## Qeios

### Peer Review

# Review of: "CAFE: An Integrated Web App for High-Dimensional Analysis and Visualisation in Spectral Flow Cytometry"

#### Simone Puccio<sup>1</sup>

1. IRGB, National Research Council (Consiglio Nazionale delle Ricerche, CNR), Italy

The submitted work, *CAFE*, introduces a new tool for analyzing high-dimensional flow cytometry data. The manuscript is well-written from both technical and scientific perspectives. The authors have provided clear documentation, and the code is well-structured and publicly available on GitHub, showcasing a commitment to transparency and reproducibility. Additionally, "toy data" are supplied to enable users to test the tool. However, one notable limitation is that the tool requires intermediate to advanced programming skills, which may hinder accessibility for researchers who are not proficient in coding. This limitation is particularly significant given the use of the term "web tool," which implies a user-friendly, graphical interface accessible to a broader audience.

The primary concern with *CAFE* lies in its lack of novelty. The field of high-dimensional flow cytometry already has several established pipelines that integrate diverse clustering algorithms, offering robust and often innovative solutions for data analysis. In contrast, *CAFE* does not introduce a fundamentally new analytical approach or algorithm, which limits its distinction from existing tools. While the implementation of known methods in a streamlined pipeline can be valuable, it does not substantially advance the state of the art.

Another significant concern is the choice of batch effect removal methodology. The use of Combat, while widely recognized and effective in certain contexts, is not well-suited for high-dimensional flow cytometry data. Combat's general approach risks removing biological variability alongside technical variability, which could compromise the interpretation of results. More specialized methods, such as Cycombine or Cytonorm, have been specifically designed for flow cytometry and would have

been better suited for preserving biological signals while addressing batch effects. Incorporating such tailored methods would strengthen the reliability and applicability of the pipeline.

Lastly, there is room for improvement in the feature selection process. The reliance on the number of principal components (PCA) for dimensionality reduction is a common and effective approach in single-cell genomics; however, in the context of flow cytometry, this method may fail to address specific technical issues. For instance, PCA-based selection does not account for markers that exhibit technical bias during staining, which can introduce noise or confounding factors into the analysis. A more nuanced feature selection strategy, capable of identifying and excluding problematic markers, would greatly enhance the robustness of the analysis.

In conclusion, while *CAFE* provides a well-implemented and accessible tool for users with programming expertise, its lack of methodological innovation, non-specific batch correction, and limited feature selection strategy detract from its overall impact. Addressing these issues would significantly improve the utility and scientific contribution of the tool.

### Declarations

Potential competing interests: No potential competing interests to declare.