

2017-06-09

# viSNE fine-tuning enables better resolution of cell populations

---

Anna Belkina, Chris Ciccolella, Jennifer Snyder-Cappione. (2017). viSNE fine-tuning enables better resolution of cell populations, presented at CYTO 2017 International Society for Advancement of Cytometry Annual Meeting.

<https://hdl.handle.net/2144/27023>

*Downloaded from DSpace Repository, DSpace Institution's institutional repository*

## viSNE fine-tuning enables better resolution of cell populations

**Anna Belkina<sup>1</sup>, Chris Ciccolella<sup>2</sup>, Jennifer Snyder-Cappione<sup>1</sup>**

<sup>1</sup>Flow Cytometry Core Facility, Boston University School of Medicine, Boston, MA, United States, <sup>2</sup>Product Development, Scientific Applications, Cytobank, Inc, Santa Clara, CA, United States

t-Distributed Stochastic Neighbor Embedding (t-SNE or viSNE) is a dimensionality reduction algorithm that allows visualization of complex high-dimensional cytometry data as a two-dimensional distribution or “map”. These maps can be interrogated by human-guided or automated techniques to categorize single cell data into relevant biological populations and otherwise visualize important differences between samples. The method has been extensively adopted and reported in the literature to be superior to traditional biaxial gating. The analyst must carefully choose the parameters of a t-SNE computation, as incorrectly chosen parameters might create artifacts that make the resulting map difficult or impossible to interpret. The correct choice of algorithm parameters is complicated by a lack of agreed-upon quantitative framework for assessing the quality of algorithm results. Gauging result quality currently relies on subjective visual evaluation by an experienced t-SNE user.

To overcome these limitations, we used Cytobank viSNE engine for all t-SNE analyses and employed 18-parameter flow cytometry data as well as 32-parameter mass cytometry data of varying numbers of events to optimize t-SNE parameters such as total number of iterations and perplexity. We also investigated the utility of Kullback-Liebler (KL) divergence as a metric for map quality as well as SPADE clustering as an indirect measure of multidimensional data integrity when flattened into t-SNE coordinates.

We have established the imperative requirement for the number of t-SNE analysis optimization steps (‘iteration number’) to be scaled with the total number of data points (events) in the set, suggesting that a number of existing software solutions produce unclear t-SNE maps of flow and mass cytometry data due to built-in user control restrictions. We also evaluated lower-level parameters within the t-SNE code that control the ‘early exaggeration’ stage initially introduced into t-SNE algorithm for better map optimization. These parameters are not available as part of the standard algorithm interface, but we found that they can be tuned to produce high quality results in shorter periods of time, avoiding unnecessary increases of both analysis duration and computation cost. Therefore, our approach allows to fine-tune the t-SNE analysis to ensure both optimal resolution of t-SNE low-dimensional maps and better faithfulness of their presentation of high-parameter cytometry data.