FlowScatt: enabling volume-independent flow cytometry data by decoupling fluorescence from scattering

Ruud Stoof[1], Lewis Grozinger[1], Huseyin Tas[2], and Ángel Goñi-Moreno [1,3]

[1] *School of Computing, Newcastle University, Urban Sciences Building,Science Square, Newcastle upon Tyne NE4 5TG, U.K.*
[2] *Systems Biology Department, Centro Nacional de Biotecnología-CSIC, Campus de Cantoblanco, Madrid 28049, Spain*
[3] *Centro de Biotecnología y Genómica de Plantas (CBGP, UPM-INIA), Universidad Politénica de Madrid (UPM) , Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Campus de Montegancedo-UPM, 28223 Pozuelo de Alarcón, Madrid, Spain*

July 23, 2020

Abstract

Motivation: Measuring fluorescence by flow cytometry is fundamental for characterising single-cell performance. While it is known that fluorescence and scattering values tend to positively correlate, the impact of cell volume on fluorescence is typically overlooked. This makes of fluorescence values alone an inaccurate measurement for high-precision characterisations.
Results: We developed FlowScatt, an open-source software package that removes volume-dependency in the fluorescence channel. Using FlowScatt, flourescence values are re-calculated based on the unified volume per cell that arises from scattering decomposition.
Availability: FlowScatt is openly available as a Python package on https://github.com/rstoof/FlowScatt. Experimental data for validation is available online.
Contact: angel.goni-moreno@newcastle.ac.uk

1 Introduction

The single-cell characterisations of targeted molecular dynamics via flow cytometry (FC) is at the heart of systems and synthetic biology. From assisting the selection of genetic components to build novel bio-computing circuits (Nielsen *et al.*, 2016), to the analysis of complex phenotypes for metabolic engineering applications (Tracy *et al.*, 2010), FC is revealed as a fundamental method. Yet, the potential of this technique is often limited by the analysis capabilities of the software package of choice. As a response to this issue, many computational tools are being developed to, among other things, aid with the calibration (Beal *et al.*, 2019), stochastic analysis (Tiberi *et al.*, 2018) or gating (Becht *et al.*, 2019) of FC data.

Gene expression noise (i.e., fluctuations in the concentration of a protein being expressed by a gene) constitutes the *fingerprint* of information processing in a cell, and its characterisation is often the target of FC analysis. The features of noise patterns (e.g., their distribution or signal-to-noise ratios) are fundamental to understand key microbial mechanisms, such as promoter dynamics (Goñi-Moreno *et al.*, 2017) or infection types (García-Betancur *et al.*, 2017). However, noise is not merely a consequence of stochasticity, but also a signal that sometimes (positively) correlates to specific dynamics of the cell. That is the case of cell volume, since the average number of proteins per cell scales with volume (Kempe *et al.*, 2015). Therefore, the fluorescence signal of a cell (i.e., the noise of a specific reporter protein), as measured by FC, needs to be calculated based on its corresponding scattering value (FSC), as a proxy for cell volume, for the sake of accuracy. Otherwise, effects of cell volume, an important source of signal variability, would lead to misinterpretations of FC data.

Here we present FlowScatt, a software tool that re-calculates fluorescence information based on scattering values. As a result, the output is presented as volume-independent florescence signals and distributions which can be compared with other experiments.

2 FlowScatt

In order to reliably compare FC data between experiments, FlowScatt re-calculates fluorescence distributions for a single value of scattering. To minimize the impact of this process this value is chosen as the mean scattering for the experiments to be compared.
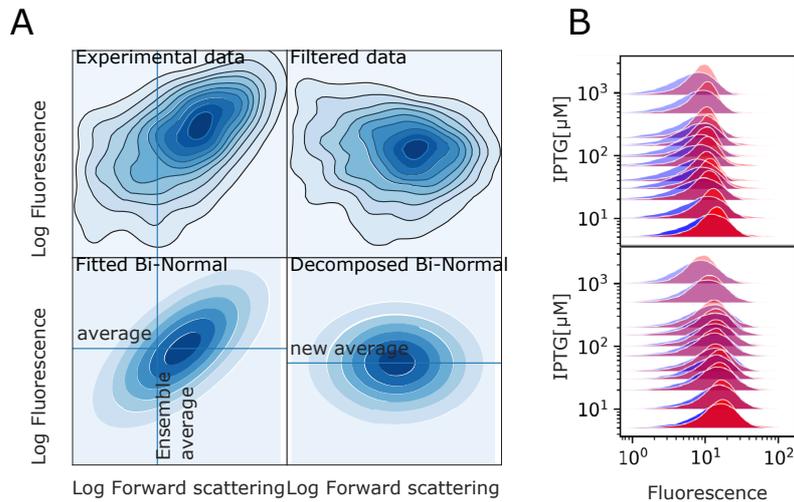
Figure 1: Calculating volume-independent fluorescent data. **A.** Experimental data is fitted to a binormal distribution, then re-calculated based on scattering decomposition (using average) to obtain new, volume-independent, fluorescence data. **B.** Overlap between experimental (blue) and filtered (red) data. Differences in overlap are due to the top experiments being further below the ensemble average for scattering than the bottom.

As a first step, the scattering and fluorescence data are log-transformed. The Probability Density Function of this transformed data is estimated and then fit to a bi-variate normal distribution:

$$f_{F,S} = \mathcal{N}(\mu_{\mathrm{F}}, \mu_{\mathrm{S}}, \sigma_{\mathrm{F}}, \sigma_{\mathrm{S}}, \rho) \tag{1}$$

Where f is the Probability Density Function described by the binormal distribution $\mathcal{N}$ and $\mu_{\mathrm{x}}, \sigma_{\mathrm{x}}$ and ρ are the inferred means, variances and covariance of the distribution. F and S indicate values for fluorescence and scattering, respectively.

Next, the average scattering of all experiments to be compared, $\bar{\mu}_{\mathrm{S}}$ is calculated. Then for each experiment, an estimation of the fluorescence distribution at scattering value $\bar{\mu}_{\mathrm{S}}$ can be obtained using a conditional bivariate log-normal distribution:

$$f_{\mathrm{F|S}=\bar{\mu}_{\mathrm{S}}} = \frac{\mathrm{F}\sigma_{\mathrm{S}} + \rho\sigma_{\mathrm{F}}(\bar{\mu}_{\mathrm{S}} - \mathrm{S})}{\sigma_{\mathrm{S}}} \tag{2}$$

In our example we corrected against the forward scattering height channel to negate volume variation effects, but for small organisms such as bacteria side scattering could be a good measure as well (Felip *et al.*, 2007).

3   Results

Using FC data as input, FlowScatt uses a filtering process to mitigate the impact of cell volume on fluorescence measurements and adjust their observed distribution appropriately, such that they may be used for comparison across experiments.

We exemplify the utility of FlowScatt by application to FC data from genetic NOT gates in bacterial cells described by Tas *et al.* (2020). Figure 1A shows an individual example experiment which is part of a group of experiments to be compared. The input data (top left) displays a correlation between the chosen fluorescence and scattering channels. If scattering channel distributions vary, for instance due to variation in cell volumes, this correlation will affect comparison between fluorescence distributions of different experiments. This is indeed the case for this example; the binormal distribution fit to this data shows that its mean scattering value (μ_{S}) lies above that of the group of experiments as a whole ($\bar{\mu}_{\mathrm{S}}$). FlowScatt adjusts for this variation to produce a decomposed binormal distribution and filtered data, as shown in Figure 1A (top and bottom right).

In Figure 1 B, differences between observed fluorescence distributions (blue) and adjusted distributions (red) defined by (2) can be seen for two groups of experiments, for which concentration of IPTG is the independent variable. Variability in the extent of overlap between these distributions reflects the variability of the observed scattering distributions for individual experiments.

It is possible that this first iteration of FlowScatt will not completely remove the influence of cell size on fluorescence (if it deviates from a power-law relation). This can be seen when filtered data and the scattering values still correlate (the top right of Figure 1). The tool can be re-run on the new data, however future work could generalise to other experimentally seen distributions.

4 Conclusion

Decoupling fluorescence from cell volume produces measurements where signal variability is minimally affected by the growth stage of the microbe. Moreover, results are comparable, not only within the same experiment, but across experiments with varying conditions that may affect the distribution of cell sizes. We advocate the use of FlowScatt as an addition to the efforts directed towards the standardisation of fluorescence measurements.

Funding

References

Beal, J. *et al.* (2019). Tasbe flow analytics: A package for calibrated flow cytometry analysis. *ACS synthetic biology*, **8**(7), 1524–1529.

Becht, E. *et al.* (2019). Reverse-engineering flow-cytometry gating strategies for phenotypic labelling and high-performance cell sorting. *Bioinformatics*, **35**(2), 301–308.

Felip, M. *et al.* (2007). Suitability of flow cytometry for estimating bacterial biovolume in natural plankton samples: comparison with microscopy data. *Applied and environmental microbiology*, **73**(14), 4508–4514.

García-Betancur, J.-C. *et al.* (2017). Cell differentiation defines acute and chronic infection cell types in staphylococcus aureus. *Elife*, **6**, e28023.

Goñi-Moreno, Á. *et al.* (2017). Deconvolution of gene expression noise into spatial dynamics of transcription factor–promoter interplay. *ACS synthetic biology*, **6**(7), 1359–1369.

Kempe, H. *et al.* (2015). The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. *Molecular biology of the cell*, **26**(4), 797–804.

Nielsen, A. A. *et al.* (2016). Genetic circuit design automation. *Science*, **352**(6281).

Tas, H. *et al.* (2020). Contextual dependencies expand the re-usability of genetic inverters. *bioRxiv*.

Tiberi, S. *et al.* (2018). Bayesian inference on stochastic gene transcription from flow cytometry data. *Bioinformatics*, **34**(17), i647–i655.

Tracy, B. P. *et al.* (2010). Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Current opinion in biotechnology*, **21**(1), 85–99.