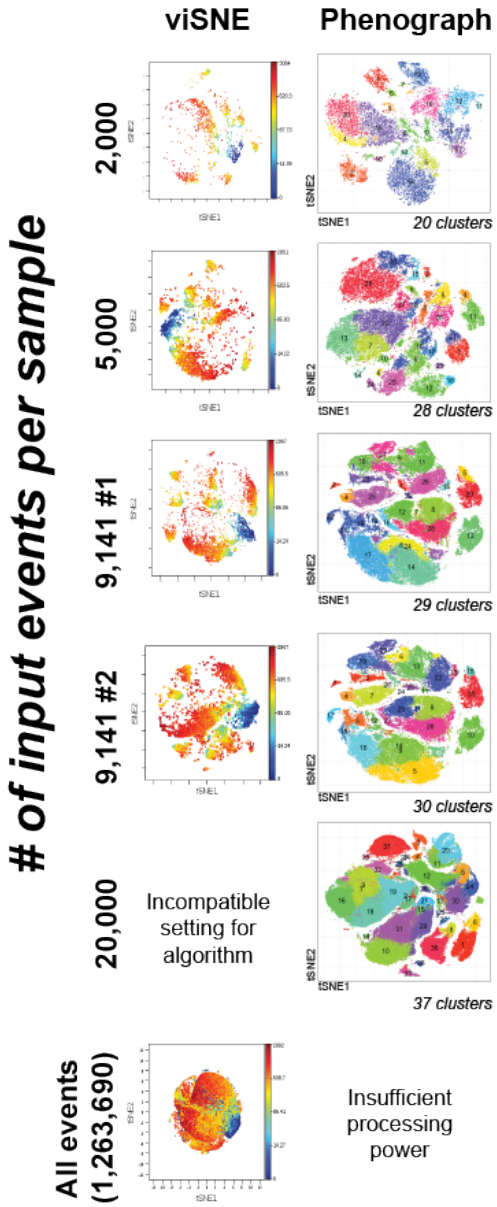
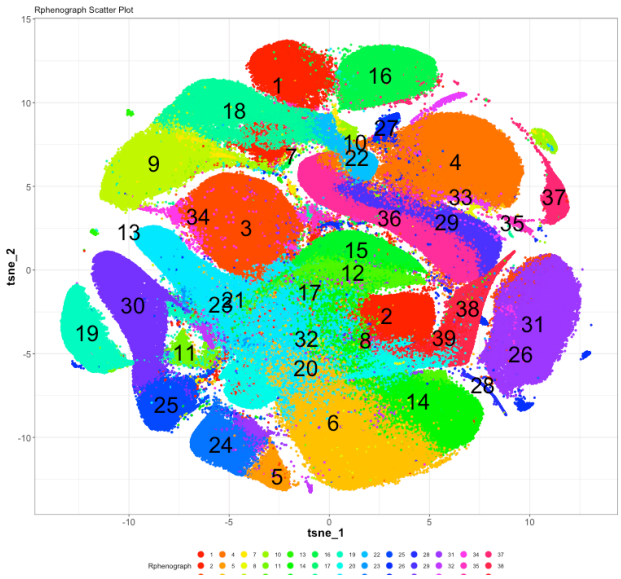
Q: How many events should I input into viSNE/PhenoGraph?

A: The answer to this question really depends on the quality of your experiment, your research question, and the processing power of the computer you are using for analysis. The main issue you'll run into with trying to run viSNE or PhenoGraph on a large amount of cells is: overcrowding your tSNE plot and the length of computing time. Notice that when trying to cluster all events (~1.2 million events total) islands on the viSNE plot are impossible to discern and for PhenoGraph the R program crashed (Figure 9F)!



Kimball, AK … ET Clambey, J Immunol 2018

Here is an example of an overcrowded tSNE plot:



Notice how the clusters start stacking on top of each other, making analysis very difficult…

Here are some of my recommendations for PhenoGraph analyses:

- If you acquired a small amount of events (due to cell loss, issues with your machine, etc.) you will most likely want to use all of the available events for analysis.
    - *In some cases I have concatenated events from multiple replicates together in order to increase the # of events in each condition for analysis. By doing this I have sacrificed the statistical power of my experiment, but I've increased the chances of having a successful PhenoGraph run that could reveal shifts in cellular abundance/expression.*
- If you acquired a large amount of events (<10,000 events per sample) I would recommend using the merge method: "ceiling" and using a fixed number of 10,000 events. So, if you are analyzing 10 samples the total number of events that will be clustered and displayed on the tSNE plot will be 100,000 events.
    - *If you have a lot of replicates I wouldn't go over a total number of ~1 million, adjust your fixed number accordingly.*

    - *If you have a range in events amongst your sample you can select the merge method: "min", this way all samples will be equally sampled but you can maximize the number of events analyzed.*
- If you are interested in a specific type of cell I would recommend further gating on live, singlet events prior to PhenoGraph analysis. For example, if you are looking for memory T cells, consider gating your live, singlet events for CD3+ events in FlowJo. Export this population and input into cytofkit. Then select the merge method "min" and run a PhenoGraph analysis on all relevant parameters (not including the CD3 marker as all input events will be positive for CD3). This will allow you to find this rare population without overcrowding your tSNE plot.

Please note that this is an issue you must consider when utilizing viSNE in Cytobank or the tSNE plugin in FlowJo, the solutions are identical to those above but the names of settings may be slightly different. The new algorithm "UMAP" promises to ameliorate this issue, but until recently UMAP has been inaccessible to most bench scientists as it is computationally intensive and written in Python. However, cytofkit2 (the newest version of Cytofkit) has UMAP built in. I am currently working on getting this package successfully running at Anschutz.

I have also collaborated with TICR on campus to try to modify the PhenoGraph code so the analysis can be run through the super computer on campus (removing the computational power limitation), please contact me if you're interested in pursuing this project further.

Figuring out the right settings for your experiment and your computational resources can take some time. I often do many PhenoGraph runs on a single dataset with a variety of settings. As you analyze your data it's natural that your experimental question changes and thus your PhenoGraph settings will too.