Spectracular: minimizing spectral overlap in multicolor flow cytometry experiments

Authors:
Mike Bogetofte Barnkob[1], Alfredo Benso[2], Gianfranco Politano[2], Lars Rønn Olsen[3,*]

[1]Centre for Cellular Immunotherapy of Haematological Cancer Odense (CITCO), Department of Clinical Immunology, Odense University Hospital, University of Southern Denmark, Odense, Denmark
[2]Computer and Control Engineering Department, Politecnico di Torino, Italy
[3]Department of Health Technology, Technical University of Denmark
[*]Corresponding author

Contact information
Department of Health Technology, Technical University of Denmark, Kemitorvet, Building 204, 2800 Kongens Lyngby, Denmark
lronn@dtu.dk

Funding
This work was funded by the Independent Research Fund Denmark (grant 8048-00078B to LRO).

Abstract

Designing multicolor antibody panels for flow cytometry is complex and time-consuming. Poorly designed panels can result in repeated rounds of optimization and irreproducible results due to spectral overlap between fluorochromes. However, few tools exist to help researchers when designing panels, and none automatically considers spectral overlap. Here we present Spectracular, a web server which utilizes a complexity optimization algorithm to suggest fluorochrome combinations with minimal or near-minimal spectral overlap. Spectracular is available at https://biosurf.org/spectracular.html.

Introduction

Fluorescent flow cytometry remains a core methodology in cell biology research and clinical diagnostics. Primary usages include cell phenotyping, fluorescence-activated cell sorting, and sensitive validation of protein expression[1]. While these applications have traditionally been handled with panels of a dozen or less antibody-fluorochrome conjugates, advanced instruments with multiple lasers and detectors now enable the parallel measurement of up to 28-40 fluorochromes simultaneously[2,3]. This allows for novel biological insights into single cells in both health and disease[4], but to truly leverage panels of this size, they must be designed with minimal overlap between the fluorochromes' emission spectra.

Using multiple fluorescent proteins in the same panel will invariably increase the risk of emission spectra overlapping at unacceptable levels, and in turn lead to increased spill-over spreading error, which can confound the interpretation of the results[5]. Poor fluorochrome selections can also cause excessive background signal and errors when compensating post acquisition[6], and ultimately lead to irreproducibility between instruments, users, and laboratories[7]. As flow cytometers incorporate more lasers and users design larger panels of antibodies, minimizing spectral spillover between fluorochromes becomes increasingly important but also increasingly complicated[8,9].

While fluorochrome selection can be done manually by researchers when designing panels for a few antibodies, it quickly becomes time-consuming and technically challenging when including several antibodies[10]. In our experience, few flow cytometry users quantify the spectral overlap of their panels, but instead rely on online spectral viewers and personal experience. However, simply visually inspecting complex density curves is a poor way of identifying conflicting spectra, and humans perform suboptimally or even poorly, when attempting to solve combinatorial optimization problems of this sort[11]. Even approaching the problem with brute force computation is impossible in many cases, as the combinatorial complexity increases approximately exponentially with size of the pool of fluorochromes to choose from (e.g. choosing 20 fluorochromes out of a pool of 133 produces ~2.74×10^{23} different combinations to consider). With the advent of spectral flow cytometry, which enables even larger panels, the combinatorial complexity is bound to increase further.

To enable higher reproducibility and better panel design for flow cytometry, we have designed a complexity optimization algorithm to suggest optimal or near-optimal fluorochrome combinations on an easy-to-use web server.

Materials and Methods

Spectral data

Emission and excitation spectra were downloaded from Chroma Technology (www.chroma.com/), Semrock (www.semrock.com), and FPbase (www.fpbase.org/)[12]. In total, 133 commercially available fluorochromes were selected for the Spectracular web server. Based on the laser configuration submitted by the user, the emission spectra for all fluorochromes are calculated. A fluorochrome will be assumed to be excited by the laser which leads to the highest emission signal.

Calculation of overlap in *n* choose *k* spectra

The overlap between two emission spectra, *X* and *Y*, is calculated using the Szymkiewicz–Simpson coefficient (also known as the overlap coefficient):

$$overlap(X, Y) = \frac{|X \cap Y|}{min(|X|, |Y|)}$$

Where *X* and *Y* are two spectra to be compared. Selecting a combination of *k* choose *n*, has a combinatorial complexity of:

$$C(n,k) = \frac{n!}{(k!(n-k)!)}$$

In this case, *n* is the number of fluorochromes to choose from and *k* is the number of desired fluorochromes in the panel. In cases where the number of combinations is less than 10^7, all solutions are brute force calculated, thus providing the optimal solution. For cases with a greater combinatorial complexity, we lower the computational cost by approximating the optimal solution. This is done by first clustering the spectra into *k* clusters with hierarchical clustering using the 1st Wasserstein distance between the spectra. The clustered spectra can then be represented as a complete *k*-partite graph, by connecting vertices (representing fluorochromes) to all vertices not in the same cluster, where the edge weights correspond to the overlap coefficient. Then, the optimization problem can be formulated as a generalized minimal clique problem, with the approximate combinatorial space of:

$$C(n,k)_{approx} \approx \left(\frac{n}{k}\right)^k$$

This can be solved in a reasonable time by brute force computing all combinations in the fluorochrome set when efficiently parallelizing the computation.

Performance evaluation

To evaluate the performance of Spectracular, we calculated the maximum overlap coefficient for all spectra in a given solution, for *n* choose *k*, where *n*=30 and *k*=]1;30[, using the following 30 spectra: BUV395, Brilliant Violet 421, Horizon V450, Pacific Blue, Brilliant Violet 480, Horizon V500, BUV496, BUV563, Brilliant Blue 515, Brilliant Violet 510, FITC, PE, Brilliant Violet 605, PE-Dazzle 594, PE-CF594, Brilliant Violet 650, APC, BUV661, PE-Cy5, PerCP-Cy5.5, Brilliant Blue 700, Brilliant Violet 711, BUV737, Brilliant Violet 750, PE-Cy7, PE-Fire 780, APC-H7, APC-Fire 750, Brilliant Violet 786, and BUV805 (we were forced to perform this test on a subset of 30 fluorochromes, as increasing *n* will rapidly increase the combinatorial space to a size where it is infeasible to brute force calculate the optimal). For comparison with Spectracular, we calculated the brute force optimal solution, where the optimal solution for *k* fluorochromes was defined as the combination with the lowest maximum overlap coefficient. We also calculated the brute force worst solution, which was defined as the solution with the highest mean overlap coefficient. Finally we calculated the mean of 100 random solutions for each *k*. Similarly, we recorded the runtime for both brute force and Spectracular when run on a 3.1 GHz Intel Core i7 2017 macbook pro, with computation parallelized on 8 cores.

Comparison to expert-made panels

In order to compare results from Spectracular with human-made panels, we turned to the collection of Optimized Multicolor Immunofluorescence Panels (OMIPs) that have been published since 2010[13]. This collection of flow cytometry panels are made and reviewed by flow cytometry experts and the methodology is described in great detail. We collected the fluorochrome panels and instrument laser configurations for 63 panels (see **Supplementary Table 1**) that were aimed at multicolor flow cytometry (excluding mass cytometry panels and panels without complete information on instrumentation, laser setup and fluorochrome usage).

Implementation

Spectracular was implemented in R[14], and the web server was developed using Shiny[15]. While users can currently choose from 133 different fluorochromes, a list of 55 pre-selected fluorochromes are suggested to users initially. These were selected by comparing the emission overlap between all fluorochromes and choosing the ones with least overall emission overlap. A minimal excitation ratio for proposed fluorochromes is set to 25%, but can be changed by the user.

Results

Spectracular features

Spectracular provides solutions to several flow cytometry panel design challenges. To use the basic functionality of the tool, users input which lasers their flow cytometry instruments have, how many antibodies they wish to detect, and which fluorochromes they want to choose from. Spectracular uses publicly available spectral data from multiple sources including the majority of fluorochromes and dyes commonly used in flow cytometry (currently 133 fluorochromes, but additional will be added as necessary). Based on this information, Spectracular approximates the combination of *k* fluorochromes (where *k* is the desired panel size), with minimal spectral overlap.

Spectracular also offers more advanced functionalities. Users can select certain fluorochromes to always be included in a panel, and ask for suggestions for additional fluorochromes to be incorporated. This is useful when part of a panel has already been developed but a researcher wishes to incorporate extra antibodies or when a researcher wants to include an antibody that is only available with limited fluorochromes.

Performance

In order to test the computational performance of Spectracular, we calculated the mean overlap coefficient for all spectra in 28 cases (30 choose [2;29]) and compared this to the brute force optimal solution, the brute force worst, and the mean of 100 random combinations. Spectracular provides optimal or near-optimal solutions to any *k* in the test (**Figure 1A**) and provides this solution in approximately 8 seconds regardless of the combinatorial space. In contrast, the brute

force optimal solution becomes computationally expensive to calculate as the combinatorial space increases, taking more than 41 minutes to calculate for *k* = 15 (**Figure 1B**). While this may be reasonable to some, in a case where a user would want to brute force compare 60 fluorochromes for a 28 multicolor panel, the computational power needed fare exceeds what is available to most researchers (it would take approximately 50 years to calculate on a standard laptop). Spectracular will also approximate the solution to this problem in less than 10 seconds.

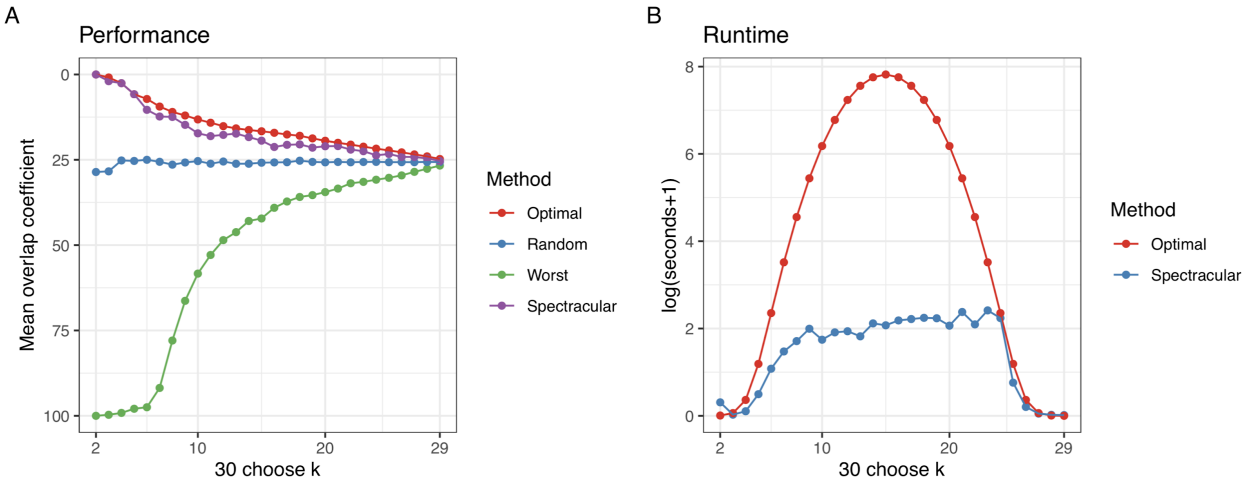

Figure 1: The performance and runtime of Spectracular of 30 choose 2-29 fluorochromes. A) Performance as measured by the mean overlap coefficient between all spectra in a given solution. The random method is the mean of 100 random combinations. B) Runtime of the brute force optimal solution and Spectracular in log(seconds+1).

We next wanted to compare panels suggested by Spectracular with panels designed and reviewed by flow cytometry experts. We therefore analyzed 63 experiments described within the OMIP collection[3]. These panels utilize between 2-5 lasers in order to excite between 5-39 antibody-fluorochromes conjugates (see **Supplementary Table 1**). In all cases Spectracular was able to find fluorochrome combinations that led to a lower mean overlap coefficient (**Figure 2**). It should be noted that in our analysis we assume to have access to antibodies with any fluorochrome conjugate. Still, it is clear from the detailed design considerations behind these panels, that researchers use days and weeks to iteratively screen and test antibody/fluorochrome combinations, Spectracular is able to provide on par or better combinations in a matter of seconds.

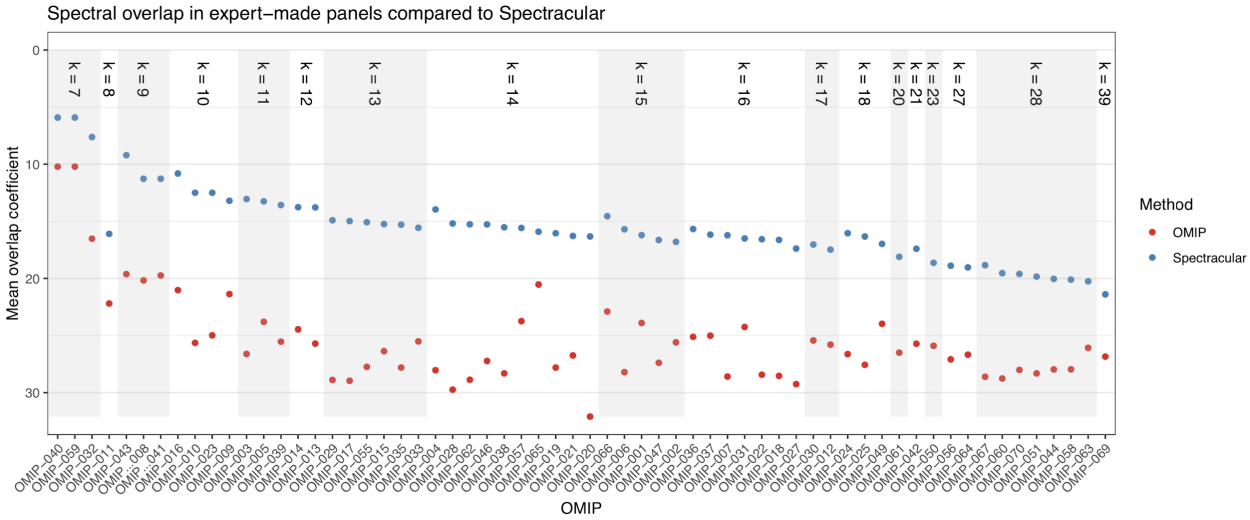

Figure 2: Analysis of the OMIP panels. The mean overlap coefficient of the fluorochromes in each OMIP was calculated and plotted in red. The mean overlap coefficient of the Spectracular solution of the same size was calculated and plotted in blue. OMIPs were ordered by the size of the panel.

Spectracular does not take into account users' flow cytometer filter settings, brightness of fluorochromes, abundance of antibody target, nor the natural decay of reagents such as tandem-dyes. These are all important factors to consider, but we specifically aimed for a simple, easy-to-use tool, which enables instantaneous solutions to traditionally time consuming tasks, while providing a near-optimal result.

Conclusion

Designing panels for multicolor flow cytometry is a complex and time consuming task. To address this challenge and thereby aid researchers and clinicians to build better panels faster, we have developed Spectracular, which utilizes complexity optimization algorithms to facilitate the rapid design of optimal or near-optimal multicolor flow cytometry panels from all commonly-used fluorochromes. We believe the performance and features of Spectracular will allow users to avoid repeated rounds of trial and error, and thus significantly reduce the amount of time and resources laboratories spend on designing new antibody panels.

Spectracular is not only useful for generating new panels but also updating existing ones. Our analyses show that improvements could be made to OMIP panels that are designed and reviewed by flow cytometry experts, when incorporating most currently available fluorochromes in the analysis.

To our knowledge, Spectracular is the first tool that enables a quick, easy and near-optimal computational solution to the complex problem of combining fluorochrome spectra. The tool is freely available for academic use at https://biosurf.org/spectracular.html.

References

1. Robinson, J. P. & Roederer, M. HISTORY OF SCIENCE. Flow cytometry strikes gold. *Science* **350**, 739–740 (2015).
2. Liechti, T. & Roederer, M. OMIP-060: 30-Parameter Flow Cytometry Panel to Assess T Cell Effector Functions and Regulatory T Cells. *Cytometry A* **95**, 1129–1134 (2019).
3. Park, L. M., Lannigan, J. & Jaimes, M. C. OMIP-069: Forty-Color Full Spectrum Flow Cytometry Panel for Deep Immunophenotyping of Major Cell Subsets in Human Peripheral Blood. *Cytometry A* **97**, 1044–1051 (2020).
4. Bendall, S. C., Nolan, G. P., Roederer, M. & Chattopadhyay, P. K. A deep profiler's guide to cytometry. *Trends Immunol.* **33**, 323–332 (2012).
5. Nguyen, R., Perfetto, S., Mahnke, Y. D., Chattopadhyay, P. & Roederer, M. Quantifying spillover spreading for comparing instrument performance and aiding in multicolor panel design. *Cytometry A* **83**, 306–315 (2013).
6. Maecker, H. T. & Trotter, J. Flow cytometry controls, instrument setup, and the determination of positivity. *Cytometry A* **69**, 1037–1042 (2006).
7. McLaughlin, B. E. *et al.* Nine-color flow cytometry for accurate measurement of T cell subsets and cytokine responses. Part II: Panel performance across different instrument platforms. *Cytometry A* **73**, 411–420 (2008).
8. Ferrer-Font, L. *et al.* Panel Design and Optimization for High-Dimensional Immunophenotyping Assays Using Spectral Flow Cytometry. *Curr. Protoc. Cytom.* **92**, e70 (2020).
9. Brummelman, J. *et al.* Development, application and computational analysis of high-dimensional fluorescent antibody panels for single-cell flow cytometry. *Nat. Protoc.* **14**, 1946–1969 (2019).
10. McLaughlin, B. E. *et al.* Nine-color flow cytometry for accurate measurement of T cell subsets and cytokine responses. Part I: Panel design by an empiric approach. *Cytometry A* **73**, 400–410 (2008).
11. Murawski, C. & Bossaerts, P. How humans solve complex problems: the case of the knapsack problem. *Sci. Rep.* **6**, 34851 (2016).
12. Lambert, T. J. FPbase: a community-editable fluorescent protein database. *Nat. Methods* **16**, 277–278 (2019).
13. Roederer, M. & Tárnok, A. OMIPs--Orchestrating multiplexity in polychromatic science. *Cytometry A* **77**, 811–812 (2010).
14. Team, R. C. R: A Language and Environment for Statistical Computing. (2020).
15. Chang, W. *et al.* shiny: Web Application Framework for R. (2021).