

flowMeans: Rapid Cell Population Identification in Flow Cytometry Data

Nima Aghaeepour^{1,2}, Radina Nikolic^{1,3}, Holger Hoos⁴, Ryan Brinkman^{1,5}

¹Terry Fox Laboratory, BC Cancer Agency, Vancouver, British Columbia, Canada ²Department of Bioinformatics, University of British Columbia, British Columbia, Canada ³Department of Statistics, University of Oxford, Oxford, United Kingdom ⁴Department of Computer Science, University of British Columbia, British Columbia, Canada ⁵Department of Medical Genetics, University of British Columbia, British Columbia, Canada

The K-means clustering algorithm was the first automated data analysis approaches applied to FCM data. However, the adoption of K-means has been restricted, because it requires the number of populations to be pre-identified, it is sensitive to its initialization, and it is limited to modelling spherical cell populations. We have developed flowMeans, a time-efficient and accurate method for automated identification of cell populations in flow cytometry (FCM) data based on K-means clustering. Unlike traditional K-means, flowMeans can identify concave cell populations by modelling a single population with multiple clusters. Our framework uses a change point detection algorithm to determine the number of sub-populations, enabling the method to be used in high throughput FCM data analysis pipelines. flowMeans is publicly available as an open source R package through Bioconductor.

Acknowledgements:

This project was supported by a Michael Smith Foundation for Health Research Scholar Award to RRB, a MSFHR/CIHR scholarship to NA, a University of British Columbia's graduate fellowship to NA, by NIH grant 1R01EB008400, Canadian Cancer Society grant #700374, and by an NSERC Discovery Grant held by HH. This research has been enabled by the use of computing resources provided by the Western Canada Research Grid (WestGrid) and Compute/Calcul Canada. The authors would like to thank Greg Finak and Raphael Gottardo from the Institut de recherches cliniques de Montreal and Thomas Lumley from the Department of Biostatistics, University of Washington for their comments on the methodology and, Nishant Gopalakrishnan from the Fred Hutchinson Cancer Research Center for his comments on the flowMeans Bioconductor package.