Loughborough University

This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

# Evaluating flow cytometry automated data analysis software [poster]

PLEASE CITE THE PUBLISHED VERSION

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Cheung, Melissa, Jonathan Campbell, Julian Braybrook, Robert James Thomas, and Jon N. Petzing. 2019. “Evaluating Flow Cytometry Automated Data Analysis Software [poster]”. figshare. https://hdl.handle.net/2134/38352.

Wolfson School of Mechanical, Electrical and Manufacturing Engineering

# Evaluating flow cytometry automated data analysis software

Melissa Cheung[1], Dr Jonathan Campbell[2], Prof Julian Braybrook[2], Prof Rob Thomas[1] and Dr Jon Petzing[1]

Loughborough University

## Introduction

### What is flow cytometry?

Flow cytometry is a technique used for single cell analysis. Cell subsets are identified by their intensities of fluorescent markers.

### Why we automate data analysis

Manual cell quantification varies between operators, has limited reproducibility, and is liable to bias[1]. Automated software is thought to be more reliable[2].

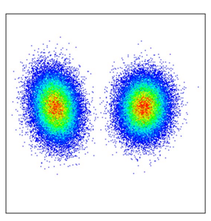
### What is the problem?

Variation potentially remains in many automated algorithms available. Hence, assurance in cell measurements from automated flow cytometry analysis is needed.

### Aims

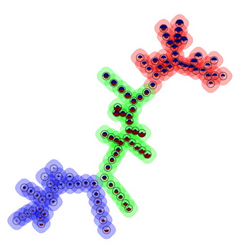
To define the confidence in flow cytometry automated data analysis software platforms in the context of cell therapy manufacturing

## Methods



Design synthetic flow cytometry datasets with controlled cell properties

Run datasets through automated software Flock2, FlowSOM and SPADE3[3-5]



Compare cell clusters before and after automated clustering

Analyse cell population accuracy and precision

## Results

### Multiple cluster testing

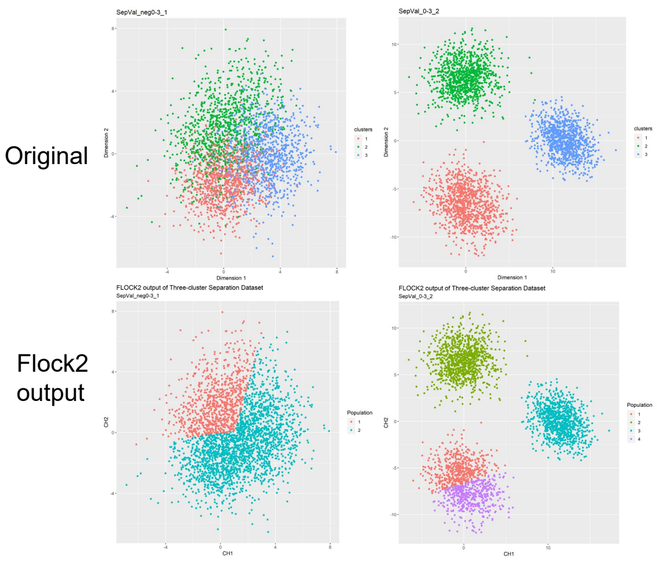



**Figure 1.** Example of incorrect number of clusters detected by Flock2

We generated datasets with two or three equal-sized clusters far apart from each other, then tested if different flow cytometry software obtained the correct number of clusters. This could not be tested on software which required the number of clusters to be defined. We found Flock2 had a fail rate of 21%.

### Cell separation testing

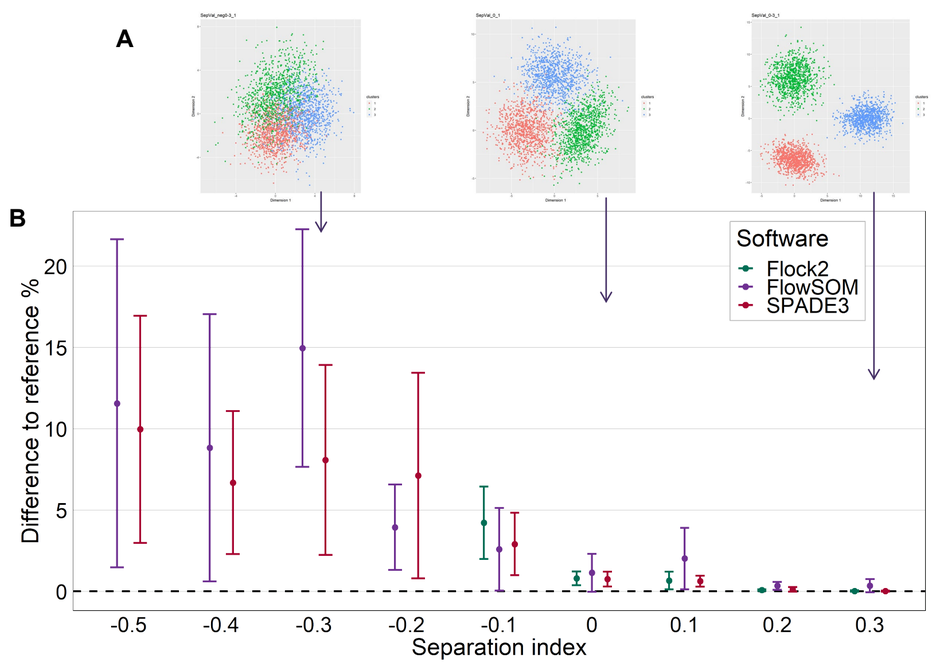



**Figure 2.** (**A**) Synthetic clusters with varying degrees of separations. (**B**) Difference in software output of cell population percentage compared with our known reference values. N=9, mean ±SD. Separation index range is -1 to 1, higher is more separated.

We designed datasets with clusters with varying separation indices[6], from far apart to merged, then tested the software ability to assign cells to their correct populations and quantify them. We found the accuracy and precision of software decreases as clusters get closer together.

### Rare cell testing

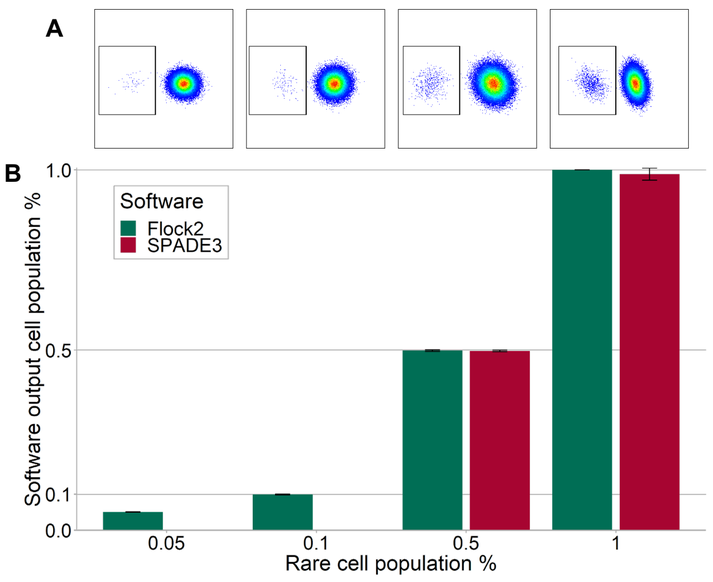



**Figure 3.** (**A**) Simulated rare cell populations, manually gated. (**B**) Detection of rare cells by automated software.

We simulated rare cell populations from 1% to 0.01% among $10^5$ total events and tested software sensitivity to our rare cells and the accuracy of the detection. We found SPADE3 did not detect events at 0.1%, and Flock2 accurately identified cells until 0.05%.

## Conclusions

This study is developing a set of synthetic datasets for evaluating the performance of clustering algorithms when identifying cell populations.

Our results allow users of the automated flow cytometry data analysis software the ability to define confidence in their data analysis and better understanding of the benefits and limitations of the algorithms.

For cell therapy manufacturers and regulators, this will help to inform decision making during in-process and final quality control of regenerative medicines.

**REFERENCES**

1. Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nature Reviews Immunology*. 2016 Jul;16(7):449.
2. Aghaeepour N, Finak G, Hoos H, Mosmann TR, Brinkman R, Gottardo R, Scheuermann RH, FlowCAP Consortium, Dream Consortium. Critical assessment of automated flow cytometry data analysis techniques. *Nature methods*. 2013 Mar;10(3):228.
3. Qian Y, Wei C, Eun-Hyung Lee F, Campbell J, Halliley J, Lee JA, Cai J, Kong YM, Sadat E, Thomson E, Dunn P. 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;78(S1):S69-82.
4. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, Saeys Y. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A*. 2015 Jul;87(7):636-45.
5. Qiu P. Toward deterministic and semiautomated SPADE analysis. *Cytometry Part A*. 2017 Mar;91(3):281-9.
6. Qiu W, Joe H. Generation of random clusters with specified degree of separation. *Journal of Classification*. 2006 Sep 1;23(2):315-34.

**ACKNOWLEDGEMENTS**

This research is kindly supported by the UK EPSRC/MRC Doctoral Training Centre for Regenerative Medicine at Loughborough University (EP/L105072/1), and LGC Ltd

**AUTHOR AFFILIATIONS**

[1] Centre for Biological Engineering, Loughborough University, UK. [2] LGC, Teddington, UK.

**CONTACT INFORMATION**

Wolfson School of Mechanical, Electrical and Manufacturing Engineering
Loughborough University
Leicestershire LE11 3TU UK
E. m.cheung@lboro.ac.uk
www.lboro.ac.uk/department/meme