

Integrating Computational and Visual Analytics for Cell-Based Biomarker Discovery from Polychromatic Flow Cytometry Data

Aishwarya Mandava¹, Ivan Y. Chang¹, Kim D. Lu², Frank Zaldivar², Dan M. Cooper², Richard H. Scheuermann^{1,3}, Yu Qian¹

¹Department of Informatics, J. Craig Venter Institute, La Jolla, CA; ²Department of Pediatrics, Pediatric Exercise and Genomics Research Center (PERC), University of California, Irvine ;³Department of Pathology, University of California, San Diego, CA

Motivation

Traditional Manual gating analysis on 2-dimensional plots is subjective and time consuming.

Computational auto-gating methods such as FLOCK(Flow Clustering without K) use data driven approaches and have successfully replaced manual gating analysis.

However, validating and evaluating the clusters in an interpretable way remain a challenge especially while identifying novel and rare cell populations.

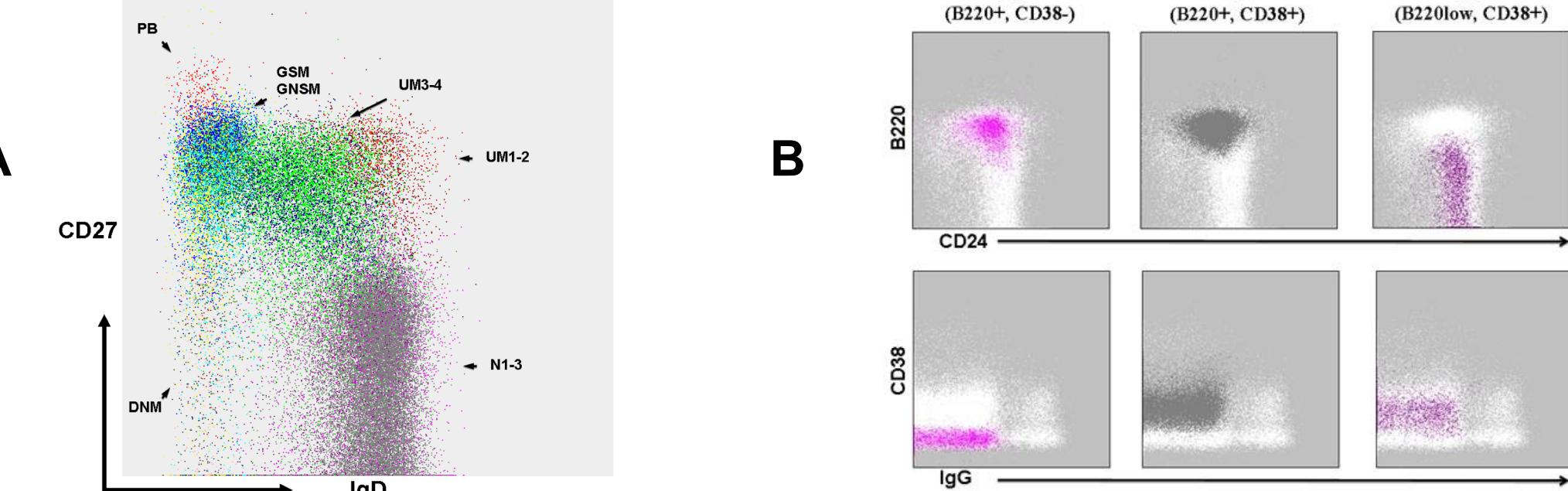


Figure 1: A) Example of an auto-gating method, FLOCK(Flow Clustering without K), which is a density based clustering approach to identify cell populations. B) FLOCK identified clusters on other marker combinations

Approach-Advanced Visualization

Our group at JCVI has incorporated a visual analytics tool into the computational pipeline that integrates visualizing methods such as t-SNE or t-Stochastic Neighbor Embedding(t-SNE), 2-Dimensional and 3-Dimensional plots that enable to validate and evaluate the results.

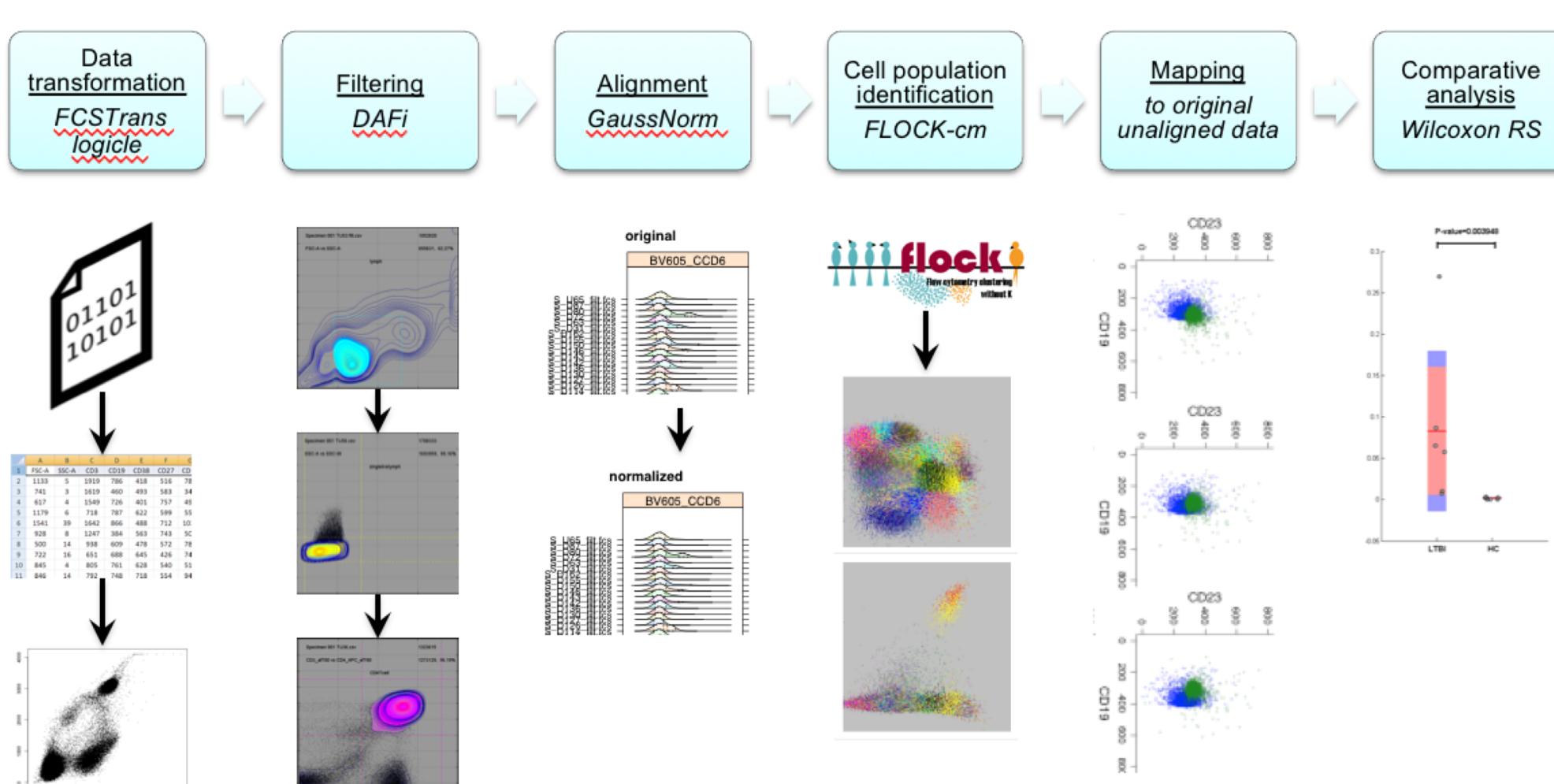


Figure 2: The computational pipeline that has assembled a series of algorithms, including an initial Logicle transformation using FCSTrans, base population filtering using DAFI(directed automated filtering and identification of cell populations), cross sample alignment using GaussNorm, population clustering using FLOCK(Flow clustering without K), and finally non-parametric statistical comparison using the Wilcoxon Rank Sum test.

Application 1

Application 1: Sexual dimorphism in glucocorticoid receptor (GR) expression in human leukocytes

Aim: This study aims to compare Glucocorticoid receptor expression on circulating leukocytes and to demonstrate sexual dimorphism in leukocyte Glucocorticoid receptor expression. We attempted to re-analyze this data using the manual gating hierarchy

Re-analysis: DAFI base population filtering was done following manual gating hierarchy. After DAFI pre filtering, FLOCK clustering was applied for population identification.

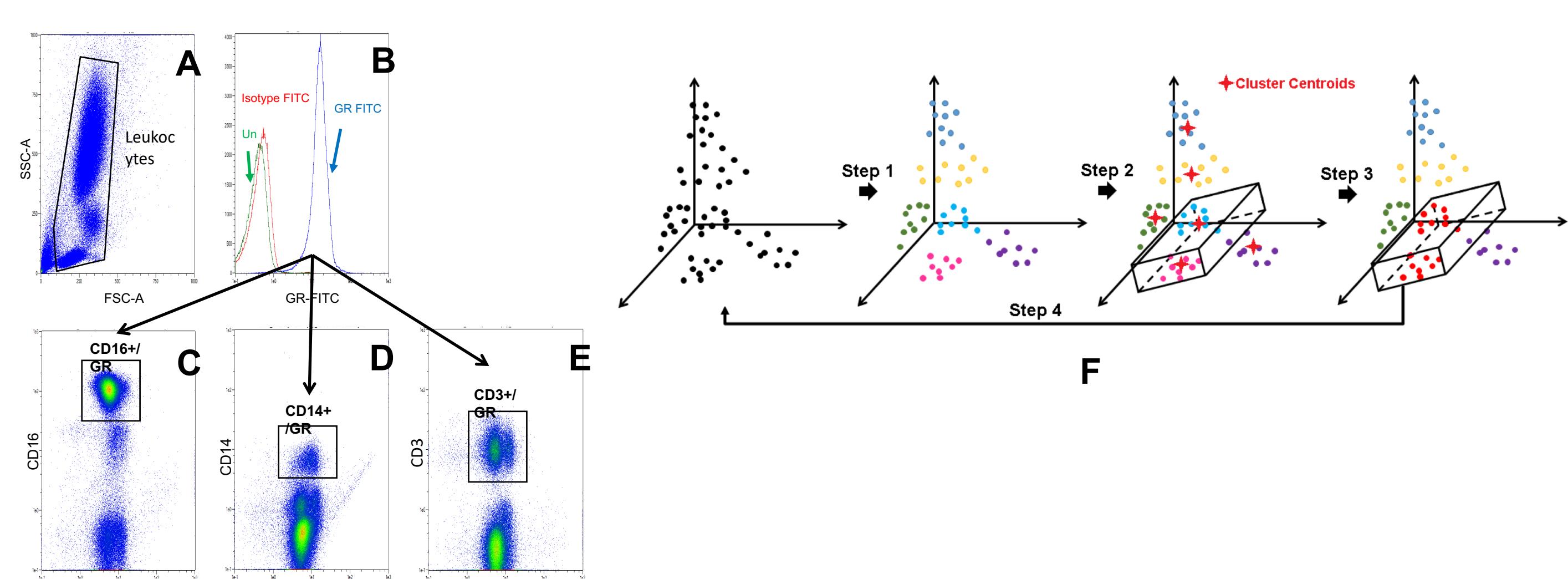


Figure 3: Manual gating hierarchy to identify GR+ leukocytes A: Leukocytes B: GR+ leukocytes C: GR+ Granulocytes D: GR+ Monocytes E: GR+ T-lymphocytes F: DAFI base population filtering

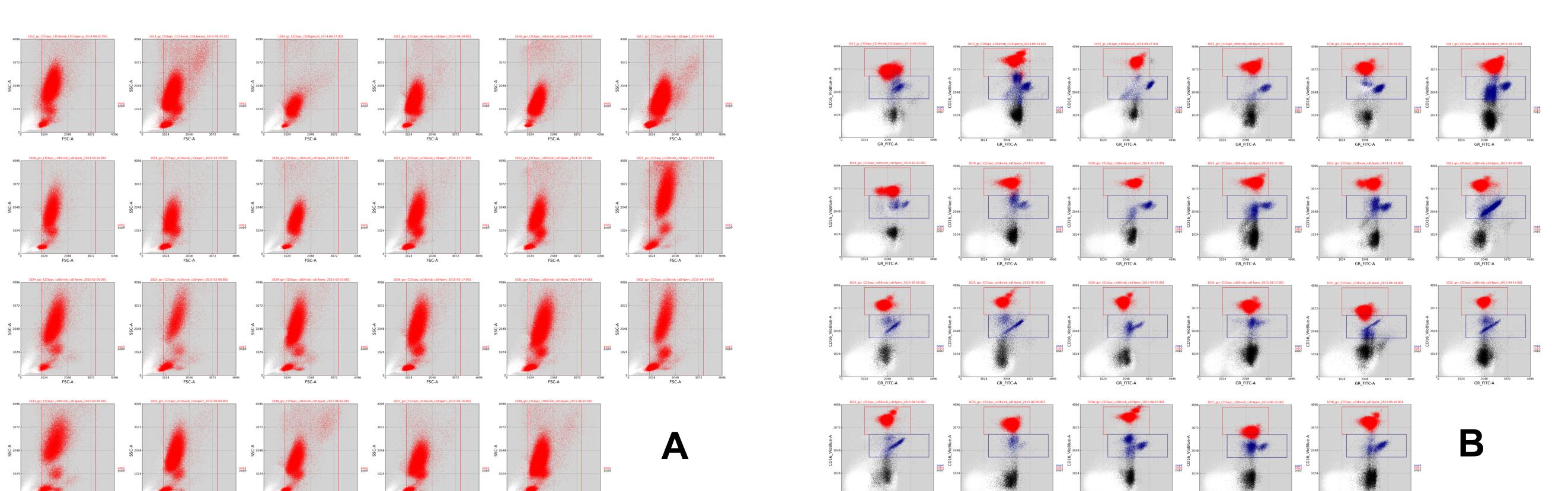


Figure 4: DAFI gating results for identifying A) Leukocytes B) GR+ CD16+ population

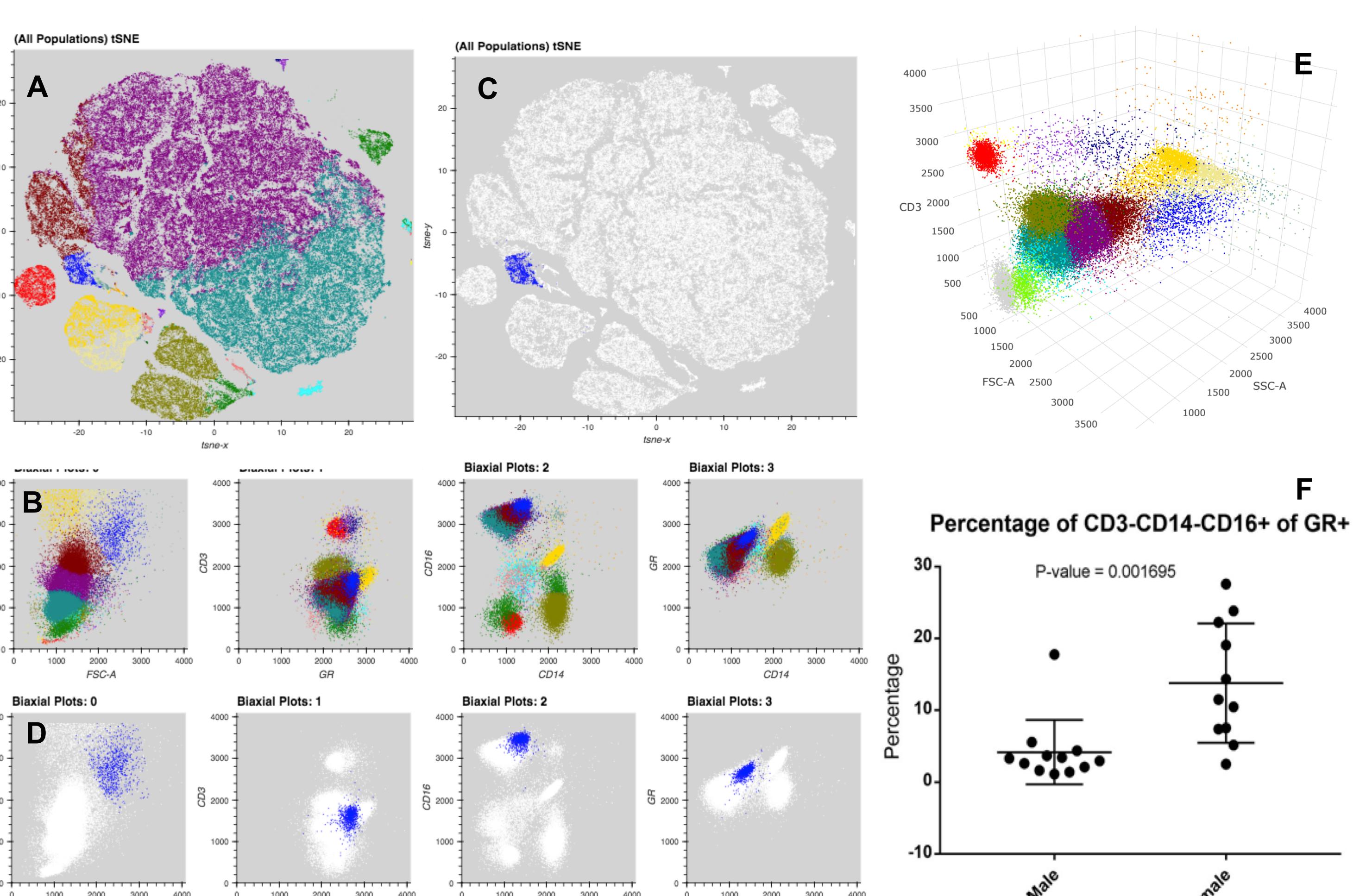


Figure 5: A) t-SNE map projecting the GR+ events on to the t-SNE transformed x and y axes. Clusters are color coded with FLOCK cluster IDs. B) The same clusters shown on the biaxial plots C) Selecting FLOCK identified cluster on the t-SNE map auto-refreshed on the biaxial plots in D. D) Selecting FLOCK identified cluster on the biaxial plots auto refreshed on t-SNE map in C. E) 3-Dimensional plots showing FLOCK identified clusters. F) Re-analysis using the visual analytics tool identified GR+ CD3-CD14-CD16+ subset that differed significantly between male and female subjects.

Application 2

Application 2: Identification of cell-based biomarker in latent tuberculosis infection(LTBI)

Aim: This study aims to identify rare T cell subsets that are significantly different between Latent Tuberculosis infected donors and Healthy controls

Re-analysis: DAFI base filtering followed by FLOCK clustering for population identification

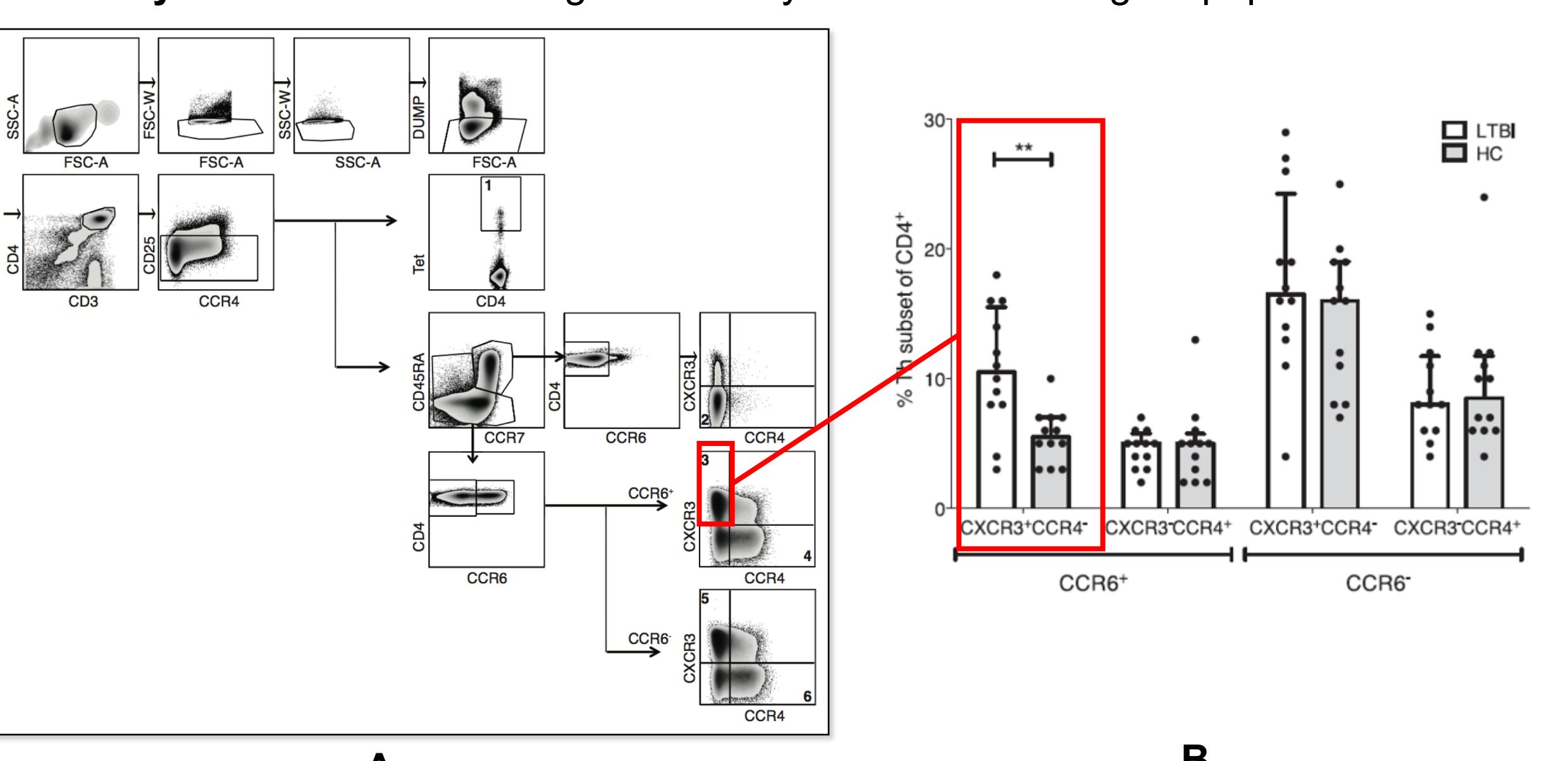


Figure 6: A) Manual gating hierarchy focused on CD4+ memory T cells. B) Previously identified populations: CCR6+ CXCR3+ CCR4-; CCR6+ CXCR3- CCR4-; CCR6- CXCR3+ CCR4-; CCR6- CXCR3- CCR4+. CXCR3+ CCR4- subset differed significantly between LTBI and HC donors, while the other subsets did not.

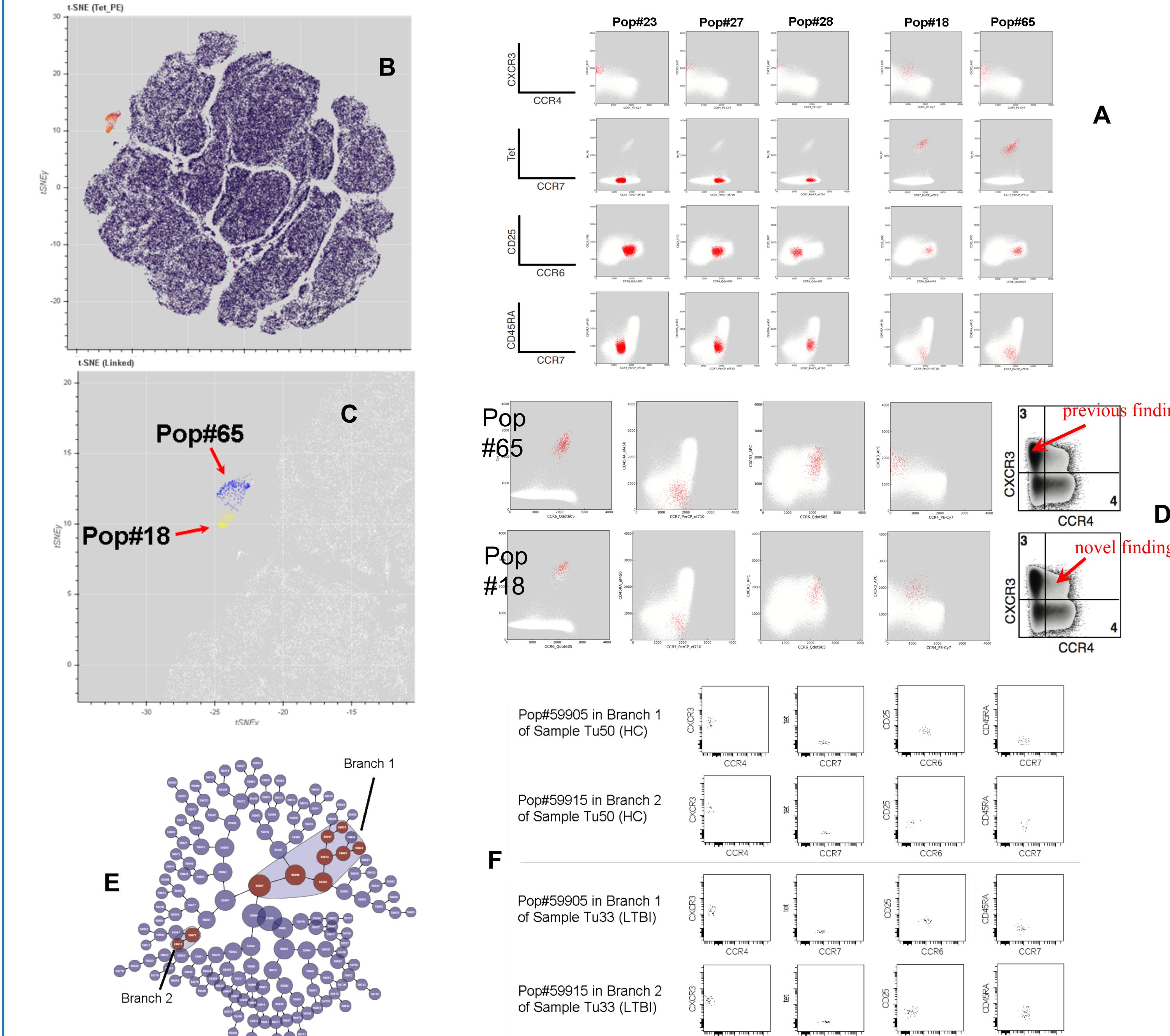


Figure 7: A) FLOCK identified populations. FLOCK identified two major tetramer positive populations 18 and 65. Population 18 is CCR4+ and population 65 is CCR4-. B) Populations 18 and 65 clusters on t-NSE map C) Zoomed-in t-SNE map showing the two populations. D) The previous study identified a single cell population CCR6+ CXCR3+ CCR4- population that comprised 23,27 and 65 . Re-analysis with DAFI using FLOCK identified two novel cell populations 18 and 28. E and F) CITRUS analysis. Branch 1 is CCR6+ CXCR3+ CCR4- corresponded to the findings from the previous analysis and branch 2 is CCR6- CXCR3+ CCR4- which matched population 28 from our results. CITRUS results could not identify population 18 from our results.

Cyberinfrastructure

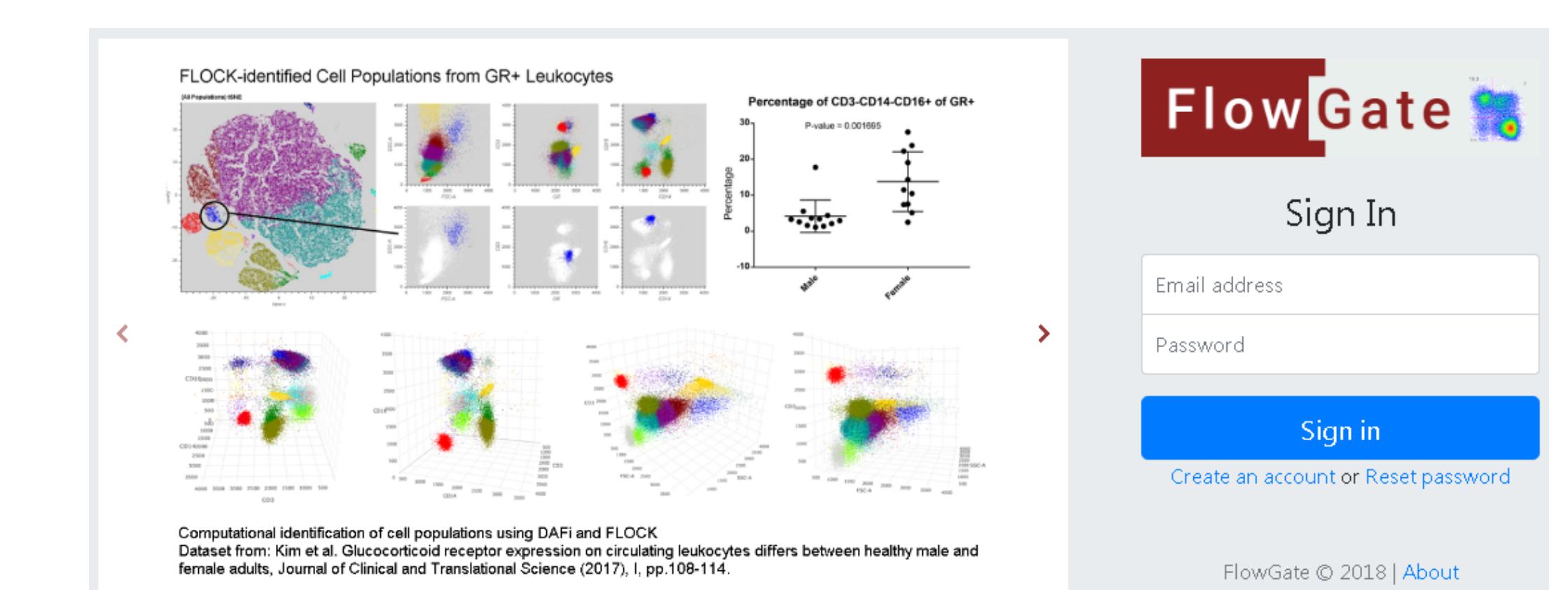


Figure 8: FlowGate infrastructure, for extensible and scalable web-based flow cytometry data analysis.

References

- Yukiko Kikuchi, Kohei Hosono, Junpei Yamashita, Yukika Kawabata, Kataaki Okubo, Glucocorticoid receptor exhibits sexually dimorphic expression in the medaka brain, General and Comparative Endocrinology, Volume 223, 2015, Pages 47-53
- Arlehamn, Cecilia Lindestam et al. "Transcriptional Profile of TB Antigen-Specific T Cells Reveals Novel Multifunctional Features." *Journal of Immunology (Baltimore, Md.)* 193.6 (2014): 2931-2940.
- Qian et al. Elucidation of Seventeen Human Peripheral Blood B cell Subsets and Quantification of the Tetanus Response Using a Density-Based Method for the Automated Identification of Cell Populations in Multidimensional Flow Cytometry Data. *Cytometry Part B, Clinical Cytometry*, 2010.
- Lee et al., DAFI: A Directed Recursive Data Filtering and Clustering Approach for Improving and Interpreting Data Clustering Identification of Cell Populations from Polychromatic Flow Cytometry Data *Cytometry Part A*, 2018
- NIH/NCATS U01TR001801 (FlowGate)