fies

th[illegible]analysis of multichromatic

[illegible][5], James Collier[5], Alexander Botzki[5], Josef Spidlen[4], Stéphanie Humblet-Baron[1,2] & Adrian Liston[1,2,3]✉

Compensating in flow cytometry is an unavoidable challenge in the data analysis of fluorescence-based flow cytometry. Even the advent of spectral cytometry cannot circumvent the spillover problem, with spectral unmixing an intrinsic part of such systems. The calculation of spillover coefficients from single-color controls has remained essentially unchanged since its inception, and is increasingly limited in its ability to deal with high-parameter flow cytometry. Here, we present AutoSpill, an alternative method for calculating spillover coefficients. The approach combines automated gating of cells, calculation of an initial spillover matrix based on robust linear regression, and iterative refinement to reduce error. Moreover, autofluorescence can be compensated out, by processing it as an endogenous dye in an

Fluorescently labeled antibodies and fl

constant across a broad range of fluorescence levels. Thus, a linear regression can be used to properly identify the slope between the two channels, that is, the spillover coefficient. As fluorescence data is heteroskedastic, owing to the effects of photon-counting

identified in particular channels, especially when populations did not conform to good separation (Fig. 2c, first and second columns). Traditional algorithms struggle in the case of poor

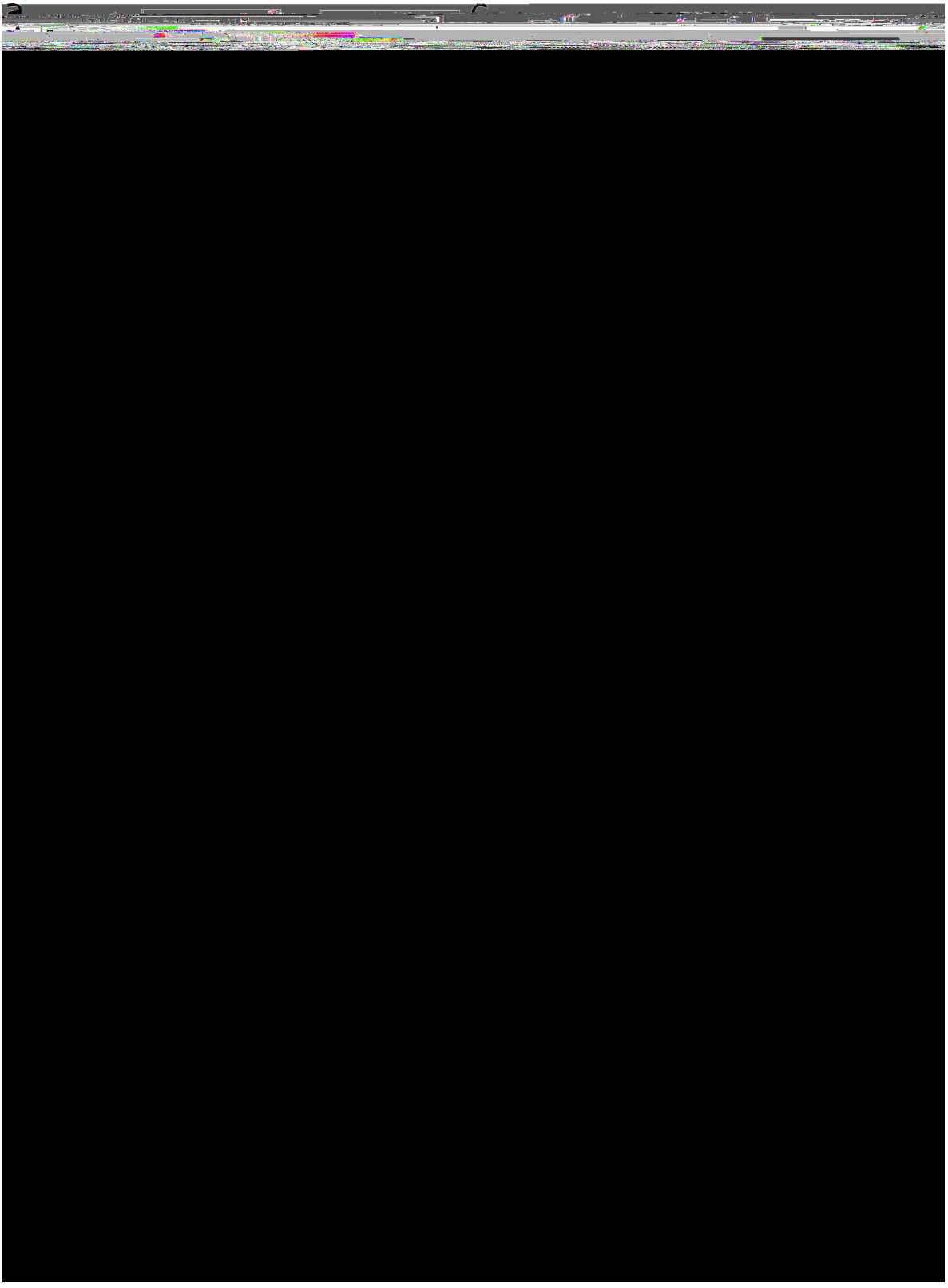
second, the spillover spreading coefficient. Coefficients deemed non-significant using an , -test were replaced with zeros, as well as any negative coefficients. The majority of quantiles were, in fact, subsamples of the traditional positive and negative

caused by σ_0 in the initial estimates, thereby allowing a more accurate estimation of the coefficients SS. If this adjustment step

were expanded through the development of new dyes and lasers, intracellular staining protocols were optimized for the detection of intracellular (and even post-translationally modifi

complete panel, that is, not just for the set of controls. In our experience, some panels still require minor modifications of the spillover matrix, which implies that the single-color controls do not fully describe the fluorescence properties of the complete panel, probably because of second-order phenomena such as secondary fluorescence or other interactions between dyes. Thus, this remains an open question.

While we demonstrate the utility of this method using eight representative datasets, the tool has been beta-tested more than 1000 times over a period of 22 months by more than 100



collaborating immunologists. This has allowed the development of a robust algorithm, designed to accommodate diverse datasets and to deal with less-than-perfect data arising in real-world experiments. The code is open source and is released with a permissive license, allowing integration into existing flow cytometry analysis pipelines in academia and industry. To increase access by research communities in immunology and other fields, we also provide a website (https://autospill.vib.be) that allows the upload of sets of single-color controls for calculating the spillover matrix with AutoSpill, produced in formats compatible with common software for flow cytometry analysis. As we have demonstrated by including AutoSpill in FlowJo v.10.7, this

algorithm is suitable for integration into commercial software, allowing for rapid and widespread uptake of superior flow cytometry compensation.

h

. Collaborating immunologists beta-tested AutoSpill over a period of 22 months, which allowed extensive testing and improvement of the algorithm for niche cases. Among these datasets, four are used as examples here, covering mouse cells, human cells, and beads. Compensation using AutoSpill, with default parameters, was carried out for each of these four sets of single-color controls: mouse splenocytes (MM1 dataset), human PBMCs (HS1 and HS2 datasets), and beads (Be1 dataset). We also analyzed four fully stained datasets, as examples of biological utility: mouse splenocytes (MM2 and MM3 datasets), and mouse microglia (MM4 and MM5 datasets). Data collection complied with all relevant ethical regulations for animal research and work with human participants. All animal experiments were performed in accordance with the University of Leuven Animal Ethics Committee guidelines or the Babraham Institute Animal Welfare and Ethics Review Body. Animal husbandry and experimentation complied with existing European Union and national legislation and local standards. Sample sizes for mouse experiments were chosen in conjunction with the ethics committees to allow for robust sensitivity without excessive use. For human experiments, written informed consent was obtained from all participants and the ethics committee of

ɜ , n y, . Splenocytes from C57Bl/6 mice were disrupted with glass slides, filtered through 100 μm mesh, and red blood cells lysed. Cells were stained with Fixable Viability Dye eFluor780 (eBioscience), fixed and permeabilized with Foxp3 transcription factor staining buffer set (eBioscience) according to the manufacturer's instructions, and stained overnight at 4 °C with the following antibodies: anti-CD4–BV421 (1:2000, clone N418), anti-CD24–BV510 (1:2000, clone M1/69), anti-Ly6G–BV570 (1:2000, clone 1A8), anti-XCR1–BV650 (1:2500, clone ZET), anti-CD19–BV785 (1:400, clone 1D3), anti-CD3–AF488 (1:1000, clone 145-2C11), anti-PDCA-1–PerCP-Cy5.5 (1:1000, clone 927), anti-CD23–PE (1:5000, clone B3B4), anti-CD64–PE-594 (1:500, clone X54-5/7.1), anti-CD172a–PE-Cy7 (1:5000, clone P84), anti-CD45–APC (1:10,000, clone 30-F11), anti-MHCII–AF700 (1:2000, clone M5/114.15.2) (all Biolegend), anti-IgE–BV605 (1:5000, clone R35-72), anti-CD93–BV711 (1:2000, clone AA4.1), anti-CD11b–BV750 (1:2000, clone M1/70), anti-CD80–BB630-P (1:2000, clone 16-10A1), anti-CD95–BB660-P2 (1:10,000, clone Jo2), anti-TCRβ–BB790-P (1:2000, clone H57-597), anti-CD103–BUV395 (1:1000, clone M290), anti-IgD–BUV496 (1:2000, clone 11-26c.2a), anti-Ly6C–BUV563 (1:500, clone AL-21), anti-Siglec F–BUV615-P (1:1000, clone E50-2440), anti-c-Kit–BUV661 (1:5000, clone 2B8), anti-CD21/35–BUV737 (1:5000, clone 7G6), anti-CD8a–BUV805 (1:500, clone 53-6.7) (all BD Biosciences), anti-IgM–PE-Cy5 (1:1000, clone Il/41) and anti-NK1.1–PECy5.5 (1:2000, clone PK136) (eBioscience). Compensation controls were stained as described in the MM1 dataset. Samples were acquired on a Symphony flow cytometer (BD Biosciences).

3 , n y, . Splenocytes from C57Bl/6 mice were disrupted with glass slides, filtered through 100 μm mesh, and red blood cells lysed. Cells were stained with Fixable Viability Dye eFluor780 (eBioscience), anti-CD90.2–BV510 (1:250, clone 53-2.1), anti-CD25–BV650 (1:200, clone PC61), anti-CD45–BUV395 (1:500, clone 30-F11) (all Biolegend), anti-CD127–PE (1:100, clone A7R34) and anti-B220–PE-Cy5 (1:200, clone RA3-6B2) (all eBioscience). Cells were fixed and permeabilized with Foxp3 transcription factor staining buffer set (eBioscience) according to the manufacturer's instructions, and stained overnight at 4 °C with the following antibodies: anti-T-bet–BV421 (1:200, clone 4B10), anti-CD8–BV785 (1:2000, clone 53-6.7), anti-NKp46–FITC (1:500, clone 29A1.4), anti-NK1.1–PE-Cy5.5 (1:2500, clone PK136), anti-MHCII–AF700 (1:2000, clone M5/114.15.2) (all Biolegend), anti-CD11b–eFluor450 (1:1000, clone M1/70), anti-GATA3–PE-Cy7 (1:100, clone L50-823), anti-CD3–biotin (1:1000, clone 145-2C11), anti-RORt–APC (1:500, clone AFKJS-9) (all eBioscience), anti-TCRβ–BB790-P (1:4000, clone H57-597), anti-CD4–BUV496 (1:500, clone GK1.5), and anti-CD19–BUV661 (1:2000, clone 1D3) (all BD Biosciences). Antibodies used for compensation controls were anti-CD25–BV421 (1:200, clone PC61), anti-CD44–BV510 (1:200, clone IM7), anti-CD3–BV650 (1:200, clone 17A2), anti-CD8–BV785 (1:2000, clone 53-6.7), anti-NK1.1–PE-Cy5.5 (1:2500, clone PK136), anti-MHCII–AF700 (1:2000, clone M5/114.15.2) (all Biolegend), anti-CD11b–eFluor450 (1:1000, clone M1/70), anti-TCRβ–FITC (1:500, clone H57-597), anti-B220–PE-Cy5 (1:200, clone RA3-6B2), anti-CD23–PE-Cy7 (1:500, clone B3B4), anti-CD8–biotin (1:200, clone 53-6.7), anti-Foxp3–APC (1:200, clone FJK-16s), anti-CD69–PE (1:200, clone H1.2F3) (all eBioscience), anti-TCRβ–BB790-P (1:4000, clone H57-597), anti-CD103–BUV395 (1:500, clone M290), anti-CD4–BUV496 (1:200, clone GK1.5), and anti-CD19–BUV661 (1:2000, clone 1D3) (all BD Biosciences). Streptavidin AF350 (1:200, Invitrogen) was used to identify biotinylated antibody. Samples were acquired on a Yeti/ZE5 flow cytometer (Propel Labs/BioRad).

ltered(1:250, 100 [illegible] Bio

, PE-4y5 (1:200, clone AFKJS-9 and anti-Biolegend,31Tf1.9461UV737antibody.

S v . In a flow cytometry system with n channels, let us consider the spillover matrix for a set of d single-color controls, that is for d dyes, with $d \le n$. We concentrate on the dye $k = 1..d$ during the following argument.

For any event in the flow cytometer, we have the following two-row vectors: the i event data x, with length d, and the i event data , with length n. On average for any level of fluorescence, true and observed events are related linearly through the $d \times n$ spillover matrix , according to

$$x = . \tag{2}$$

Classical flow cytometry systems have $n = d$, and compensation is usually achieved by inverting the spillover matrix and multiplying by the observed data . Spectral systems feature $n > d$, and compensation is usually called unmixing and is not unequivocally defined, because Eq. (2) produces an overspecified system of equations. In the following, and for simplicity, we refer to unmixing in spectral systems also as compensation.

Independently of the compensation method used, when the spillover matrix is estimated as $= + $, thus with some error , it unavoidably gives rise to incorrectly compensated data $x + e$, which verifies, on average,

$$(x + e)(+) = . \tag{3}$$

Therefore,

$$x = -e . \tag{4}$$

The vectors x and e, and the matrices , , and , have the following properties:

- Because x represents the true value of events in the single-color control for dye k, then $x_{i,k} > 0$ and $x_{i,l} = 0$, for all $l \neq k$.
- The kth row of the spillover matrix is normalized with $1 = s_{k,k} \ge s_{k,l} \ge 0$, for some $k = 1..d$ and every $l \neq k$.
- The row normalization of implies that the true value of the dye in the control, $x_{i,k}$, can always be obtained from the observed value , as Eq. (2) implies $= x_{i,k} = x_{i,k}$. Therefore, $e_{i,k} = 0$, irrespective of errors in the estimation of the spillover matrix.
- Also because of the row normalization of the spillover matrix, the estimation of the spillover coefficient $s_{k,k} = 1$ will always be exact, i.e. $s_{k,k} = 1$ and $e_{i,k} = 0$, irrespective of errors in the estimation of the spillover matrix.

Let us consider now the LHS of Eq. (4), i.e. the row vector e. Its lth coefficient, for any $l = 1..d$, equals

( )

27. Poot, M., Gibson, L. L. & Singer, V. L. Detection of apoptosis in live cells by Mito-Tracken(TM) red CMXRos and SYTO dye flow cytometry. . . **2** , 358–364 (1997).
28. Chattopadhyay, P. K., Winters, A. F., Lomas, W. E., Laino, A. S. & Woods, D. M. High-parameter single-cell analysis. . **12**, 411–430 (2019).
29. Spitzer, M. H. & Nolan, G. P. Mass cytometry: single cells, many features. **1**, 780–791 (2016).
30. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. . **14**, 865–868 (2017).
31. Shahi, P., Kim, S. C., Haliburton, J. R., Gartner, Z. J. & Abate, A. R. Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. , 44447 (2017).
32. Nolan, J. P. & Condello, D. Spectral flow cytometry. . **3**, 1.27.1–1.27.13 (2013).
33. Novo, D., Grégori, G. & Rajwa, B. Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices. . **3** , 508–520 (2013).
34. Futamura, K. et al. Novel full-spectral flow cytometry with multiple spectrally-