Algorithmic analysis of B cells uncovers new populations critical to the immune response during T. gondii infection

Adrianna Alpern[1], Scott Souza[2], Kirk Jensen[2], David Gravano[1]
[1]Stem Cell Instrumentation Foundry, UC Merced
[2]School of Natural Sciences, UC Merced

BACKGROUND

Working with large amounts of data is very difficult and time-consuming, particularly when comparing more than three variables at the same time. Since our brains are incapable of visualizing structures beyond the third dimension, a tool that can map multiple variables on a two-dimensional map is needed. T - distributed Stochastic Neighbor Embedding, or tSNE, is a method used to visualize higher-dimensional data (data beyond 3 dimensions) on a 2-D graph, also known as dimensionality reduction. tSNE is able to do this by aiding in visualization of data points by similar qualities based on the multiple variables being compared. FlowSOM, a plugin tool used in Flowjo, a flow cytometry tool, can generate population clusters that can be visualized on a tSNE map. These two tools can be used together to organize and analyze large datasets with multiple variables in an efficient, convenient way, with much more objectivity. This study discusses the process of using tSNE and FlowSOM to study differences in immune response to *Toxoplasma gondii*, a parasitic organism that is known to cause potentially lethal symptoms in chronically infected mice. The mice used in the experiment were wild-type B6 mice and mutant *Bumble* mice. *Bumble* mice are genetic mutants that cannot produce *Nfkbid*, a protein that controls B-cell activation. Data used in this study are all provided by Scott Souza et al.

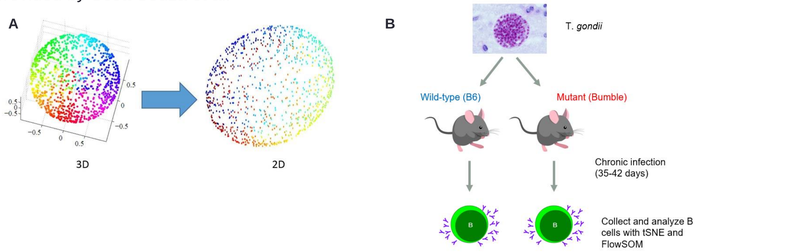


Figure 1. **A visualization of dimensionality reduction, where a three-dimensional graph is reduced to two dimensions (Rosero et al., 2017) (A). General layout of experimental method by Souza et al (B).**

METHODS

1. Collect mouse spleen and bone marrow cell data using flow cytometry (data from Scott Souza et al.).
2. Transfer data to Flowjo application.
3. Adjust parameter compensation in Flowjo.
4. Pregate for forward and side scatter, live cells, and B cells.
5. Downsample to get even sample sizes.
6. Export and concatenate data
7. Generate tSNE map.
8. Create populations using FlowSOM plugin.
9. Output Heatmap and Excel table for t-testing.
10. Select populations of interest for further analysis.

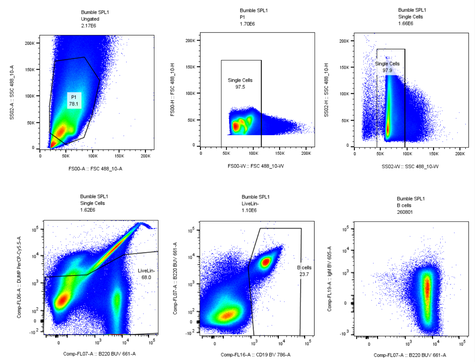


Fig. 2 : Mouse Cell Pregating method used in the Flowjo tool.

RESULTS

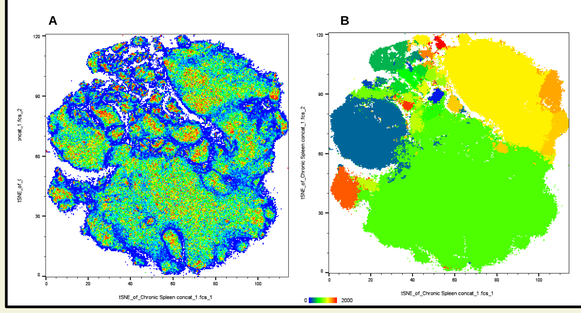
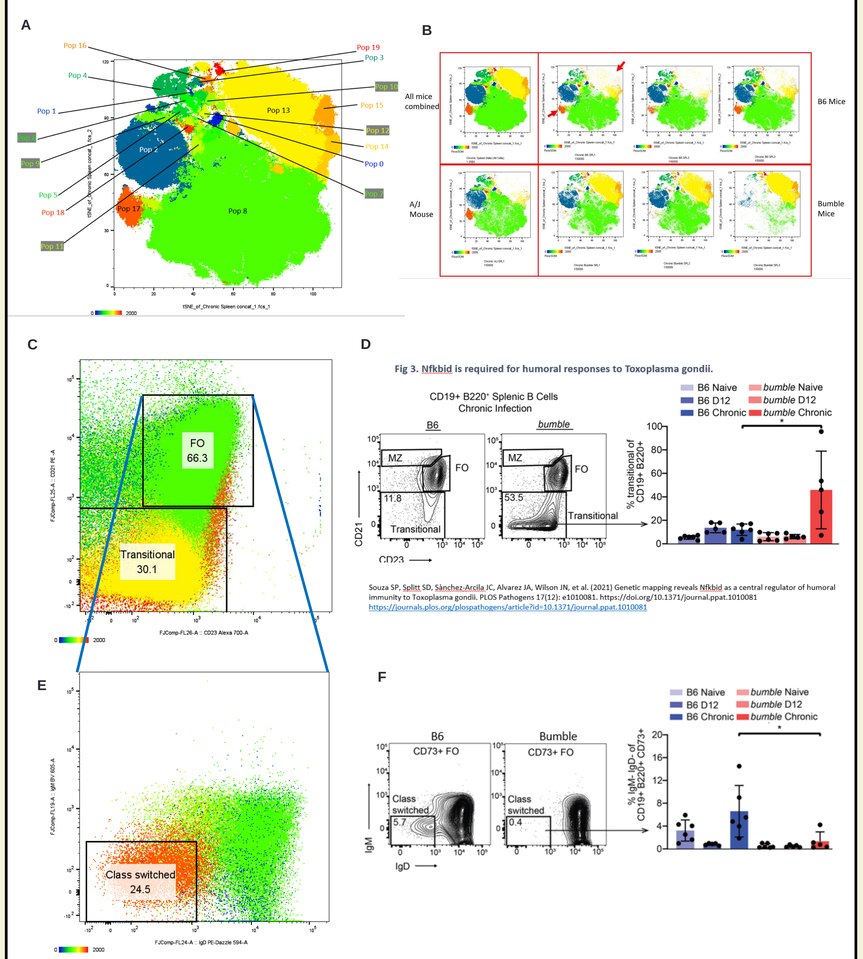


Figure 3. **tSNE and FlowSOM can be used in tandem to visualize higher dimensional data.** tSNE allows for a two-dimensional view of higher-dimensional data (A). FlowSOM can generate colored population clusters for easier viewing and analysis (B).

RESULTS



G: Populations of Transitional gate

Population number	Percent of gate	# of events
FlowSOM.Pop0	0.057	179
FlowSOM.Pop1	0.19	586
FlowSOM.Pop2	0.88	2763
FlowSOM.Pop3	0.12	368
FlowSOM.Pop4	1.60	5550
FlowSOM.Pop5	0.99	3109
FlowSOM.Pop6	0.24	766
FlowSOM.Pop7	0.050	158
FlowSOM.Pop8	8.32	26167
FlowSOM.Pop9	0.073	231
FlowSOM.Pop10	1.66	5218
FlowSOM.Pop11	0.53	1672
FlowSOM.Pop12	0.041	130
FlowSOM.Pop13	69.5	218626
FlowSOM.Pop14	5.95	18735
FlowSOM.Pop15	5.96	18757
FlowSOM.Pop16	0.60	1884
FlowSOM.Pop17	1.58	4967
FlowSOM.Pop18	0.72	2252
FlowSOM.Pop19	0.90	2845

H: Populations of FO gate

Population number	Percent of gate	# of events
FlowSOM.Pop0	0.63	4411
FlowSOM.Pop1	0.22	1513
FlowSOM.Pop2	17.3	120847
FlowSOM.Pop3	0.18	1229
FlowSOM.Pop4	0.032	224
FlowSOM.Pop5	0.042	296
FlowSOM.Pop6	0.055	386
FlowSOM.Pop7	0.97	6788
FlowSOM.Pop8	74.5	521017
FlowSOM.Pop9	1.20	8402
FlowSOM.Pop10	0.19	1319
FlowSOM.Pop11	1.71	11966
FlowSOM.Pop12	0.014	101
FlowSOM.Pop13	0.051	357
FlowSOM.Pop14	0.50	3504
FlowSOM.Pop15	2.57E-3	18
FlowSOM.Pop16	4.29E-3	30
FlowSOM.Pop17	2.30	16059
FlowSOM.Pop18	0.16	1108
FlowSOM.Pop19	1.43E-4	1

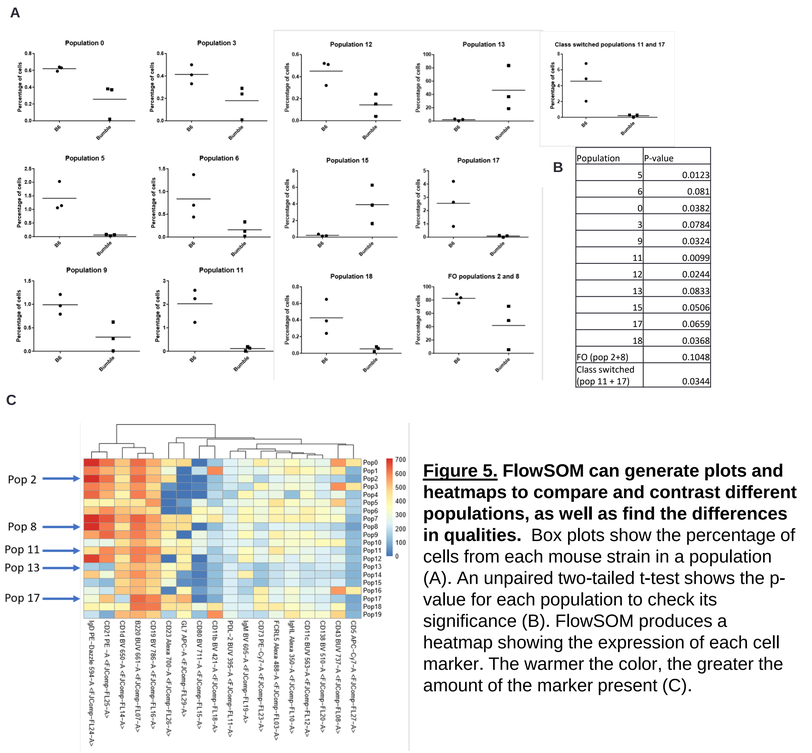
I: Populations of Class Switched Gate

Population number	Percent of gate	# of events
FlowSOM.Pop0	0.99	117
FlowSOM.Pop1	0.19	23
FlowSOM.Pop2	0.97	115
FlowSOM.Pop3	0.40	47
FlowSOM.Pop4	8.45E-3	1
FlowSOM.Pop5	0.22	26
FlowSOM.Pop6	0.18	21
FlowSOM.Pop7	0.23	27
FlowSOM.Pop8	1.34	159
FlowSOM.Pop9	2.60	308
FlowSOM.Pop10	0.63	75
FlowSOM.Pop11	10.3	1223
FlowSOM.Pop12	0	0
FlowSOM.Pop13	8.45E-3	1
FlowSOM.Pop14	8.45E-3	1
FlowSOM.Pop15	0	0
FlowSOM.Pop16	0.16	19
FlowSOM.Pop17	78.5	9290
FlowSOM.Pop18	3.27	387
FlowSOM.Pop19	0	0

Figure 4. **Using tSNE and FlowSOM to find populations of interest in mouse spleen cells.** We used FlowSOM to divide the data into 20 populations, ranging from populations 0 to 19 (A). FlowSOM clusters can also be organized based on mouse type to compare the populations of each mouse (B). Gating within CD23 vs. CD21 graph for FO and transitional cells (C). In the FO gate, we compared IgD and IgM and made another gate, class switched (E). Both tables show consistency with gating methods used in published work by Souza et al (D, F).

Analyzing FlowSOM population statistics in the FO gate and the transitional gate reveals which populations make up the majority of each gate (green boxes). According to the data, it seems that the FO gate is made up mostly of Population 8 and Population 2 (F). On the other hand, the transitional gate is made up largely of Population 13 (G). The populations that make up the class switched gate are mainly Population 11 and Population 17 (H).

RESULTS



B

Population	P-value
5	0.0123
6	0.081
0	0.0382
3	0.0784
9	0.0324
11	0.0099
12	0.0244
13	0.0833
15	0.0506
17	0.0659
18	0.0368
FO (pop 2+8)	0.1048
Class switched (pop 11 + 17)	0.0344

Figure 5. **FlowSOM can generate plots and heatmaps to compare and contrast different populations, as well as find the differences in qualities.** Box plots show the percentage of cells from each mouse strain in a population (A). An unpaired two-tailed t-test shows the p-value for each population to check its significance (B). FlowSOM produces a heatmap showing the expression of each cell marker. The warmer the color, the greater the amount of the marker present (C).

CONCLUSION

By using tSNE and FlowSOM, the data in this experiment became much easier to analyze, and the data showed clear differences in B cell activity in wild-type versus *Bumble* mice. With these tools, organizing and analyzing large amounts of data with lots of variables became much faster and more accurate. Despite theses benefits, it is important to note that this isn't a perfect algorithm, as it is prone to errors made when collecting data from flow cytometry, requires human intervention to calibrate any over- or under-compensation, and has issues with distinguishing data taken at different time points. Nonetheless, it is a great tool for understanding high-dimensional analysis, and hopefully will be utilized and refined in the future.

ACKNOWLEDGEMENTS & REFERENCES

Funding provided by the UC Merced Student Success Internship Program and ISAC SRL Emerging Leader award to David M. Gravano.

We would like to thank the UC Merced Stem Cell Instrumentation Foundry flow cytometry core for technical support.

References

- Rosero, Paul & Diaz, P. & Salazar Castro, Jose & Peña, Diego & Anaya Isaza, Andres & Alvarado Pérez, Juan & Therón, Roberto & Peluffo, Diego. (2017). Interactive Data Visualization Using Dimensionality Reduction and Similarity-Based Representations. Lecture Notes in Computer Science. 10125. 334-342. 10.1007/978-3-319-52277-7_41.
- Souza SP, Splitt SD, Sànchez-Arcila JC, Alvarez JA, Wilson JN, Wizzard S, Luo Z, Baumgarth N, Jensen KDC. Genetic mapping reveals Nfkbid as a central regulator of humoral immunity to Toxoplasma gondii. PLoS Pathog. 2021 Dec 6;17(12):e1010081. doi: 10.1371/journal.ppat.1010081. PMID: 34871323; PMCID: PMC8675933.
- Van Gassen S., et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytometry A. 2015 Jul;87(7):636-45.