions that can also be assigned to batch levels. For most applications, after the template model is loaded into the system, another batch level simply resets the model back to its previous loaded state and then does all the same analyses as the preceding batch level. This technique of resetting the model before subsequent analyses is important for advanced analysis techniques such as creating averaged models.

7 General Approach to Modeling

Fig. 4 shows the general iterative process for developing probability state models.

7.1 Background Investigation

Before beginning any modeling project, relevant literature should be obtained and studied. Also, it is important to have on hand any relevant manual gating strategies. Keep an open mind during the modeling process. Because cell type stages, branches, and subsets are difficult to visualize with traditional dot plots, literature descriptions may not be completely consistent with data analysis results.

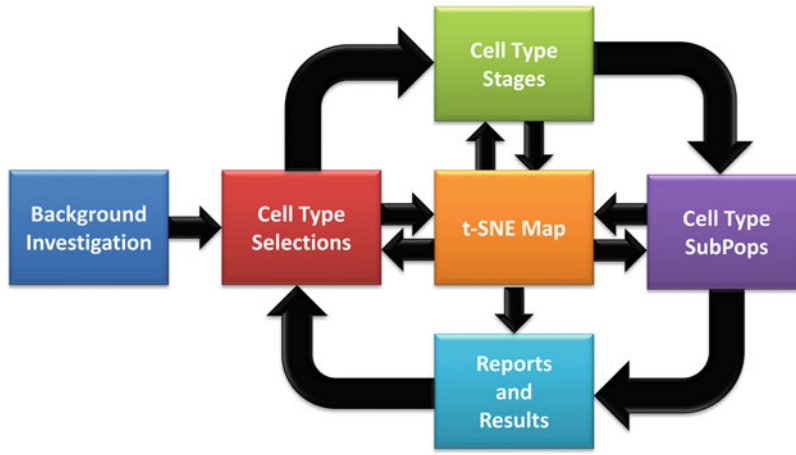

Fig. 4 General approach to modeling. This figure shows the general iterative process for developing probability state models

7.2 HD t-SNE Maps

An early step in modeling is creating high-definition t-SNE maps of all relevant measurements for a typical sample file. Using one representative file at the start makes it easier to see branched measurements and stage separations. Once you have t-SNE maps defined for a typical sample file, it can be useful to sample a number of files and create t-SNE maps of composite files. Keep in mind that composite files will inevitably have less resolution than single representative files.

7.3 Cell Type Selections

Usually, there are a number of equivalent choices for the first cell type to model. Pick the easiest and most prevalent cell type first. Since cell types compete probabilistically for events, they often serve to "clean up" events for subsequently defined cell types. Add negative selection measurements to the cell type first and then, add the best positive selection measurements. Start with canonical antibody-based measurements and then clean up selections with measurements like side-scatter, if available.

Measurements that modulate within the cell type can be added at this point, but do not create expression profiles for them yet. Populate the cell type reports with graphs that provide contextual information on the validity of cell type selections. Use HD t-SNE maps to make sure cell type selections do not include other cell types. Examine unclassified events in t-SNE maps to ensure that the model is not excluding valid cell type events.

An often overlooked modeling step is to produce a set of graphs that assess the quality of the data. On at least one cell type include a plot of time or chronology versus some positive intensity measurement to check for acquisition drifting. For fluorescence cytometers, ninety-degree light-scatter is a good choice since it is highly influenced when bubbles are present in the nozzle. A selected set of median intensity measurements is also important to report since it may indicate a problem with antibody staining or modeling. For fluorescence cytometers, always include a peak versus integral or equivalent side-scatter dot plots since the presence of aggregates is another kind of quality assessment for a sample.

Continue adding cell types in this manner until all the desired cell types are modeled with selection expression profiles. Do not begin modeling the cell type stages until all cell types have been added to the model and events have been appropriately selected for each one. Validate the selection of cell types by analyzing all test files and suitably examining their contextual plots and perhaps comparing the results with some predicate method. Usually, additional refinements of selection expression profiles are necessary after evaluating a collection of files. Always put the complete phenotypic description of the selection on the cell type report.

7.4 Cell Type Stages

This step can be skipped if the cell type does not have measurements that modulate with progression or the cell type does not have a progression. For complex cell types such as peripheral blood CD8 T

cells, the t-SNE system should be put into very high-resolution mode where the t-SNE map is restricted to a single cell type. As mentioned in Subheading 5, try first to develop the stages in an unbiased manner. Test the model stages on all the validation files. For large panels, it is advantageous to examine t-SNE maps derived from just those measurements that are involved or could be involved in staging.

Begin staging with those measurements that have simple up- or downregulating patterns that also have low line-spreads that clearly separate the different levels of measurement intensity. Model simple expression profiles first and then progress to more complex multi-level expression profiles. Use small animation regions before and after the stages or heat maps to validate stage boundaries. Choose the appropriate beginning and ending CDPs for defining stages and label them as close to the literature as possible.

If published stages are not consistent with the data, use TriCOMs to enumerate them. Create a table on the cell type report that enumerates the stage percentages and also describes them phenotypically. Archive cell types as individual files in order that future models can quickly be created by assembling appropriate cell type files.

7.5 Cell Type Subsets

In some cell type stages, there may be cellular subpopulations or subsets that have varied measurement phenotypes. Use TriCOM graphs to enumerate the phenotypically defined subpopulations. Try to use labels that are consistent with the literature and always show the phenotypic signature for each subpopulation. As mentioned above, the TriCOM system can also be used to enumerate published stages that are not consistent with model-defined stages.

7.6 Reports, Database Entries, and Alerts

The final edits to models should be to create clear and relevant graphics on generated reports. This last stage of model building is also a convenient time to select which analysis results to database. Consider storing event counts instead of percentages since they can always be used later to form percentages in programs like Excel. Finally, decide which results are capable of detecting either failures in data integrity or modeling. These databased results can be turned into alerts that inform users when something went wrong.

8 Specific Example

The following example describes model development that enumerates and displays most known populations in human peripheral blood mononuclear preparations, PBMCs. Although all the steps described in Subheading 7 were followed, this Subheading describes only a sampling of noteworthy modeling decisions for some cell types. The panel was developed by the Human Immune

Monitoring Center at Stanford University and was composed of the following antibodies and rare-earth isotope labels: (1) CD45RA Dy162, (2) CD20 Dy164, (3) CD33 Er166, (4) CD28 Er167, (5) CD24 Er168, (6) CD161 Er170, (7) CD38 Eu151, (8) CD11b Eu153, (9) CCR6 Gd155, (10) CD94 Gd156, (11) CD86 Gd157, (12) CXCR5 Gd158, (13) CCR7 Gd160, (14) CD127 Ho165, (15) CD57 In113, (16) Live/Dead In115, (17) DNA1, (18) DNA2, (19) HLA-DR Lu175, (20) CD19 Nd142, (21) CD4 Nd143, (22) CD8 Nd144, (23) IgD Nd146, (24) CD11c Nd148, (25) CD3 Nd150, (26) CD85j Sm147, (27) CD16 Sm149, (28) CD27 Sm152, (29) CD14 Sm154, (30) CXCR3 Tb159, (31) ICOS Tm169, (32) TCRgd Yb171, (33) PD-1 Yb172, (34) CD123 Yb173, (35) CD56 Yb174, and (36) CD25 Yb176. The data shown in this section were obtained from a Helios™ mass cytometer (Fluidigm Corporation).

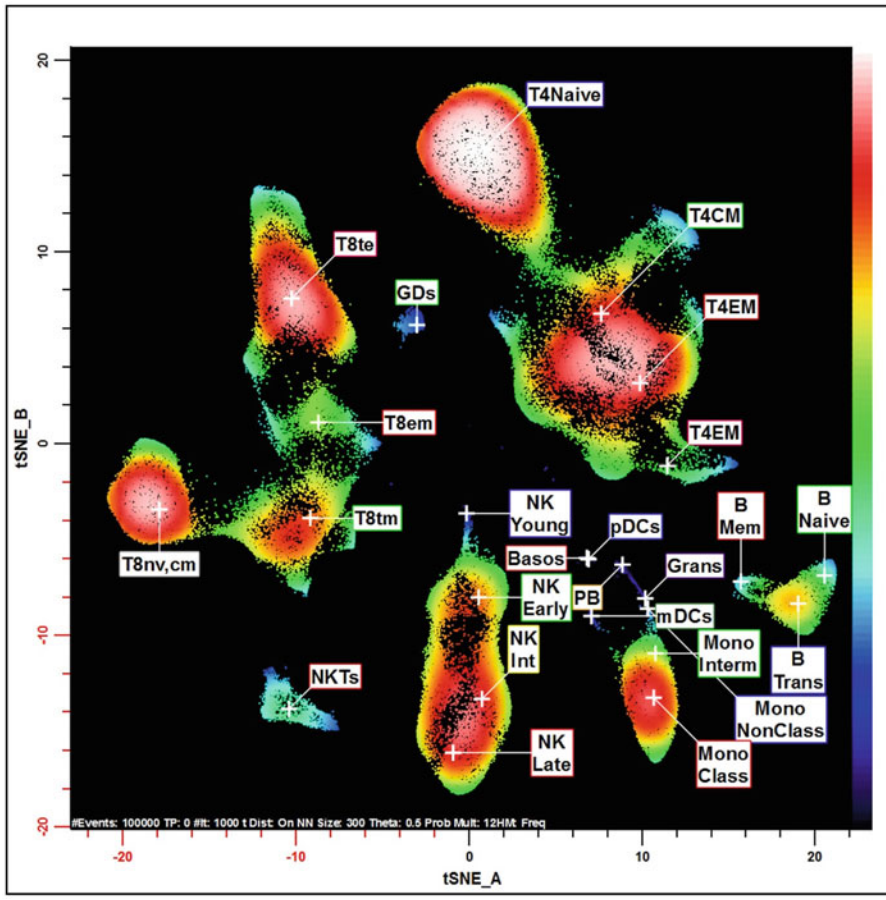

Fig. 5 PBMC HD t-SNE overview map. This figure shows a HD t-SNE map of a PBMC sample involving 36 measurements displayed as a probability heat map. All the major cell types are well separated. Many of the cell type stages are also labeled

Figure 5 shows an overview t-SNE map involving all the measurements listed above. The map separates a PBMC sample into the major cell types: CD8 T, CD4 T, NKTs, γδ (GDs), NK, Monocytes (Mono), Grans, and B cells. Many of the cell type stages are also labeled. Minor populations such as plasmacytoid dendritic cells (pDCs), myeloid dendritic cells (mDCs), and plasmablasts (PB) are not well shown with this view because their frequency is so low. Event colors are drawn from a probability heat map that is designed to allow easy inspection of cell type frequency topographies.

8.1 CD8 T Cells

Modeling CD8 T cells exemplifies the importance of validating selection and staging against high-definition t-SNE maps.

8.1.1 Selection

CD8 T cells were initially selected as $CD33^{-}CD14^{-}CD3^{+}CD8^{+}CD4^{-}$ [4, 21, 22]. A good technique for assessing cell type specificity is to only look at events classified for a specific cell type and evaluate how contained these events are within the t-SNE cell type boundary. When CD8 T cells were selected and evaluated on the HD t-SNE map, some events were inappropriately positioned in the NKT and γδ (GDs) map domains (*see* Fig. 6a, white arrows). We used colored regions to explore these NKT and γδ events in a number of expression profiles and found that CD161 and TCRgd allow for better differentiation between CD8 T cells and the NKT and γδ cell types (*see* Fig. 6b, c). This inspection only takes a few minutes since you are searching for expression profiles where blue dots, the CD8 T cells, are separated from red and green dots, the NKT and γδ events respectively.

By adding the expression profiles, Inv($CD161^{+}$) and $\gamma\delta^{-}$, to the selection phenotype, these false CD8 events were greatly reduced (*see* Fig. 6d). When HD t-SNE maps unambiguously separate cell types as in this example, this kind of selection refinement can be employed to reduce model false negatives and positives.

8.1.2 Staging

Before beginning to stage the CD8 T cells, make sure that all cell types are properly selected for all validation files. Figure 7a shows a HD t-SNE map of CD8 T cells for one donor sample using all 36 measurements. Although most of the staging information can be inferred from this map, a good technique for validating staging is to redo the t-SNE map with a limited set of measurements that appear to modulate on either side of the gaps in the overview t-SNE map. Heat map and region color analyses show that the key staging measurements are CD45RA, CCR7, CD28, CD27, and CD56 (data not shown). Figure 7b shows the same sample as shown in Fig. 7a with this limited set of involved measurements. White lines separate the map's obvious similarity domains and labels indicate the domain's key phenotypes.

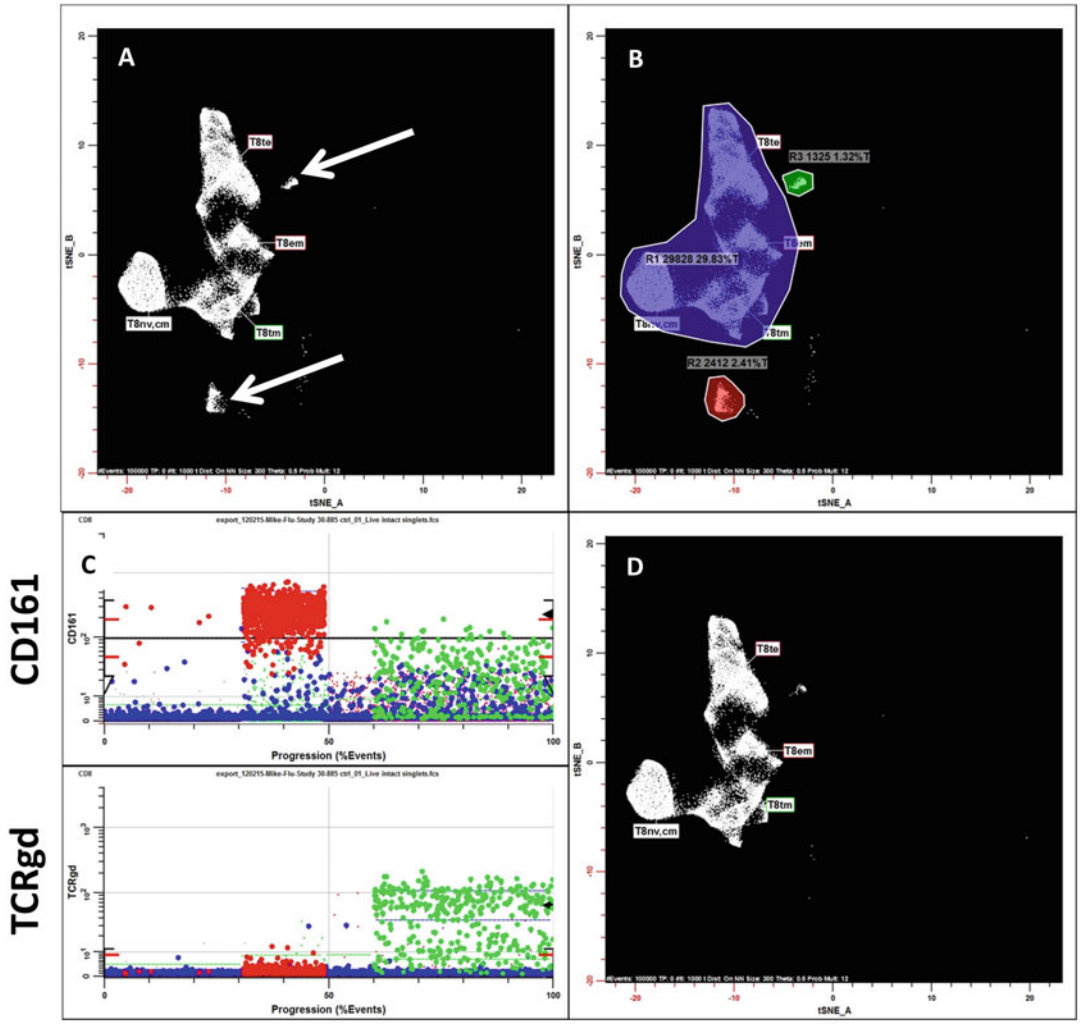

Fig. 6 Optimizing CD8 T cell type selection. (**a**) It shows CD8 T cells (*white dots*) that were selected as $CD33^{-}CD14^{-}CD3^{+}CD4^{-}$. The *white arrows* indicate areas where the CD8 T cells are inappropriately positioned within the GD and NKT domains. (**b**) It shows three different colored regions that encompass the CD8 T (*blue*), NKT (*red*), and TCRgd (*green*) t-SNE map regions. (**c**) The CD8 T cell expression profiles for CD161 and TCRgd are shown, and demonstrate that the selection phenotype can be improved by including $CD161^{-}$,d and $TCRgd^{-}$ to CD8 T cell selection phenotype. (**d**) It shows the improved specificity results by adding these additional measurements

Figure 7c–f show the associated CCR7, CD28, CD27, and CD56 heat map displays. The CD28, CD27, and CD56 measurements are clearly delineated by the gaps in the t-SNE map, whereas $CCR7^{+}$ portion of the heat map only partially fills the contained domain. The progressive nature of the heat maps confirms that they are demarcating a valid progression.

A surprising and significant finding is the gradient of CD45RA expressions (*see* Fig. 8a) through each t-SNE similarity domain.

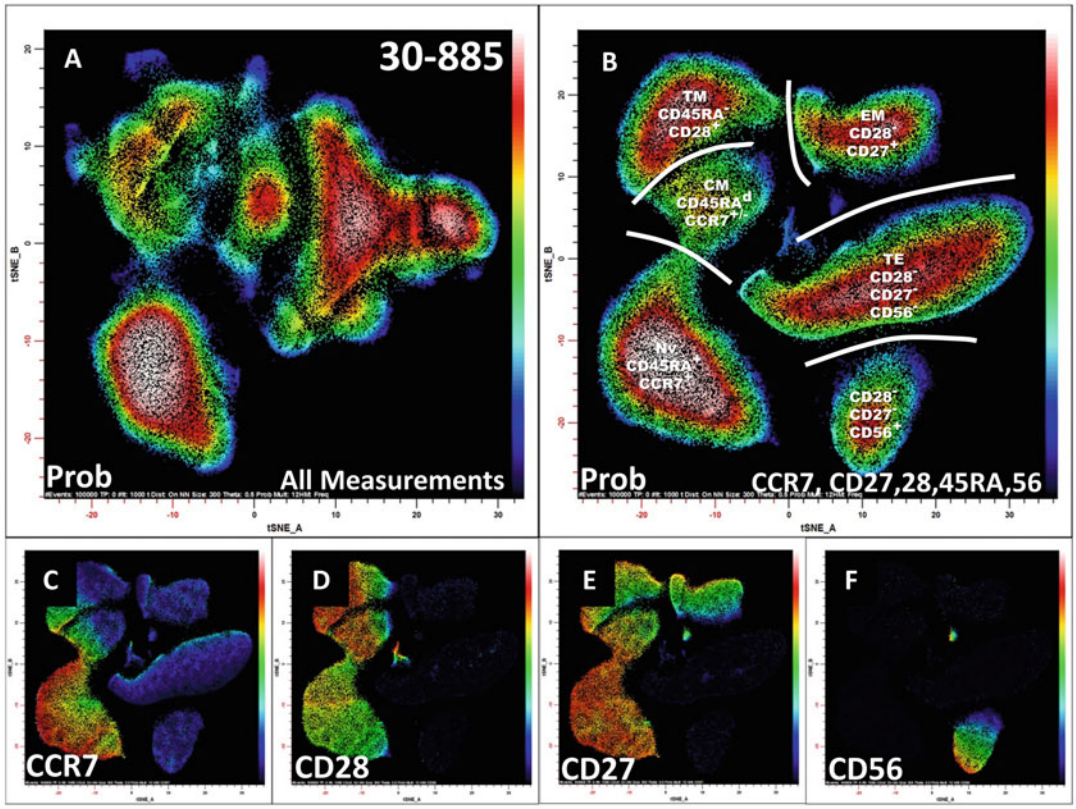

Fig. 7 Optimizing CD8 T cell staging. (**a**) It shows a HD t-SNE map of CD8 T cells for one donor sample using all 36 measurements. (**b**) It shows the same sample as shown in (**a**) with only CCR7, CD28, CD27, CD45RA, and CD56 involved in the t-SNE mapping. *White lines* separate the map's obvious similarity domains and labels indicate the domain's key phenotypes. (**c**–**f**) These show the associated CCR7, CD28, CD27, and CD56 heat map displays. The CD28, CD27, and CD56 measurements are clearly delineated by the gaps in the t-SNE map, whereas $CCR7^{+}$ portion of the heat map only partially fills the contained domain. The progressive nature of the heat maps confirms that they are demarcating a valid progression

This pattern suggests that CD45RA is complex and branched which makes it undesirable to use as a staging marker as discussed in Subheadings 2 and 5. A reasonable question to ask is what would happen if CD45RA were used as a staging marker? Most of the negative (blue and cyan) events in the TE domain would end up being re-categorized into the CM and EM stages and the positive EM domain events (red and white) would end up being in the TE stage. Because of this shuffling of events, both CD28 and CD27 would then appear branched when modeled. Also, the stages would end up crossing the t-SNE determined domains. One of the tenants of modeling is to use the simplest model that is consistent with the data. In this case, the simplest model is to not stage with CD45RA and stage with CD28 and CD27 instead. When these two markers are used as stages, the model fit is excellent (not shown) and the

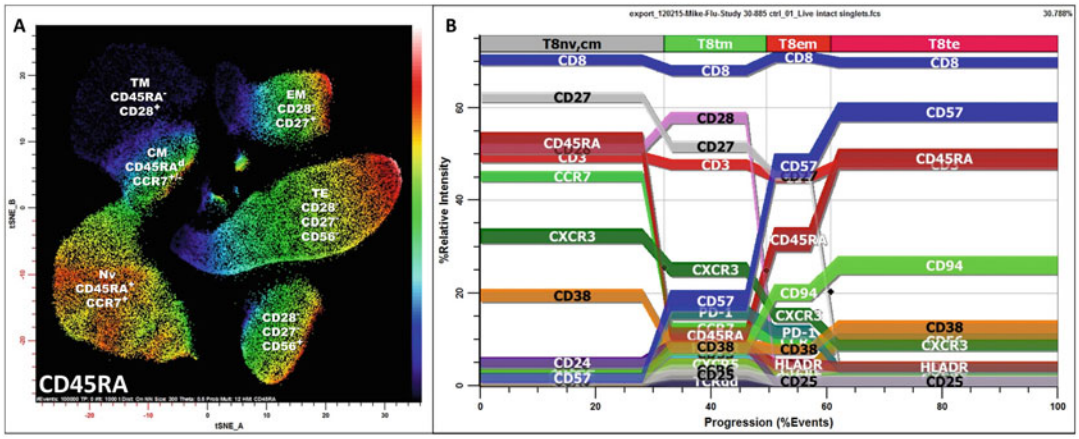

Fig. 8 Appropriate CD8 T cell type staging. (**a**) It shows a gradient of CD45RA expressions through each t-SNE similarity domain. This pattern suggests that CD45RA is complex and branched which makes it undesirable to use as a staging marker as discussed in Subheadings 2 and 5. (**b**) It shows the end result of using CCR7, CD28, and CD27 to properly stage CD8 T cells

final set of expression profiles are very similar to those presented by Appay et al. ([22]; *see* Fig. 8b).

Since the CM stage is defined as $CD45RA^{-}CCR7^{+}$ in most publications, this relatively small population can be enumerated by a TriCOM within the Nv-CM stage. The last decision that needs to be made is whether to use a step-up expression profile for CD56 or to use TriCOM to enumerate the $CD56^{+}$ population within the EM stage. If a step-up were used for CD56, then we would need to rename these stages since they would no longer be compatible with many literature-determined labels. In this case, it is best to enumerate the $CD56^{+}$ population as a TriCOM determined subpopulation.

At this point, the model needs to be tested against all validation files to make sure there was not something unusual with this single sample. Often when evaluating cell type stages and subsets as in this example, there is a discovery component to the process. A good approach to take is to be conservative in staging assessments and to use the TriCOM combinatory system as a means of enumerating these subsets. Only generalize when a reasonably large number of samples have been evaluated.

8.2 NK Cells

Traditionally, NK cells are simply $CD3^{-}CD56^{+}$ but examining the HD t-SNE in Fig. 9a shows that NK cells are far more interesting and complex than that.

8.2.1 Selection

The selection for the NKs was found to be $CD33^{-}CD14^{-}CD3^{-}CD123^{-}$Inv($CD56^{-}$).

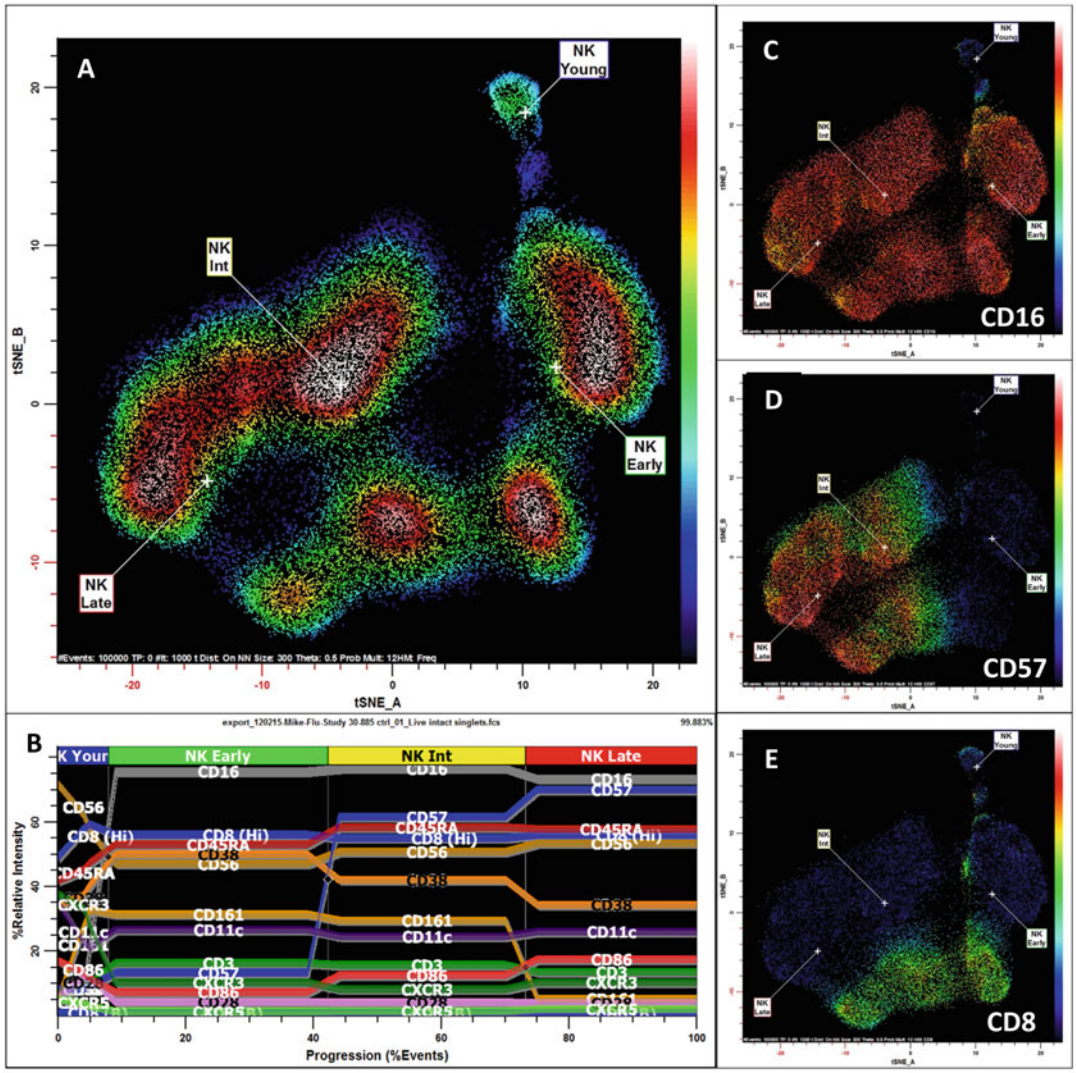

Fig. 9 Modeling NK cells. (**a**) It shows a HD t-SNE map based on all 36 measurements of NK cells. Selection of the NK cell type is $CD33^{-}CD14^{-}CD3^{-}CD123^{-}$Inv($CD56^{-}$). The population labeled as NK Young is $CD16^{-}CD56^{++}$ (*see* (**b**) for expression profile overlays and (**c**) for CD16 heat map). The next stages are (Early) $CD16^{+}CD57^{-}CD161^{+}$>(Int) $CD16^{+}CD57^{+}CD161^{+}$ and (Late) $CD16^{+}CD57^{+}CD161^{+}$ and the other stage-related marker intensities are summarized in (**b**). This sample has a biphasic expression of CD8 forming two bands in the HD t-SNE map (*see* (**a**) and (**e**))

8.2.2 Staging

The population labeled as "Young NK" is $CD16^{-}CD56^{++}$ and is thought to be immature NKs [23, 24]. Figure 9c heat map demonstrates that this population is $CD16^{-}$. CD57 then upregulates as shown by Fig. 9d. For many samples, there appears to be a branch in CD8 expression where one side of the t-SNE defined population is $CD8^{-}$ and the other $CD8^{+}$ (*see* Fig. 9e). The last stage is determined by a downregulation of CD161. The stages are therefore

defined as (Young) $CD16^{-}CD57^{-}$ > (Early) $CD16^{+}CD57^{-}CD161^{+}$ > (Int) $CD16^{+}CD57^{+}$ $CD161^{+}$ > (Late) $CD16^{+}CD57^{+}CD161^{-}$ and the other stage-related marker intensities are summarized in Fig. 9b.

9 Summary

As you no doubt noticed, modeling complex cellular populations defined by cytometry's correlated single-cell measurements is not easy. Hopefully, the framework provided will help those interested in building automated systems using modeling principles. I hope you enjoyed reading this chapter and if you ever have questions about any part, please send me emails.

Disclaimer

This author works for the company that develops and sells GemStone™. Every effort has been made to discuss general modeling concepts that would be applicable to other modeling packages if and when they become available.

References

1. Miller D, Hunsberger B, Bagwell C (2012) Automated analysis of GPI-deficient leukocyte flow cytometric data using GemStone. Cytometry B Clin Cytom 82B:319–324
2. Wong L, Hunsberger BC, Bruce Bagwell C, Davis BH (2013) Automated quantitation of fetomaternal hemorrhage by flow cytometry for HbF-containing fetal red blood cells using probability state modeling. Int J Lab Hematol 35(5):548–554. doi:10.1111/ijlh.12060
3. Wong L, Hill BL, Hunsberger BC, Bagwell CB, Curtis AD, Davis BH (2014) Automated analysis of flow cytometric data for measuring neutrophil CD64 expression using a multi-instrument compatible probability state model. Cytometry B Clin Cytom. doi:10.1002/cyto.b.21217
4. Inokuma MS, Maino VC, Bagwell CB (2013) Probability state modeling of memory CD8(+) T-cell differentiation. J Immunol Methods 397 (1-2):8–17. doi:10.1016/j.jim.2013.08.003
5. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605
6. van der Maaten L (2009) Learning a parametric embedding by preserving local structure. Paper presented at the AISTATS, TiCC, Tilburg University P.O. Box 90153, 5000 LE Tilburg, The Netherlands
7. van der Maaten L (2014) Accelerating t-SNE using tree-based algorithms. J Mach Learn Res 15:1–21
8. Bagwell CB, Hunsberger BC, Herbert DJ, Munson ME, Hill BL, Bray CM, Preffer FI (2015) Probability state modeling theory. Cytometry A. doi:10.1002/cyto.a.22687
9. Bagwell C (2001) Data analysis through modeling. Curr Protoc Cytometry Chapter 10:Unit 10.17. doi:10.1002/0471142956.cy1007s01
10. Bagwell C (2010) Probability state modeling: a new paradigm for cytometric analysis. In: Litwin V, Marder P (eds) Flow cytometry in drug discovery and development. John Wiley and Sons Inc, Hoboken, NJ, p 281
11. Bagwell CB (2011) Breaking the dimensionality barrier. Methods Mol Biol 699:31–51. doi:10.1007/978-1-61737-950-5_2
12. Bagwell C, Hill B, Wood B, Wallace P, Alrazzak M, Kelliher A, Preffer F (2015) Human B-cell and progenitor stages as determined by probability state modeling of multidimensional

cytometry data. Cytometry B Clin Cytom 88 (4):214–226
13. Loken M, Shah V, Dattilo K, Civin C (1987) Flow cytometric analysis of human bone marrow II: normal lymphocyte development. Blood 70:1316–1324
14. Loken M, Wells D (2000) Normal antigen expression in hematopoiesis. In: Stewart C, Nicholson J (eds) Immunophenotyping. Wiley-Liss Inc, Hoboken, NJ, pp 133–160
15. Hinton G (2002) Roweis S Stochastic neighbor embedding. NIPS, In, pp 833–840
16. Yianilos PN (1993) Data structures and algorithms for nearest neighbor search in general metric spaces. Proceedings of the ACM-SIAM symposium on discrete algorithms, pp 311–321
17. Meyer SL (1975) Data analysis for scientists and engineers. John Wiley & Sons, New York, NY
18. Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22(1):79–86
19. Barnes J, Hut P (1986) A hierarchical O(N log N) force-calculation algorithm. Nature 324 (4):446–449
20. Bagwell CB, Hill BL, Herbert DJ, Bray CM, Hunsberger BC (2016) Sometimes simpler is better: VLog, a general but easy-to-implement log-like transform for cytometry. Cytometry A 89(12):1097–1105. doi:10.1002/cyto.a.23017
21. Maecker HT, McCoy JP, Nussenblatt R (2012) Standardizing immunophenotyping for the Human Immunology Project. Nat Rev Immunol 12(3):191–200. doi:10.1038/nri3158
22. Appay V, van Lier RA, Sallusto F, Roederer M (2008) Phenotype and function of human T lymphocyte subsets: consensus and issues. Cytometry A 73(11):975–983. doi:10.1002/cyto.a.20643
23. Blom B, Spits H (2006) Development of human lymphoid cells. Annu Rev Immunol 24:287–320. doi:10.1146/annurev.immunol.24.021605.090612
24. Poli A, Michel T, Theresine M, Andres E, Hentges F, Zimmer J (2009) CD56bright natural killer (NK) cells: an important NK cell subset. Immunology 126(4):458–465. doi:10.1111/j.1365-2567.2008.03027.x

http://www.springer.com/978-1-4939-7344-6

Flow Cytometry Protocols
Hawley, T.S.; Hawley, R.G. (Eds.)
2018, XIV, 492 p. 107 illus., 80 illus. in color. With online files/update., Hardcover
ISBN: 978-1-4939-7344-6
A product of Humana Press