Spectral hotspot analysis reveals unmixing-dependent spreading

Peter Mage[1], Andrew Konecny[2], Florian Mair[3]

[1]**Advanced Technology Group, BD Biosciences**

[2]Fred Hutchinson Cancer Center

[3]ETH Zurich

Preprint:

bioRxiv

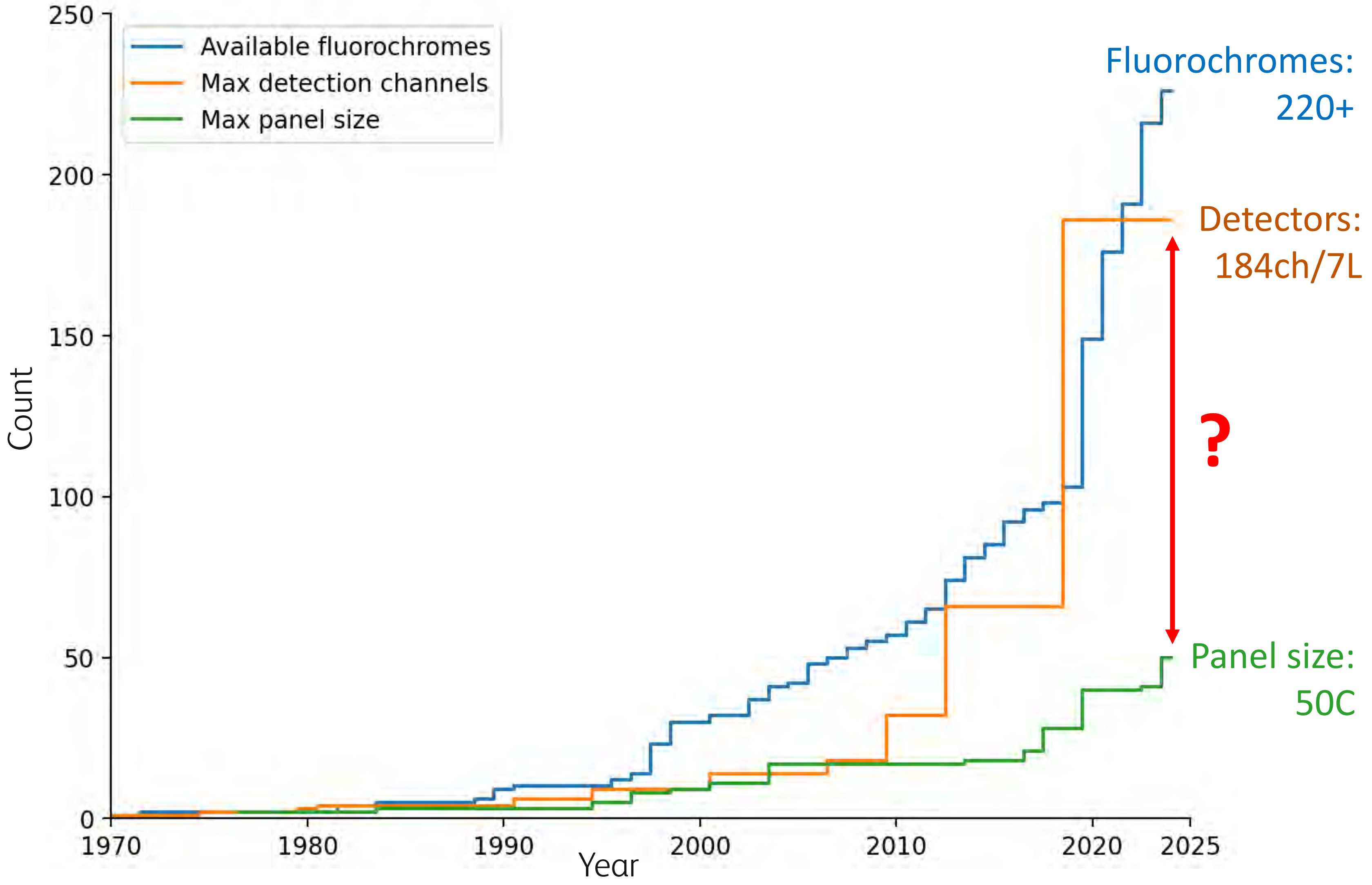
THE PREPRINT SERVER FOR BIOLOGY

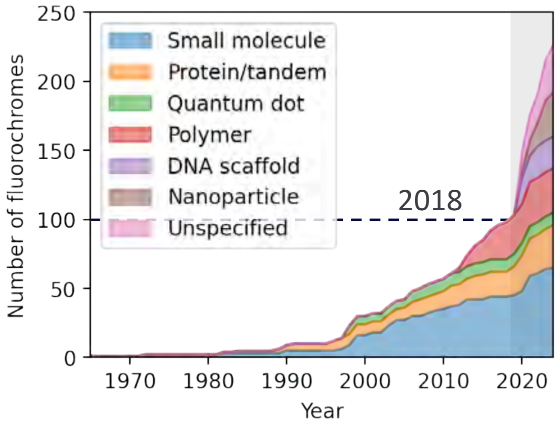
BD

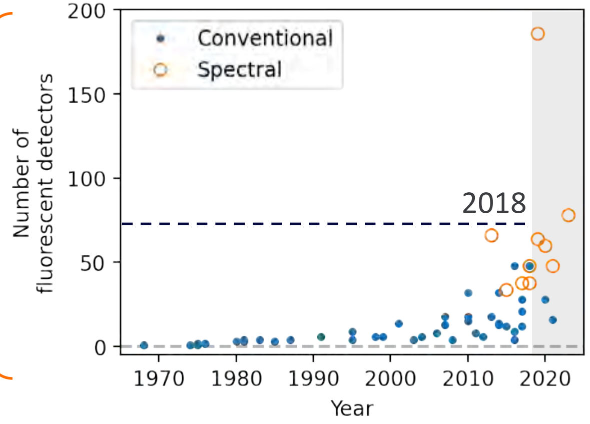
What is UDS?

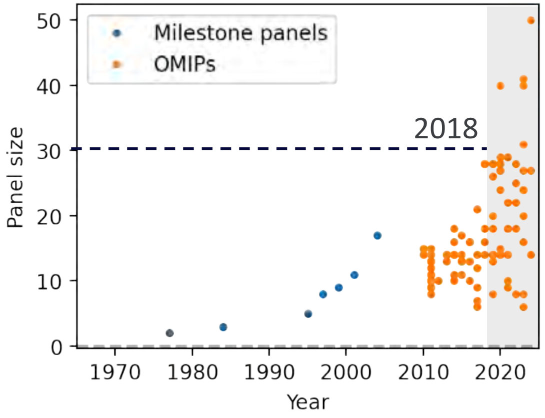
1. **What is unmixing-dependent spreading (UDS)?**
2. How can we predict UDS?
3. How can we use the Hotspot Matrix to make better panels?
4. How can we understand UDS intuitively?

Why is it hard to make a big panel?



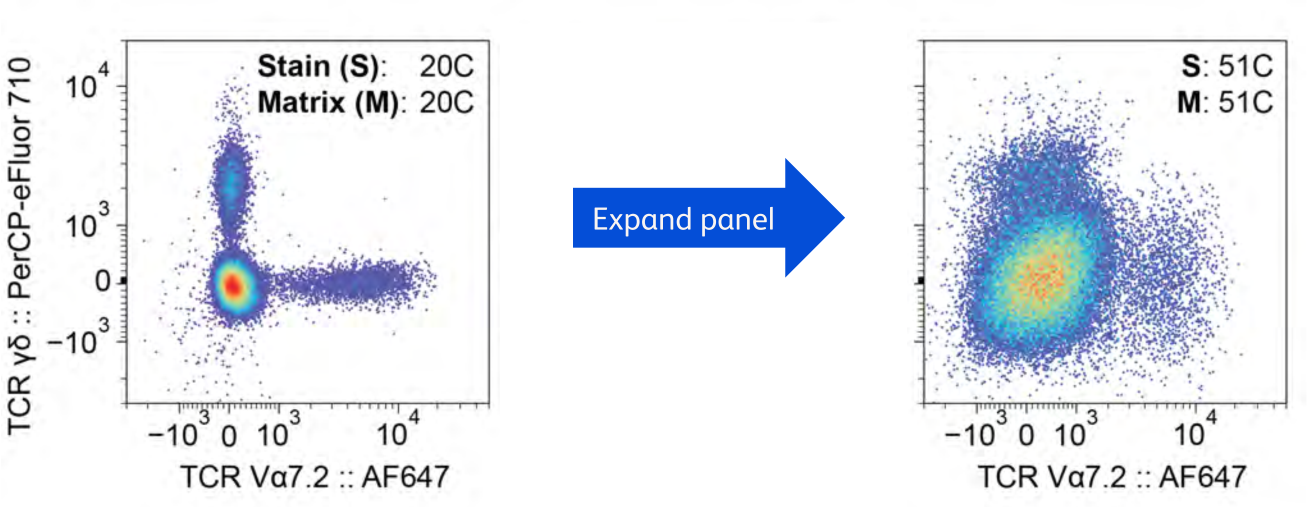






Charts indicate estimated quantities. Historical information collected from publicly available sources and internal BD records

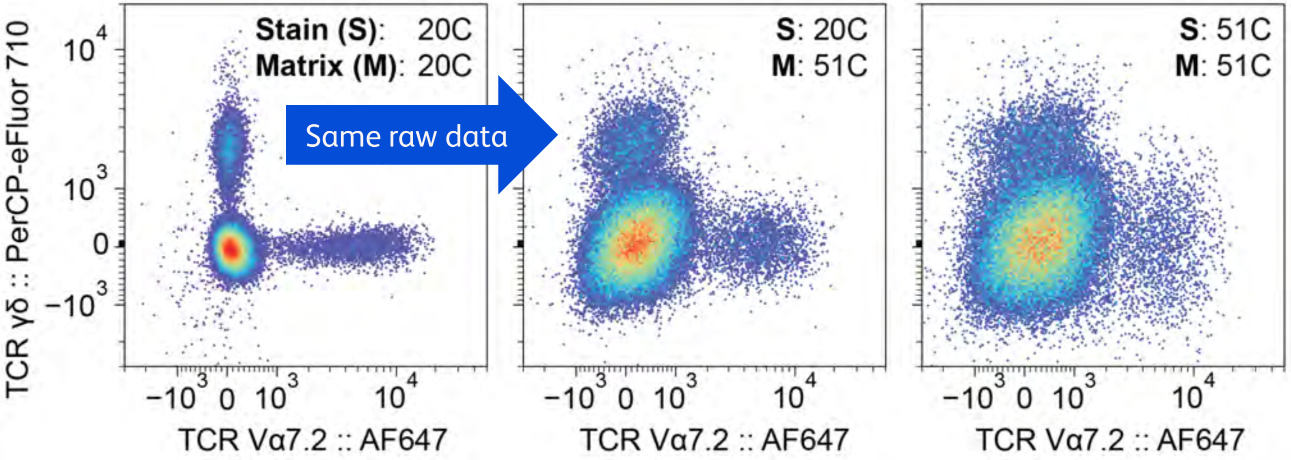
Unmixing-dependent spreading (UDS) occurs in many large spectral panels



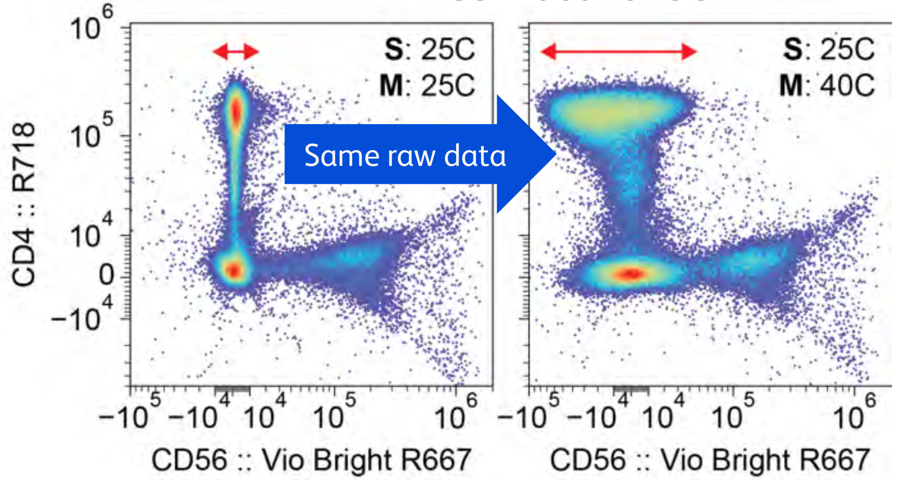
Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

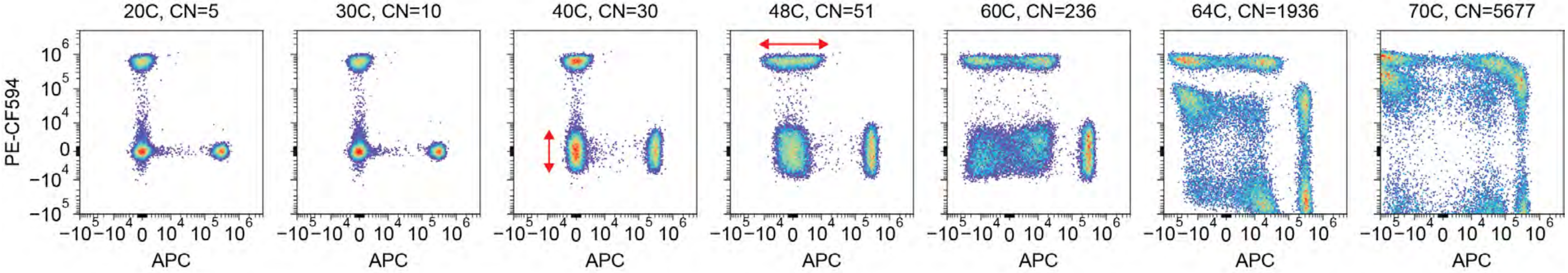
Unmixing-dependent spreading (UDS) occurs in many large spectral panels

Sony ID7000



BD FACSDiscover S8

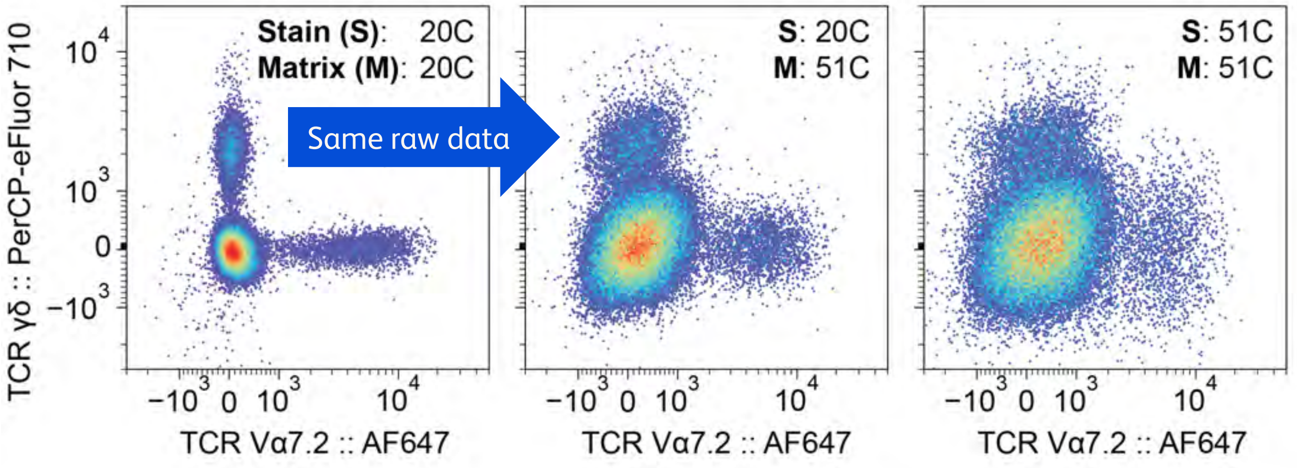




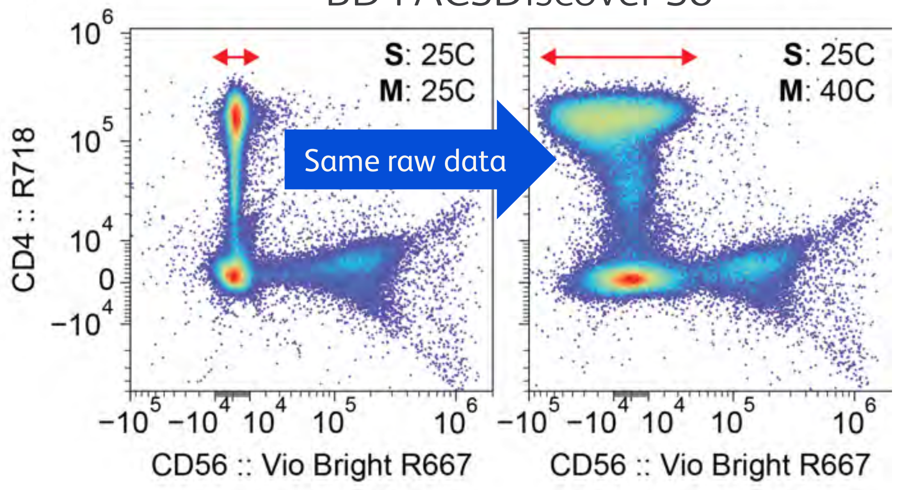
Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

Unmixing-dependent spreading (UDS) occurs in many large spectral panels

Sony ID7000



BD FACSDiscover S8



Unmixing-dependent spreading:

Change in variance (spread) of an **unmixed parameter** when the **same raw data** is unmixed using **different spectral matrices** containing:

- different set of fluorochromes and/or
- different numbers of fluorochromes

When should you be worried about UDS?

- You are making a really big panel
- You are doing something wrong by accident
- You are doing something wrong on purpose

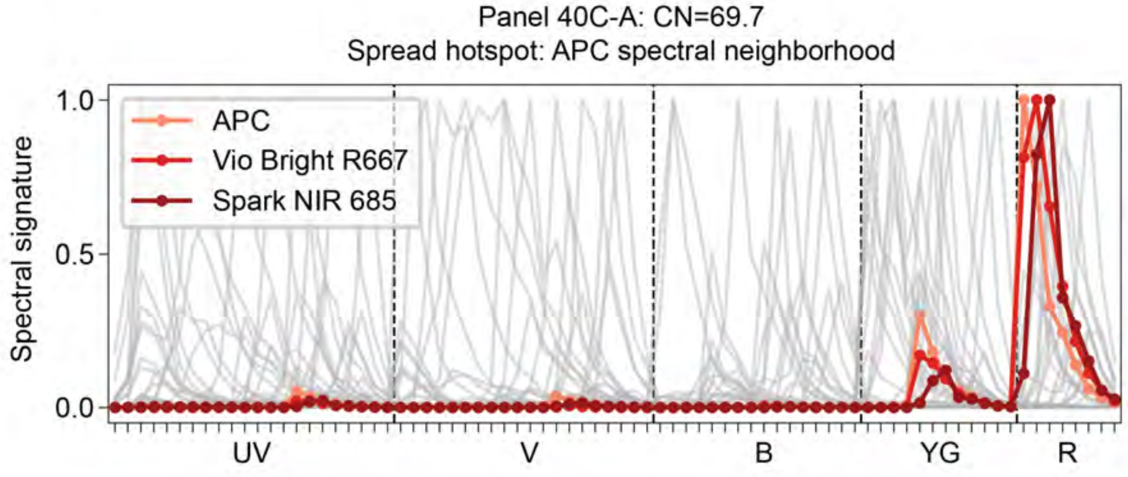
Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

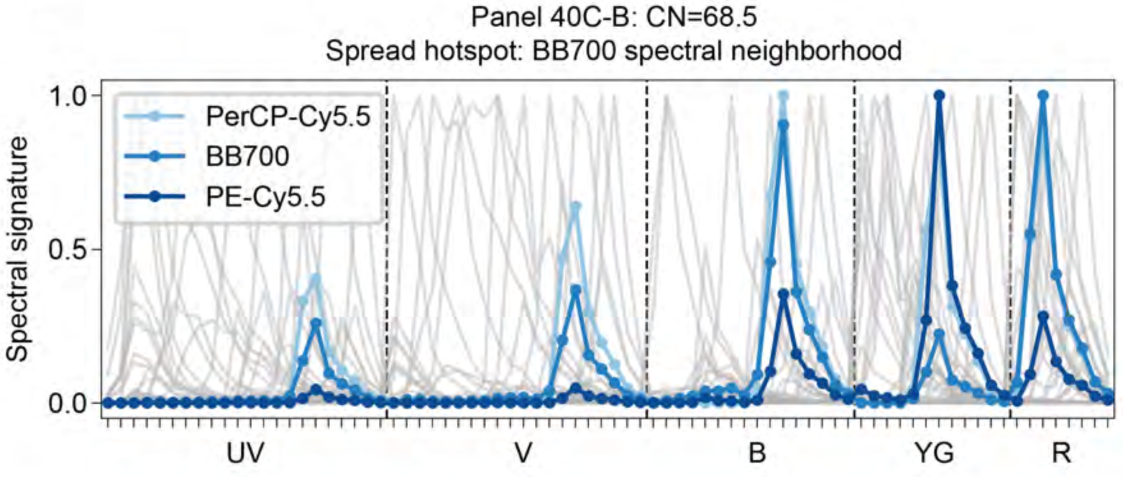
How can we predict UDS?

1. What is unmixing-dependent spreading (UDS)?
2. **How can we predict UDS?**
3. How can we use the Hotspot Matrix to make better panels?
4. How can we understand UDS intuitively?

UDS tends to occur in spectral "hotspots"

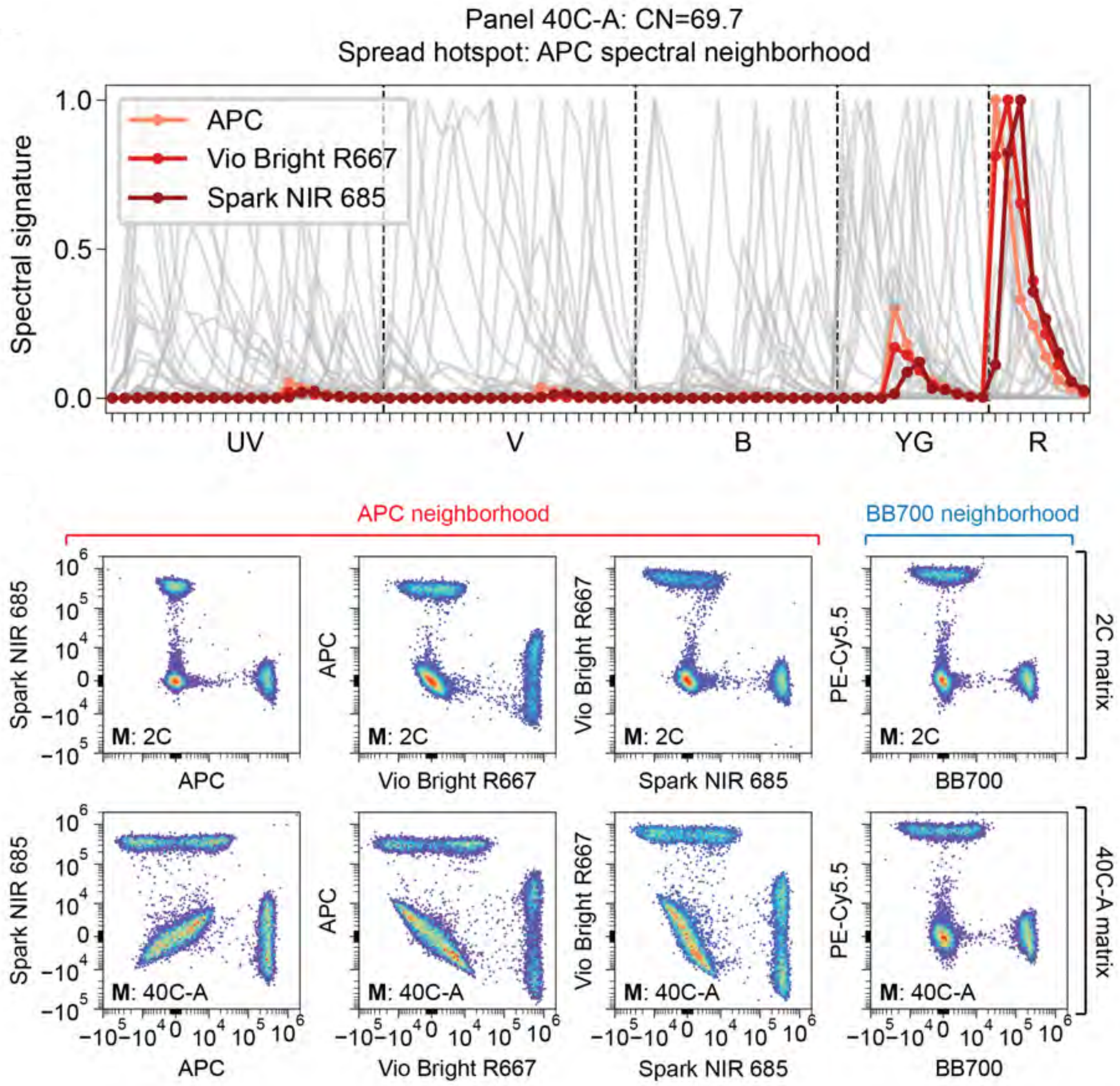
Let's take a reasonable 39C panel and try to add a color:

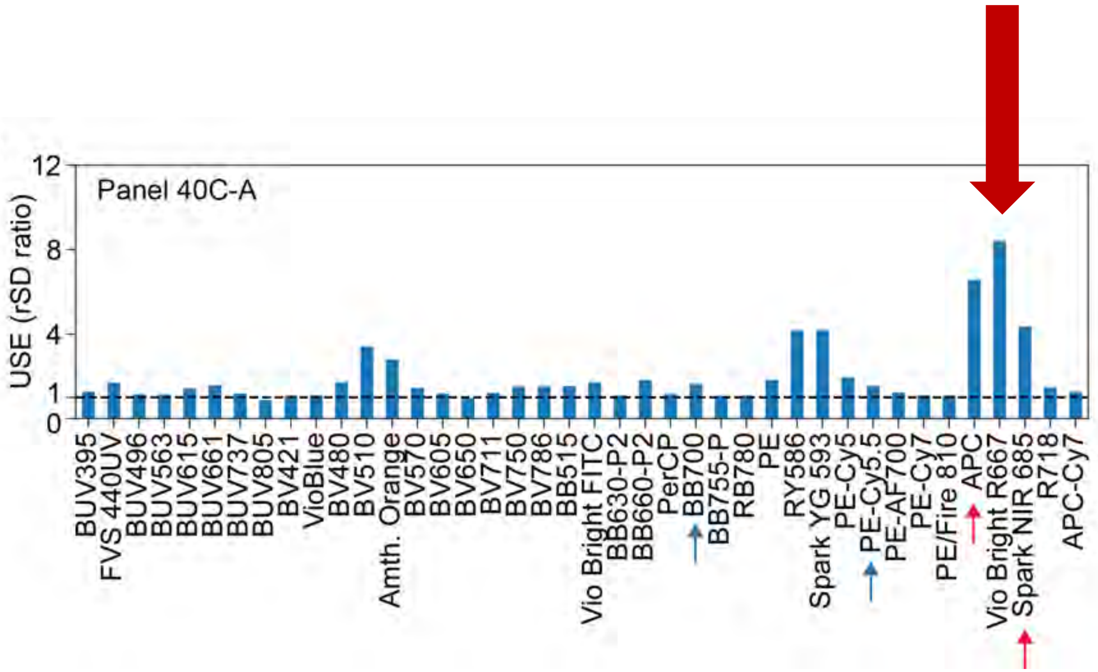




Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

UDS tends to occur in spectral "hotspots"



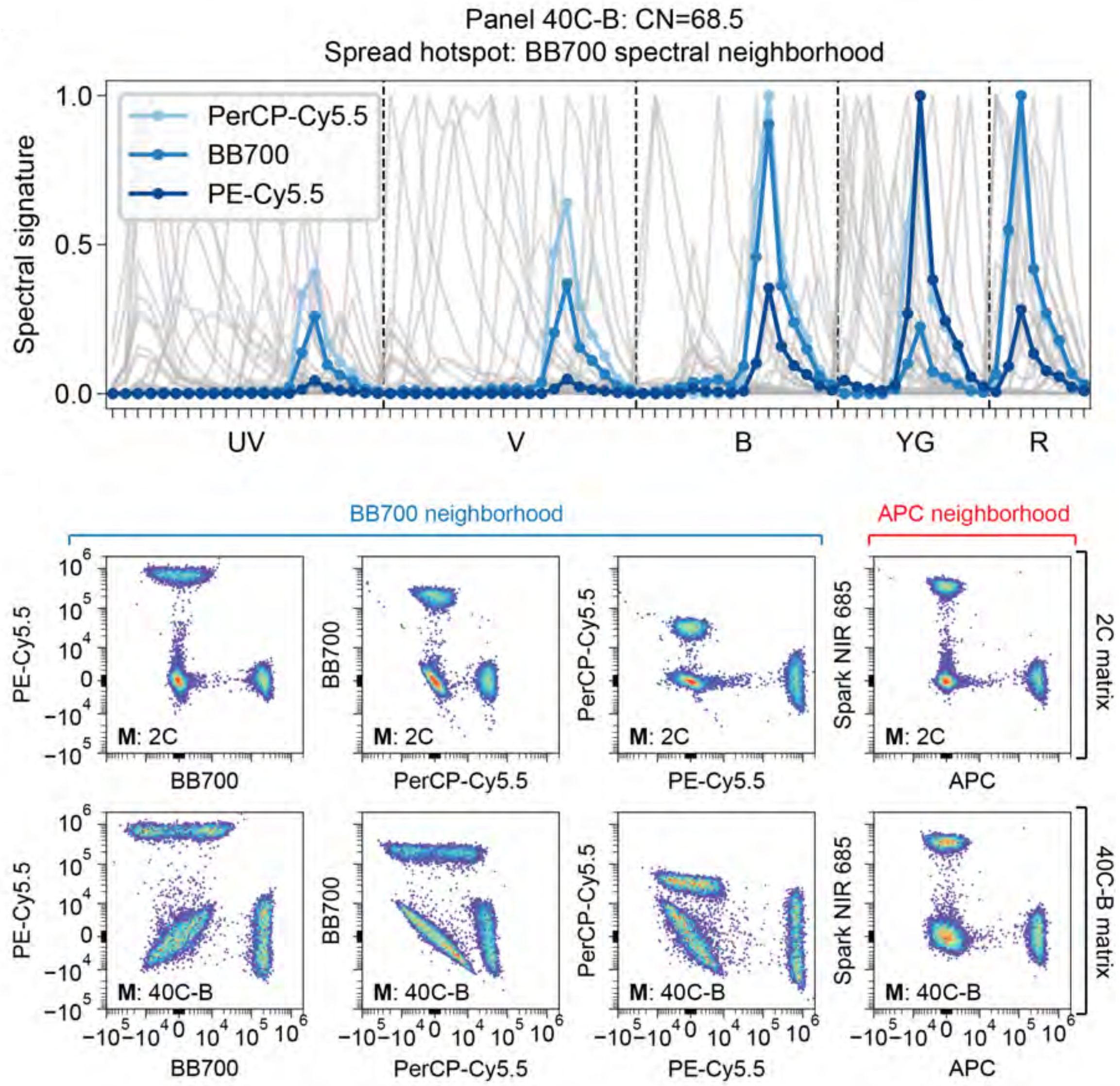


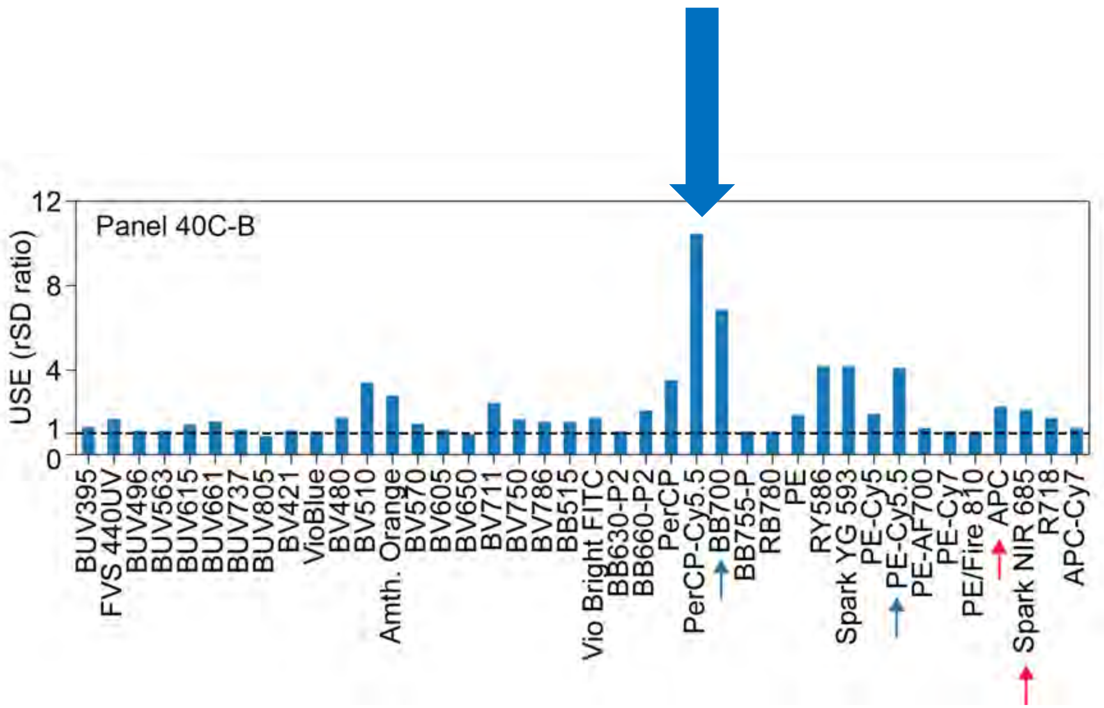
Unmixing spreading error (USE):

Ratio of unmixed rSD, 40C unmixing vs 1C unmixing

Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

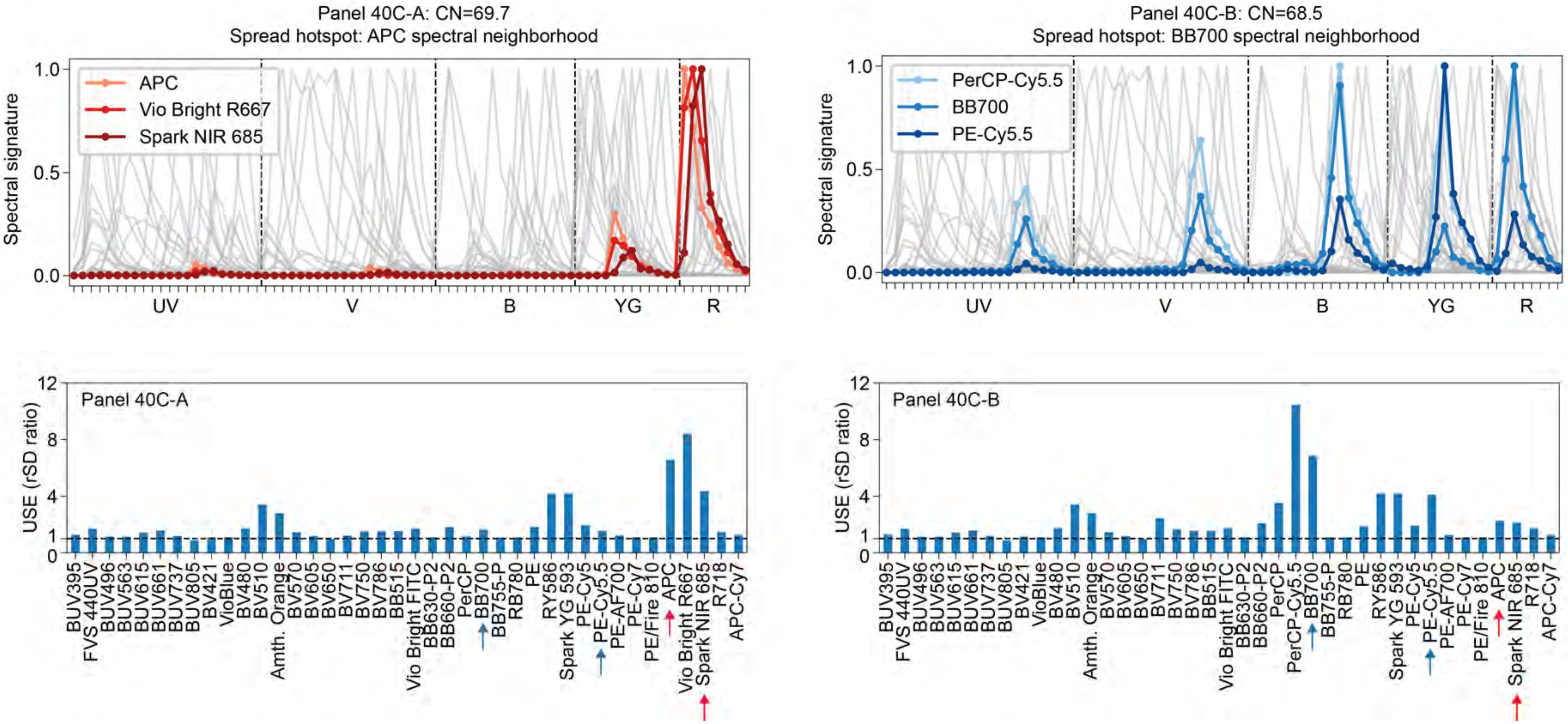
UDS tends to occur in spectral “hotspots”





Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

UDS tends to occur in spectral “hotspots”

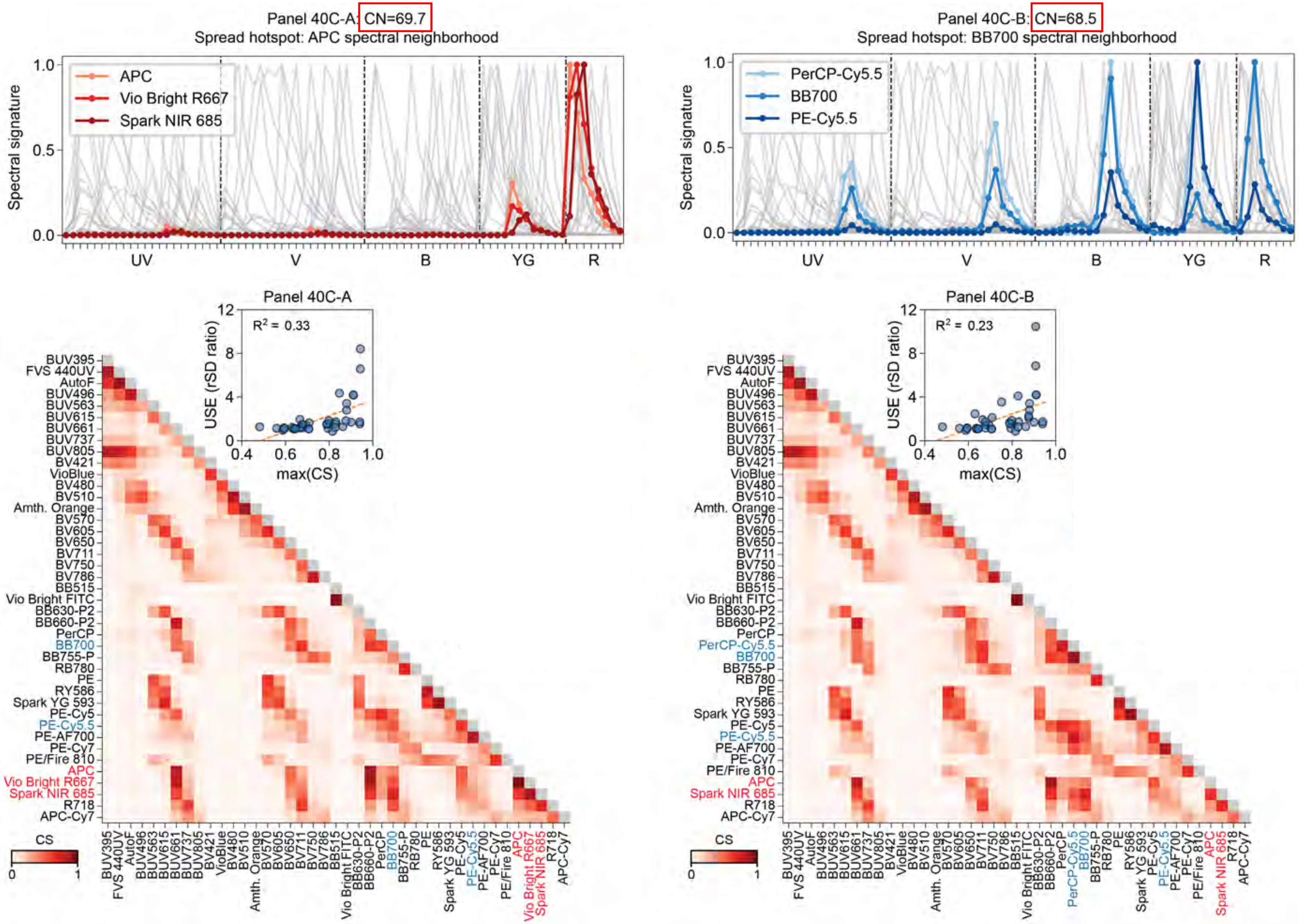


Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

UDS is difficult to predict with cosine similarity and condition number (complexity) alone

Cosine similarity tells us which pairs might have a problem – but not which ones actually do

Condition number (complexity) tells us if a problem exists, but not where it is

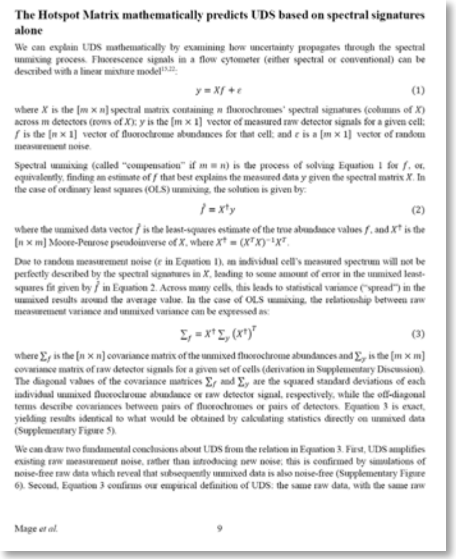


Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

The Hotspot Matrix directly predicts UDS

- Based on predicted covariance matrix
- Math available in the preprint:

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY



The Hotspot Matrix mathematically predicts UDS based on spectral signatures alone

We can explain UDS mathematically by examining how uncertainty propagates through the spectral unmixing process. Fluorescence signals in a flow cytometer (either spectral or conventional) can be described with a linear mixture model[13,22]:

$$y = Xf + \varepsilon \quad (1)$$

where X is the $[m \times n]$ spectral matrix containing n fluorochromes' spectral signatures (columns of X) across m detectors (rows of X); y is the $[m \times 1]$ vector of measured raw detector signals for a given cell; f is the $[n \times 1]$ vector of fluorochrome abundances for that cell; and ε is a $[m \times 1]$ vector of random measurement noise.

Spectral unmixing (called "compensation" if $m = n$) is the process of solving Equation 1 for f, or, equivalently, finding an estimate of f that best explains the measured data y given the spectral matrix X. In the case of ordinary least squares (OLS) unmixing, the solution is given by:

$$\hat{f} = X^{\dagger} y \quad (2)$$

where the unmixed data vector $\hat{f}$ is the least-squares estimate of the true abundance values f, and $X^{\dagger}$ is the $[n \times m]$ Moore-Penrose pseudoinverse of X, where $X^{\dagger} = (X^T X)^{-1} X^T$.

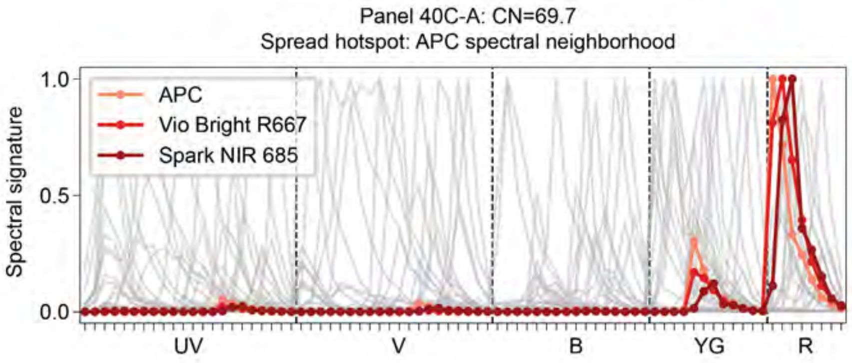
Due to random measurement noise (ε in Equation 1), an individual cell's measured spectrum will not be perfectly described by the spectral signatures in X, leading to some amount of error in the unmixed least-squares fit given by $\hat{f}$ in Equation 2. Across many cells, this leads to statistical variance ("spread") in the unmixed results around the average value. In the case of OLS unmixing, the relationship between raw measurement variance and unmixed variance can be expressed as:

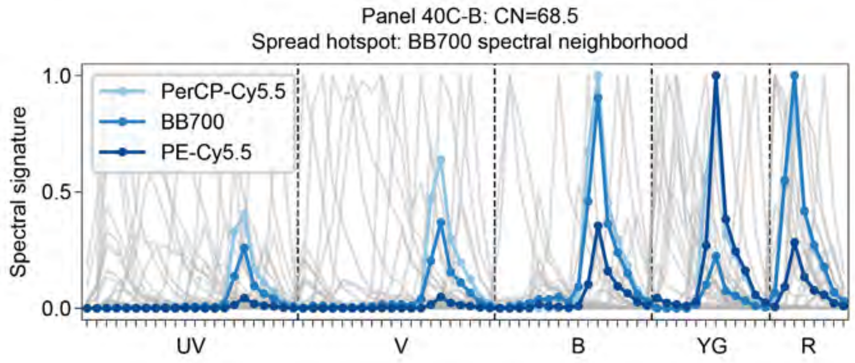
$$\Sigma_f = X^{\dagger} \Sigma_y (X^{\dagger})^T \quad (3)$$

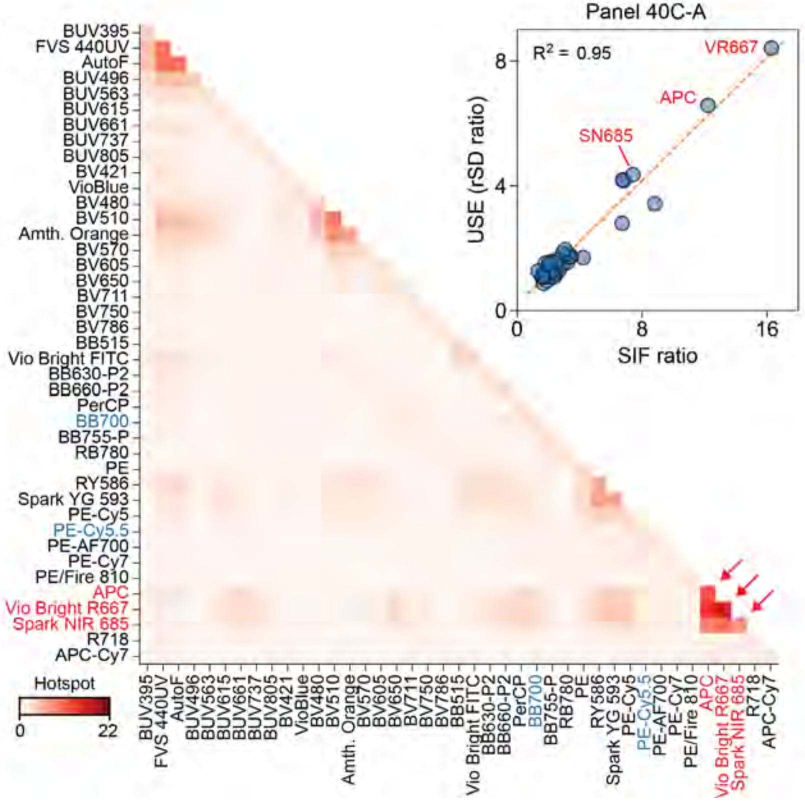
where Σ_f is the $[n \times n]$ covariance matrix of the unmixed fluorochrome abundances and Σ_y is the $[m \times m]$ covariance matrix of raw detector signals for a given set of cells (derivation in Supplementary Discussion). The diagonal values of the covariance matrices Σ_f and Σ_y are the squared standard deviations of each individual unmixed fluorochrome abundance or raw detector signal, respectively, while the off-diagonal terms describe covariances between pairs of fluorochromes or pairs of detectors. Equation 3 is exact, yielding results identical to what would be obtained by calculating statistics directly on unmixed data (Supplementary Figure 5).

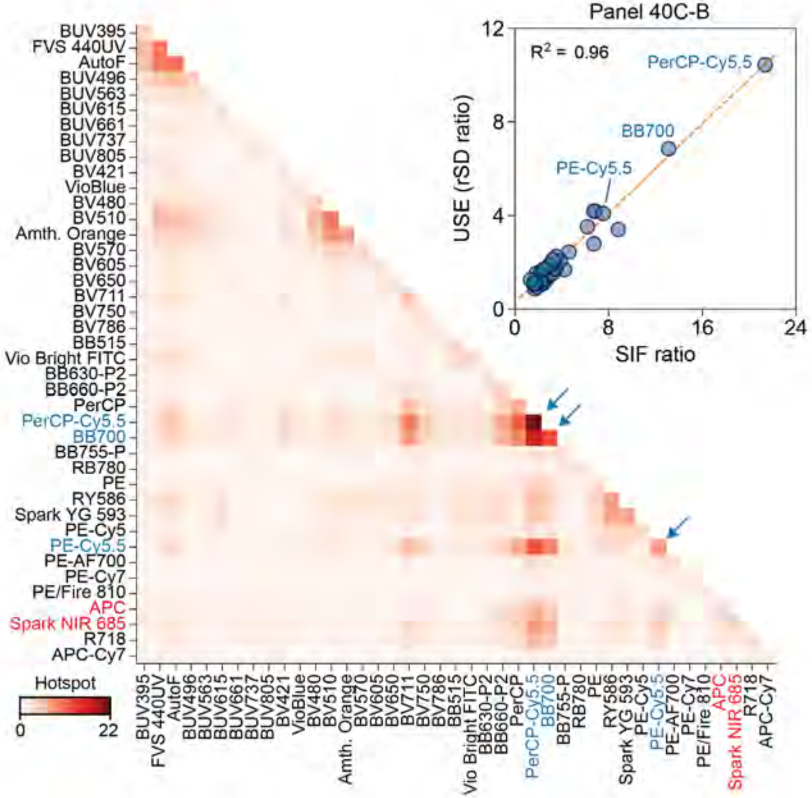
We can draw two fundamental conclusions about UDS from the relation in Equation 3. First, UDS amplifies existing raw measurement noise, rather than introducing new noise; this is confirmed by simulations of noise-free raw data which reveal that subsequently unmixed data is also noise-free (Supplementary Figure 6). Second, Equation 3 confirms our empirical definition of UDS: the same raw data, with the same raw

Mage *et al.* 9









Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

UDS has been seen before: variance inflation and collinearity in multiple regression

Spectral collinearity is the underlying cause of UDS

We conclude with an explanation of the underlying mathematical cause of UDS in terms of the statistical concepts of "variance inflation" and "collinearity." Variance inflation occurs in multiple regression when highly correlated predictor variables lead to increased uncertainty in regression parameter estimates[18,25]. Descriptions of variance inflation in the statistics literature are remarkably similar to the empirical hallmarks of UDS, such as higher-than-expected variance in regression results when using correlated predictor variables[18] (Hallmark 1) and high covariance between regression coefficients leading to visible "tilt" in bivariate plots[28,29] (Hallmark 2). Variance inflation is known to be caused by predictor variables with high collinearity, a term describing the degree of intercorrelation among a set of variables or vectors[18].

Approaches for diagnosing collinearity and variance inflation can be applied to the problem of UDS by framing spectral unmixing as a multiple regression problem. In this interpretation, unmixing is the process of fitting a linear regression model that explains a cell's measured raw signal y (the dependent "response variable" or "regressand") across m detectors ("observations") in terms of n fluorochrome spectra (independent "predictor/explanatory variables" or "regressors", summarized in matrix X). Each individual cell's estimated fluorochrome abundances $\hat{f}$ are therefore given by the regression "parameters" or "coefficients" that best fit the model to that cell's raw signal. While this statistical interpretation of unmixing

Mage *et al.* 11

is mathematically identical to the linear mixture model we described previously, it allows a new understanding of UDS as a process of variance inflation arising from collinear combinations of spectra. Just like practitioners of multiple regression work to prevent problematic variance inflation by avoiding excessive correlation among predictor variables[18,25,30], the spectral cytometrist seeks to avoid or mitigate UDS arising from collinearity among spectra in their panel.

Table of unmixing metrics, mathematical definitions, and statistical references

Metric / matrix name	Alternative names	Fluorochrome scope	Panel-specific	Data required	Cytometry description	Mathematical description	References
Cosine similarity (CS) / similarity matrix	"similarity index," "similarity scores," "similarity", "uncentered correlation matrix"	Pairwise	No	Spectral signatures only	Metric describing the degree of spectral overlap or similarity between two fluorochromes, based on their normalized spectral signatures. Spectra with CS=0 have no overlap, while spectra with CS=1 are spectrally identical.	Cosine of the angle between two fluorochromes' spectral signatures, calculated as the dot product of their signatures normalized to unit length. Equivalent to an uncentered correlation coefficient. Spans [0,1] for non-negative spectral signatures.	Kotu 2019[17], Park 2020[14]
Condition number (CN)	"complexity index," "complexity score", "complexity"	Full panel	Yes	Spectral signatures only	Single metric describing the overall difficulty of unmixing a specific combination of fluorochromes. Smaller numbers are better.	Describes the numerical sensitivity of the unmixing problem, *i.e.*, how much the unmixed result will change for a small perturbation in the input. Calculated as the ratio of the maximum to the minimum singular value of the spectral matrix. Spans [1, ∞].	Belsley 1980[18], Park 2020[14]
Hotspot Matrix	-	Pairwise and single-color	Yes	Spectral signatures only	Matrix describing the extent to which different fluorochromes are affected by UDS, and which combinations cause UDS.	Inverse of similarity matrix, or product of spectral matrix pseudoinverse with its transpose matrix. Diagonal spans [1, ∞]. Off-diagonal spans [0, ∞].	This work
Spreading Inflation Factor (SIF)	"collinearity index"[24], "standard-error inflation factor"[23]	Single-color	Yes	Spectral signatures only	Predicted increase in spread (rSD) of an unmixed fluorochrome in the context of a specific panel's matrix.	Diagonal of hotspot matrix; square root of Variance Inflation Factor. Spans [1, ∞].	This work, Stewart 1987[24], Fox 1992[23]
Spreading error (SE)	-	Single-color	Yes	Unmixed data	General term describing measured variance, or "spread," in unmixed or compensated data.	rSD or related metric describing an unmixed parameter corresponding to a specific population.	Roederer 2001[35]
Unmixing spreading error (USE)	-	Single-color	Yes	Unmixed data	An increase in spread of an unmixed parameter across all cells when unmixed with a full-panel matrix, regardless of marker expression.	Ratio of an unstained population's rSD unmixed with a full panel matrix to its rSD unmixed with a matrix containing just the single color and AF.	Konecny 2024[4], this work
Spillover spreading error (SSE) / Spillover spreading matrix (SSM)	-	Pairwise	Yes	Unmixed data	Increase in the spread of one fluorochrome's unmixed signal due to emission from a second fluorochrome that spectrally overlaps the first, normalized to the intensity of the second fluorochrome "causing" spread.	Square root of the difference in unmixed variance for one fluorochrome between a positive and a negative population for a second fluorochrome, divided by the difference in MFI of the second fluorochrome between both populations.	Nguyen 2013[8]
Total spreading error (TSE) / Total spreading matrix (TSM)	-	Pairwise	Yes	Unmixed data	Same as SSE, but not normalized to the intensity of the fluorochrome "causing" spread.	Same as SSE, but not normalized to the delta-MFI of the positive fluorochrome.	Corselli 2023[36]

Table 1: Summary of published spectral panel design metrics related to unmixing-dependent spreading (UDS). The "Fluorochrome scope" column indicates whether the metric refers to single fluorochromes, pairs of fluorochromes, or full panels of fluorochromes. The "Panel-specific" column indicates whether or not the value of the metric depends on the total panel of fluorochromes in the spectral matrix. The "Data required" column indicates what input data is needed in order to calculate the metric.

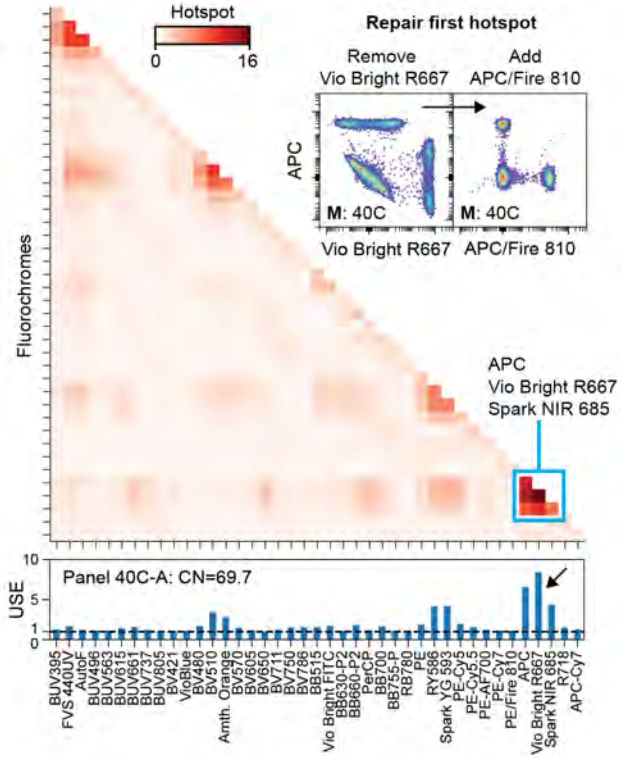
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

How can we use the Hotspot Matrix to make better panels?

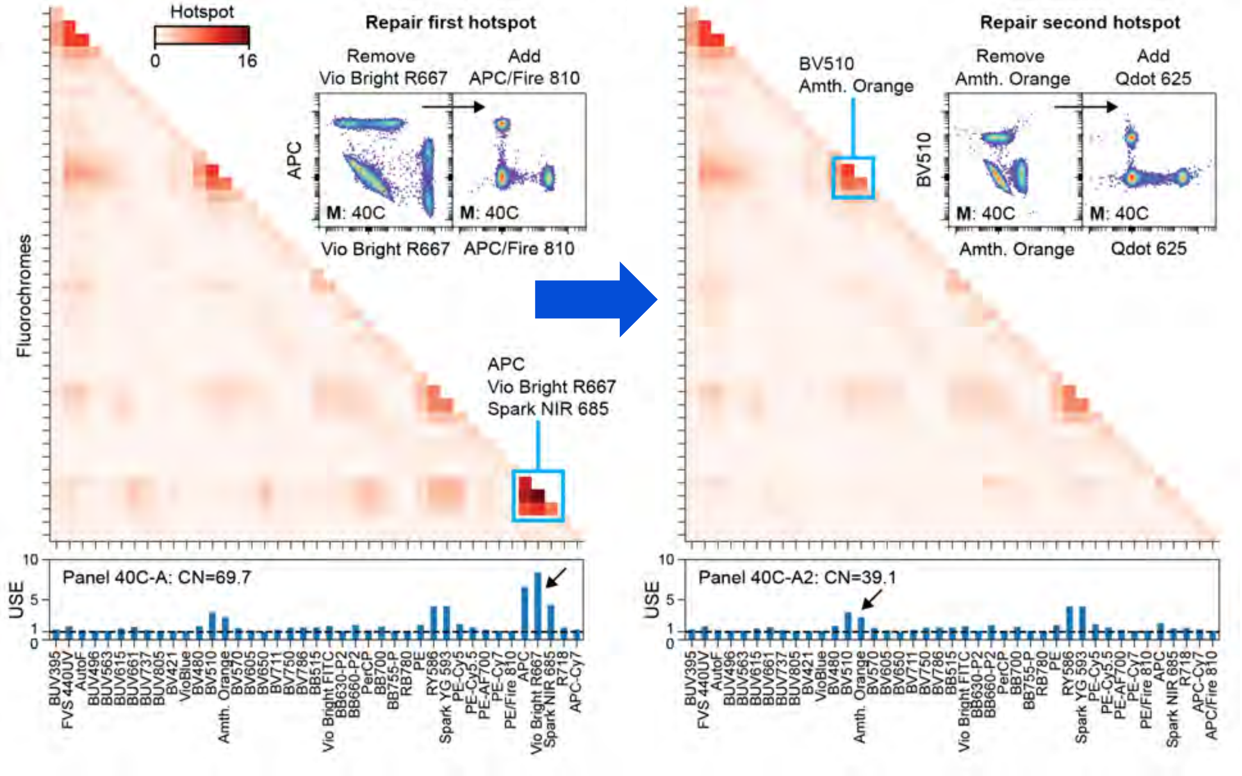
1. What is unmixing-dependent spreading (UDS)?
2. How can we predict UDS?
3. **How can we use the Hotspot Matrix to make better panels?**
4. How can we understand UDS intuitively?

Iterative improvement and expansion with Hotspot Analysis



Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

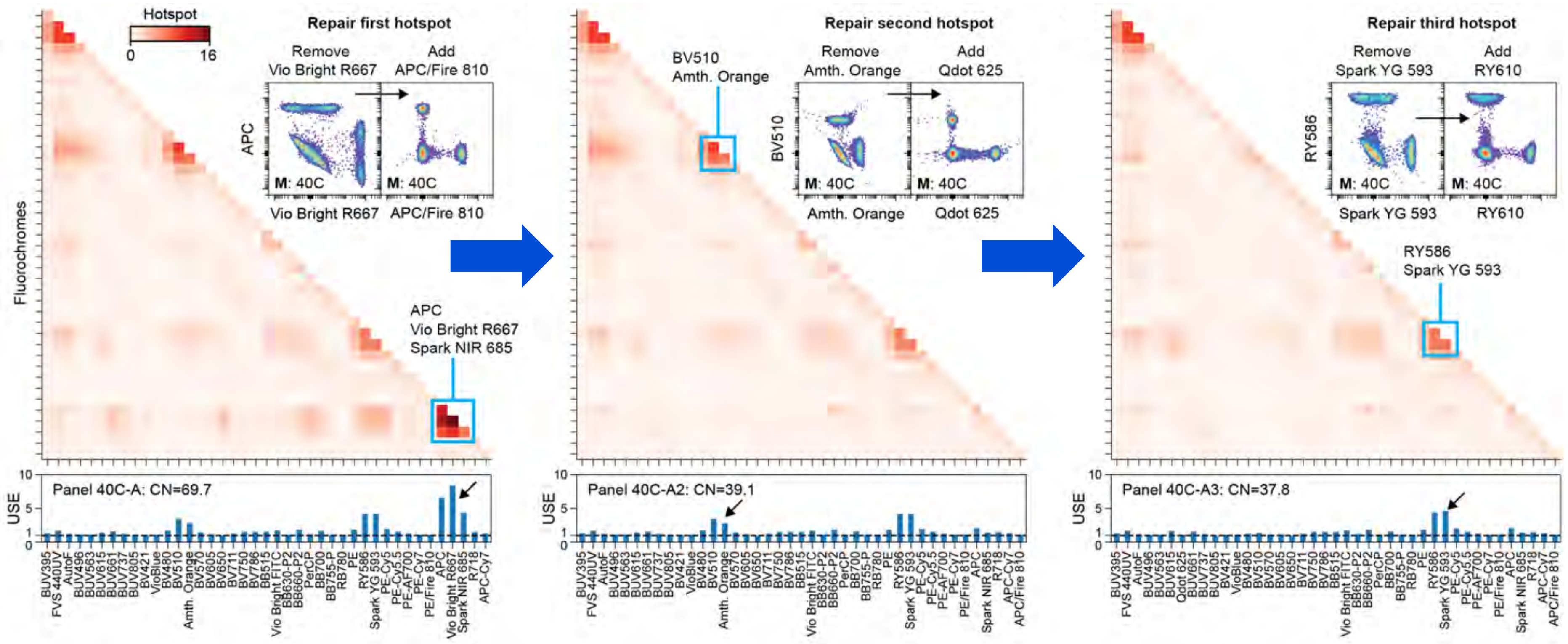
Iterative improvement and expansion with Hotspot Analysis



Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

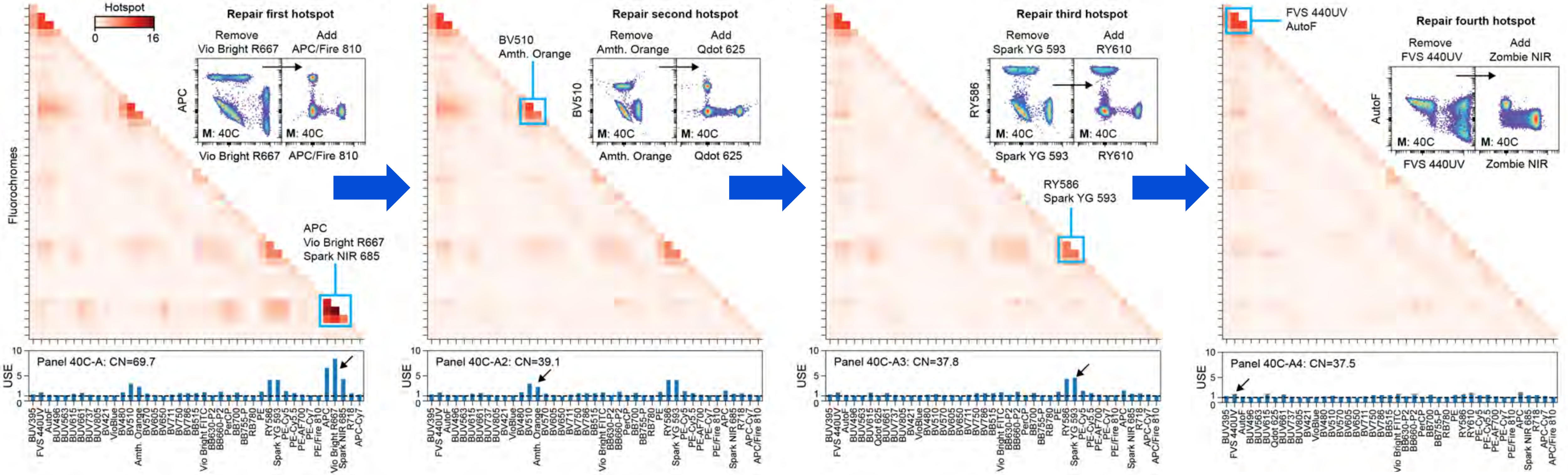
For research use only.
Not for use in diagnostic or therapeutic procedures.

Iterative improvement and expansion with Hotspot Analysis



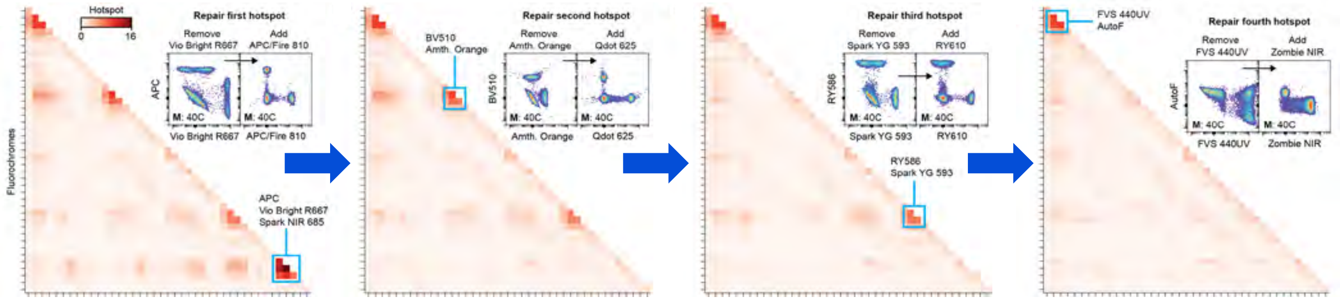
Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

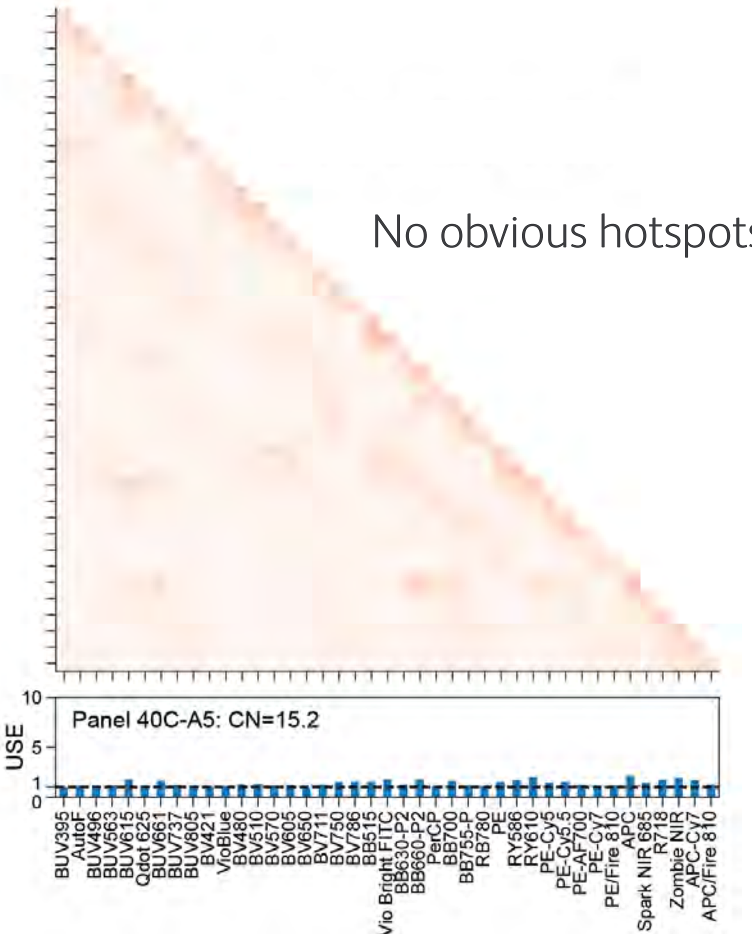
Iterative improvement and expansion with Hotspot Analysis



Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

Iterative improvement and expansion with Hotspot Analysis

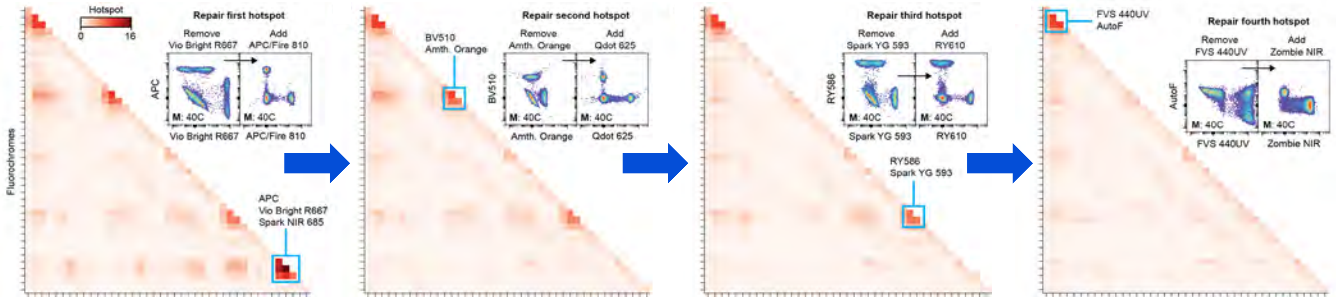


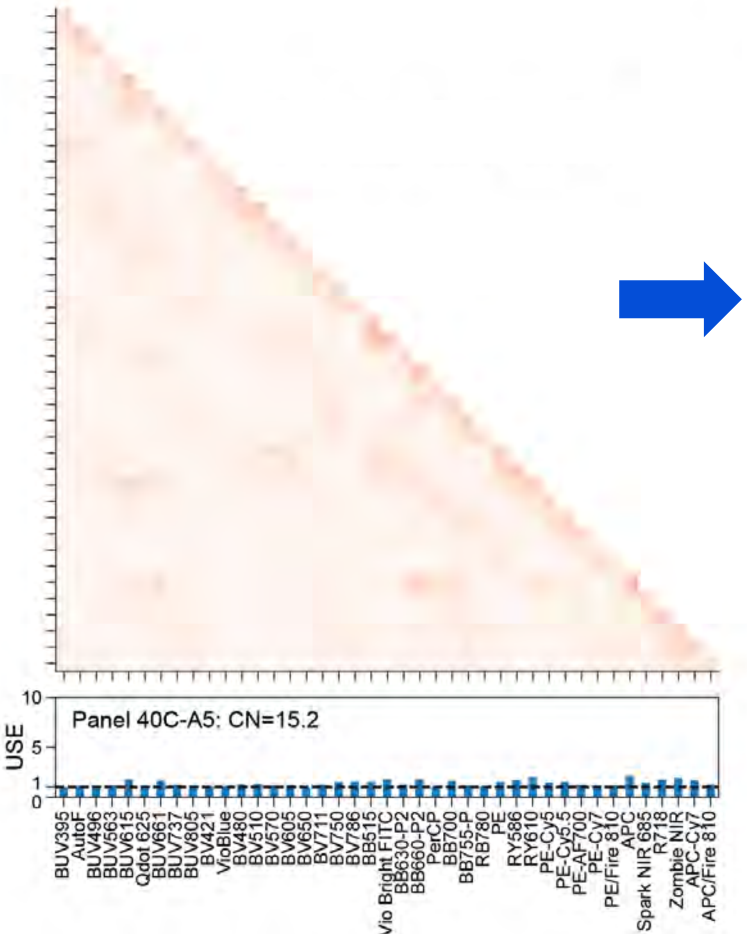


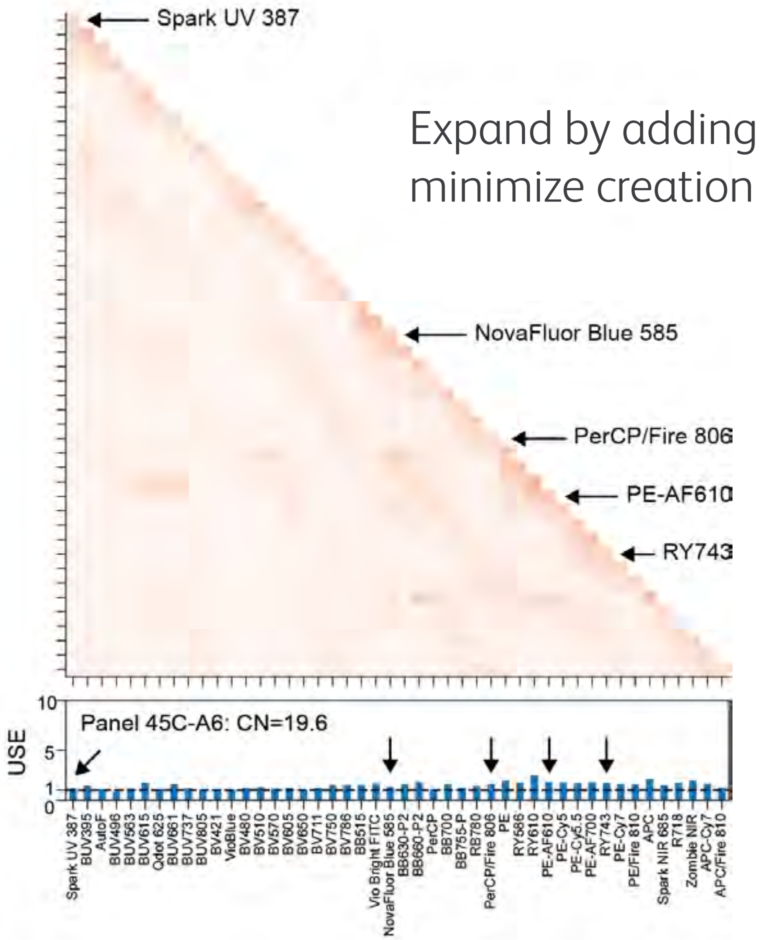
No obvious hotspots remain

Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

Iterative improvement and expansion with Hotspot Analysis







Expand by adding 5 colors that minimize creation of new hotspots

Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

How can we understand UDS intuitively?

i.e., why does it happen in the first place?

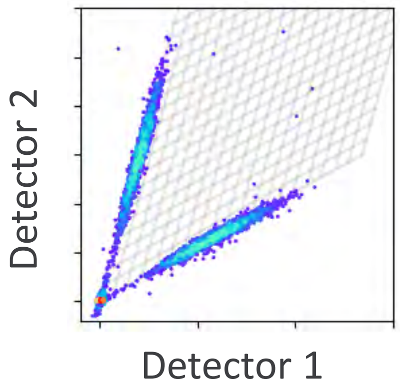
1. What is unmixing-dependent spreading (UDS)?
2. How can we predict UDS?
3. How can we use the Hotspot Matrix to make better panels?
4. **How can we understand UDS intuitively?**

Collinearity in 2D – Too much stretching “stretches” out our noise

Unmixing stretches space to
decorrelate data

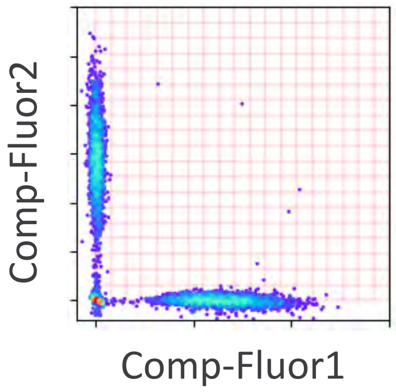
“Detector space”

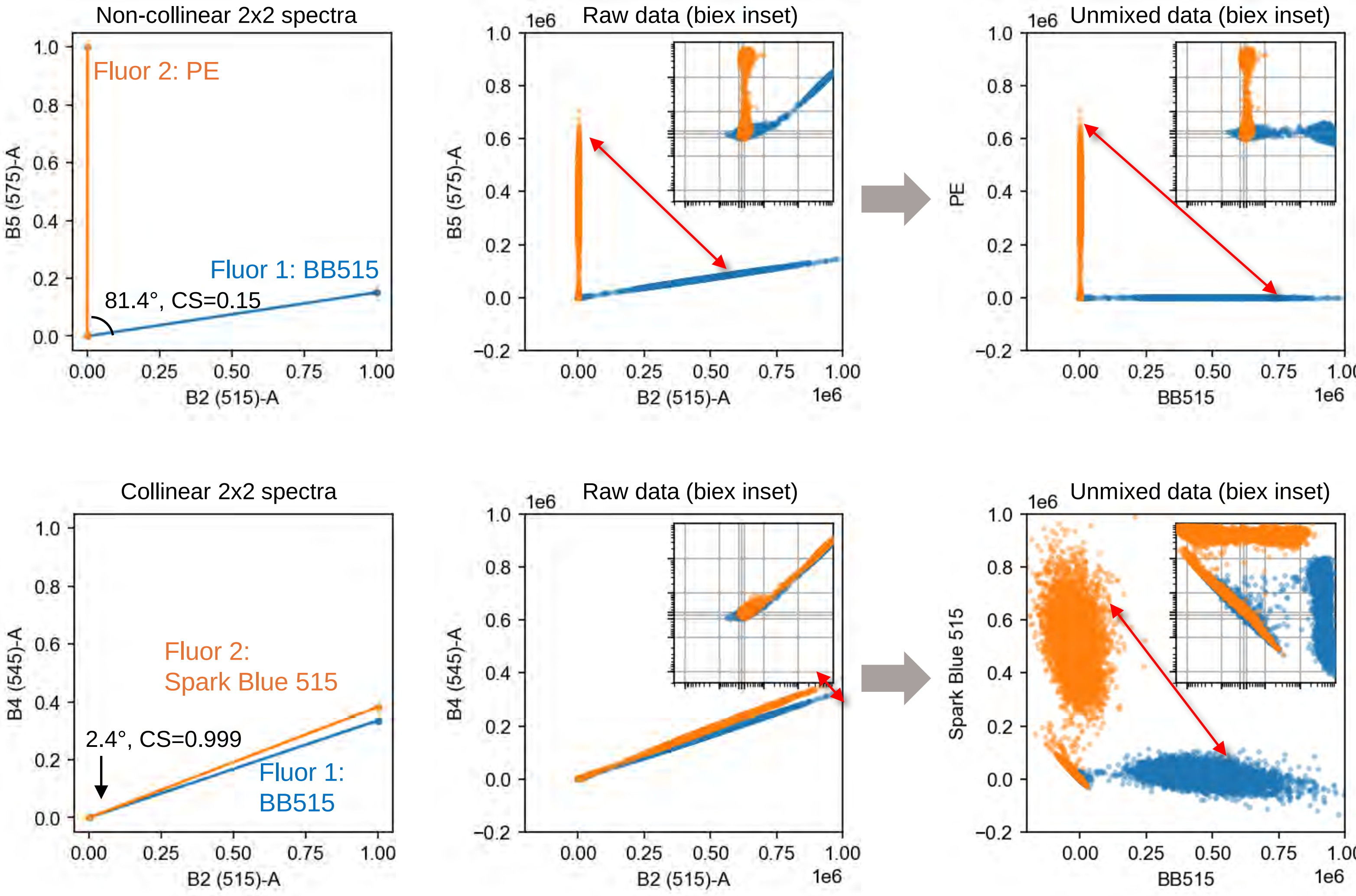
$$y = Xf$$



“Fluor space”

$$\hat{f} = X^{-1}y$$



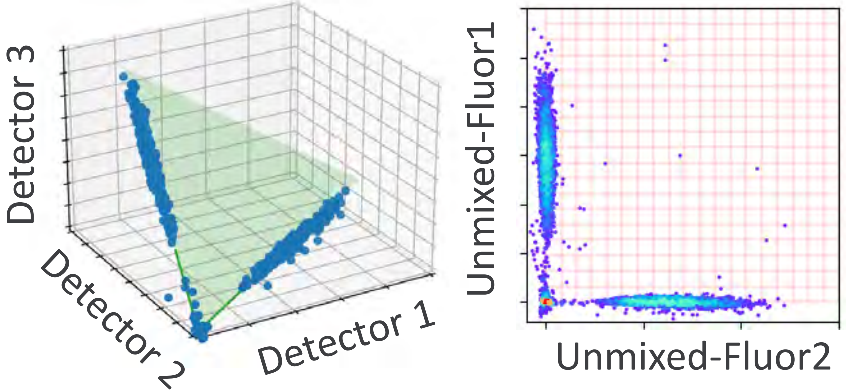


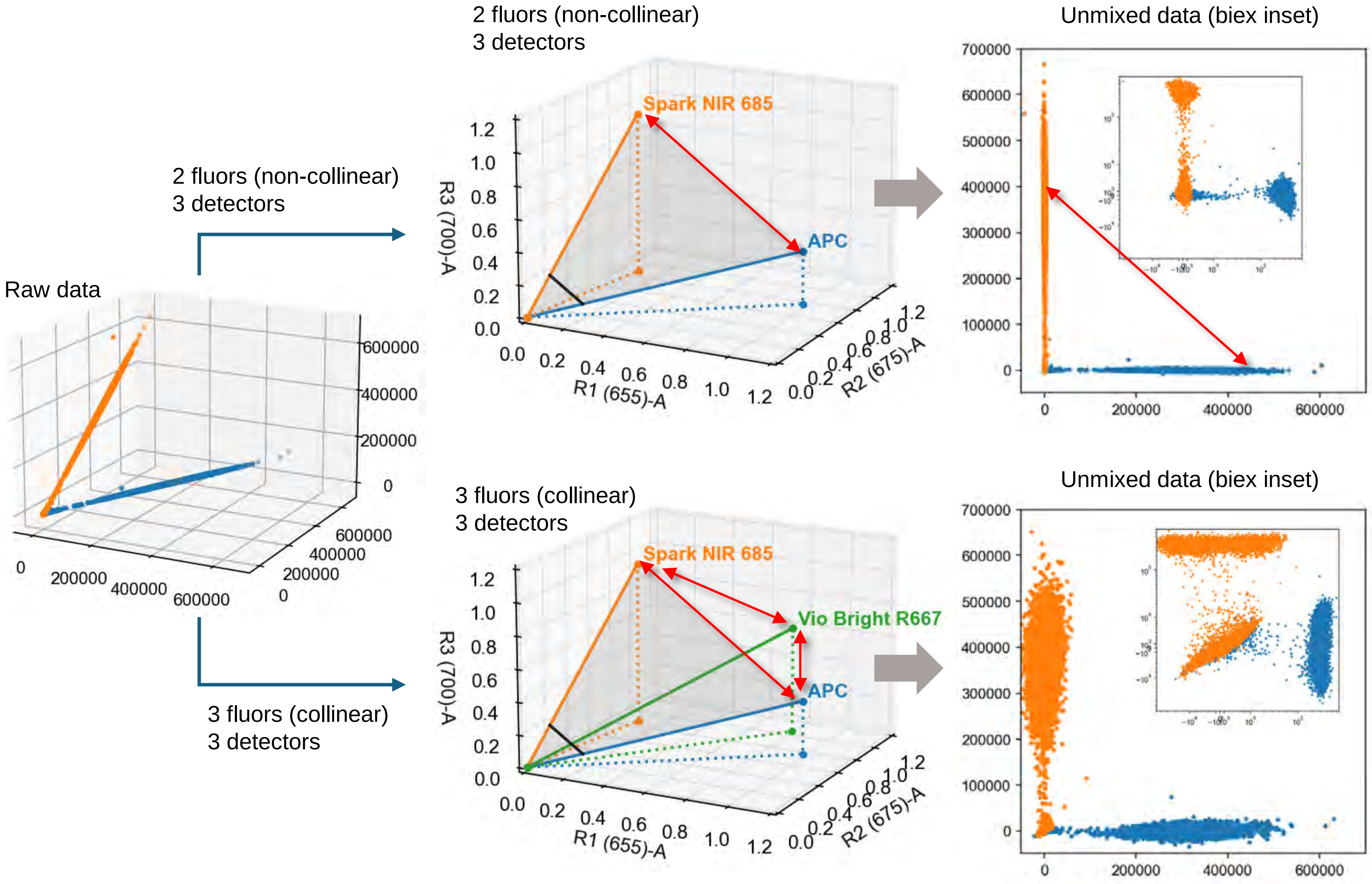
Collinearity in 3D – Too much stretching “stretches” out our noise

Unmixing stretches space to decorrelate data

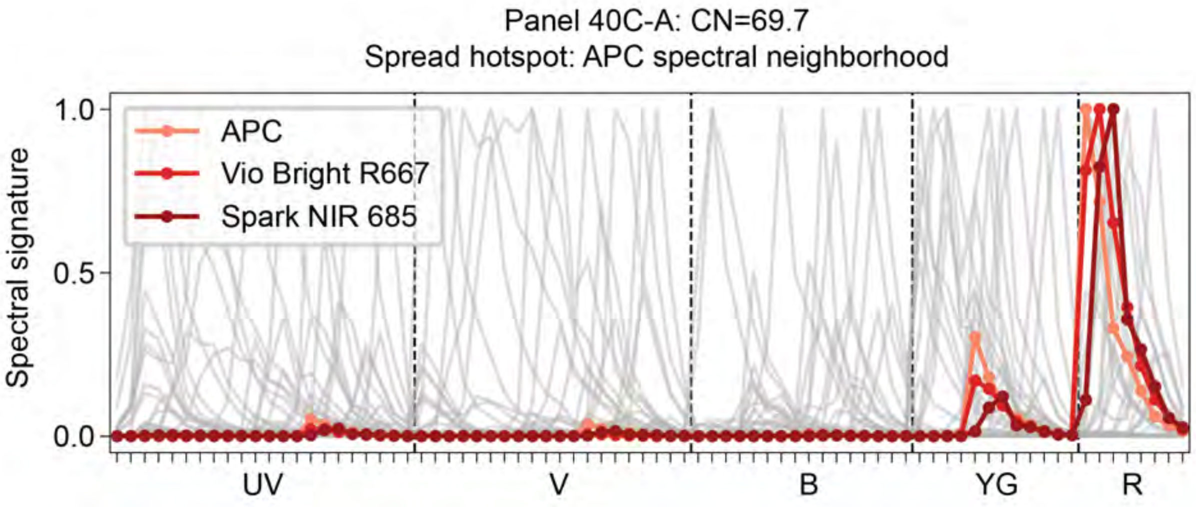
“Detector space” $y = Xf$

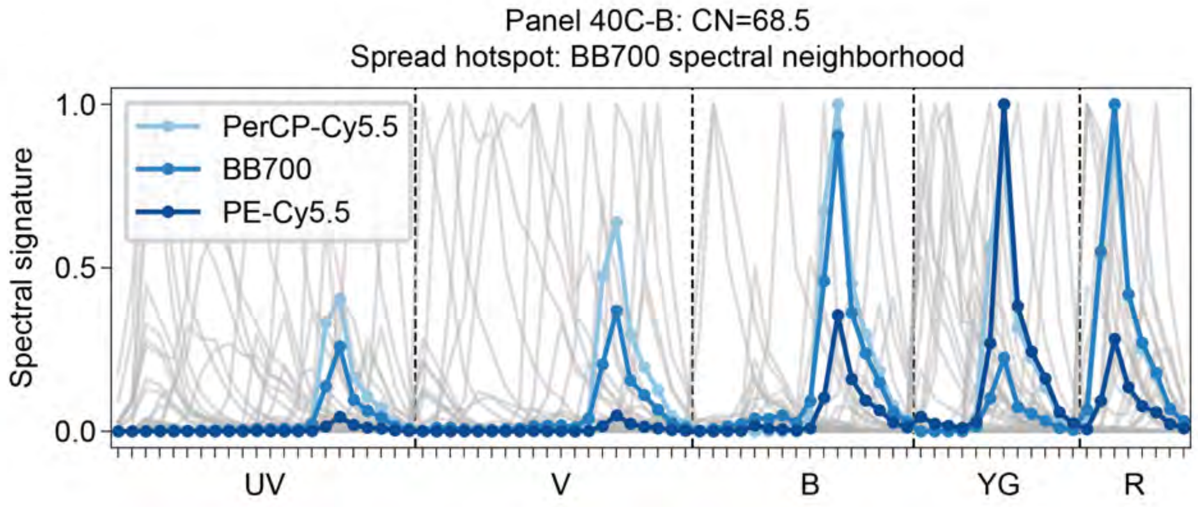
“Fluor space” $\hat{f} = X^{\dagger} y$



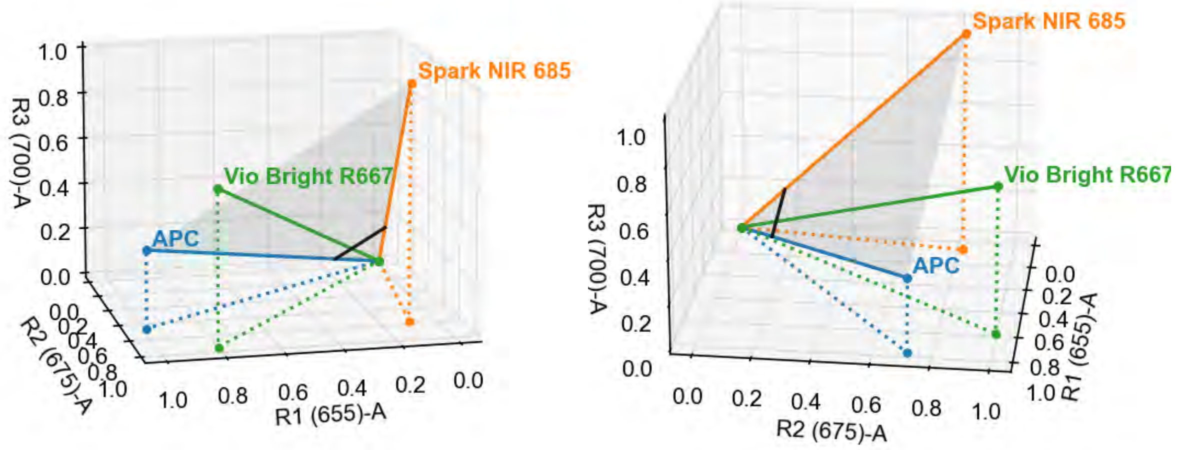


Collinearity occurs when N spectra can be described with <N dimensions

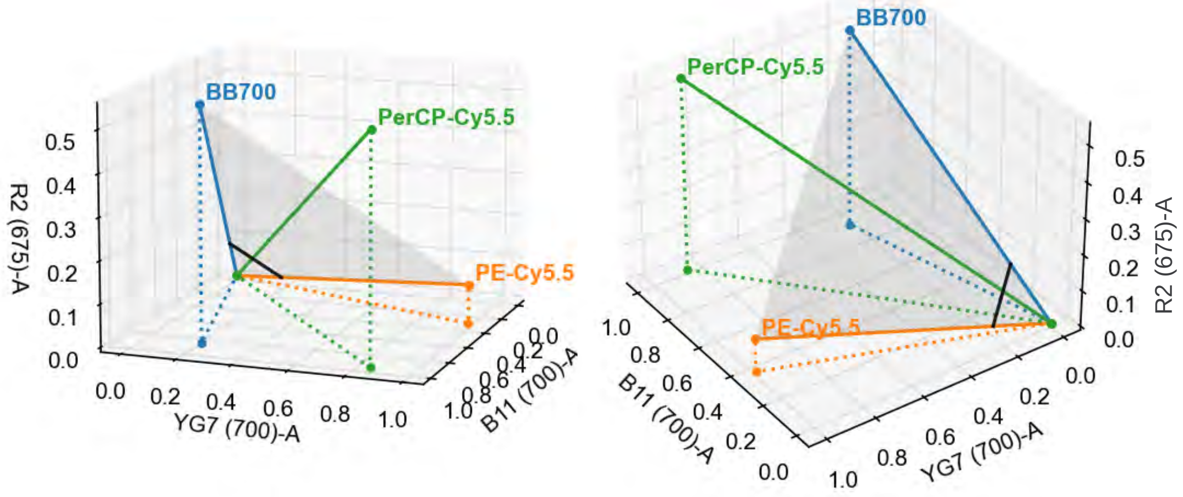




Collinear spectra (red neighborhood)

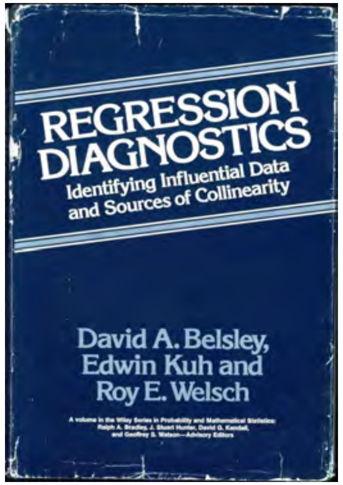


Collinear spectra (blue neighborhood)



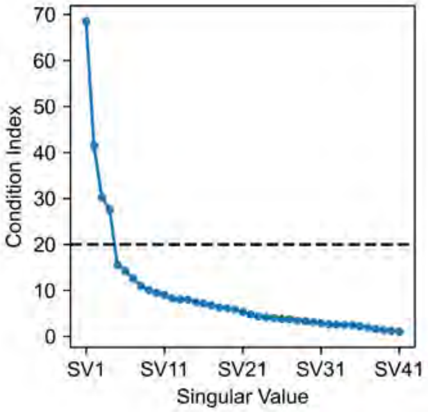
Understanding UDS in terms of collinearity lets us do cool math

Variance Decomposition Proportion analysis (VDP)



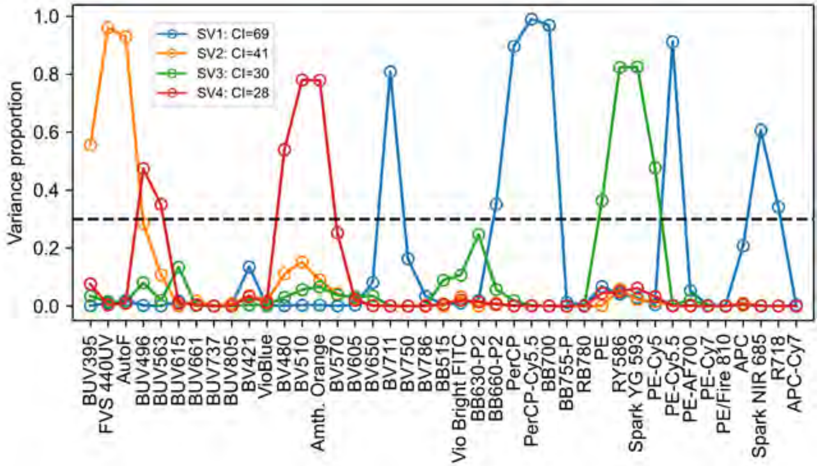
1

Find the singular value decomposition (SVD) components that are involved in collinearity



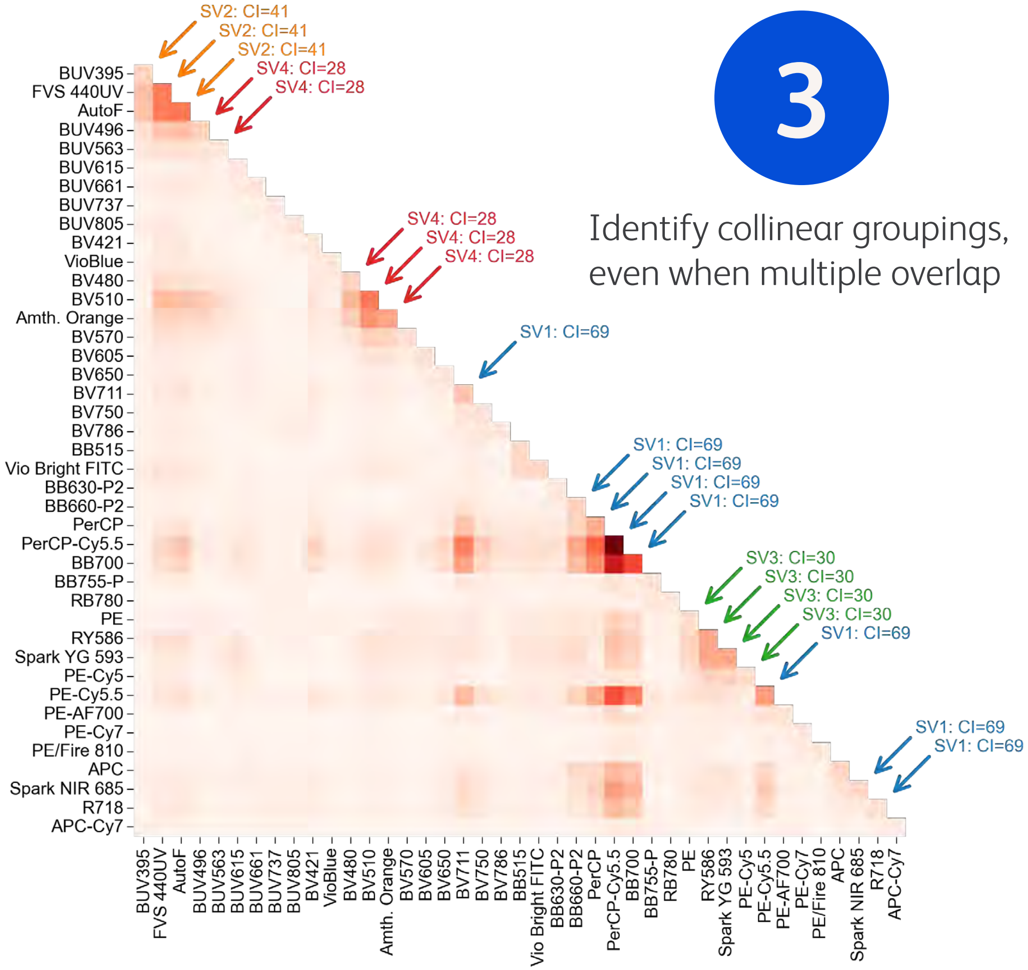
2

Check which fluorochromes have most of their variance explained by a common SVD component



3

Identify collinear groupings, even when multiple overlap



Mage, Konecny and Mair, preprint (2025). **doi:** https://doi.org/10.1101/2025.04.17.649396

Thank you!

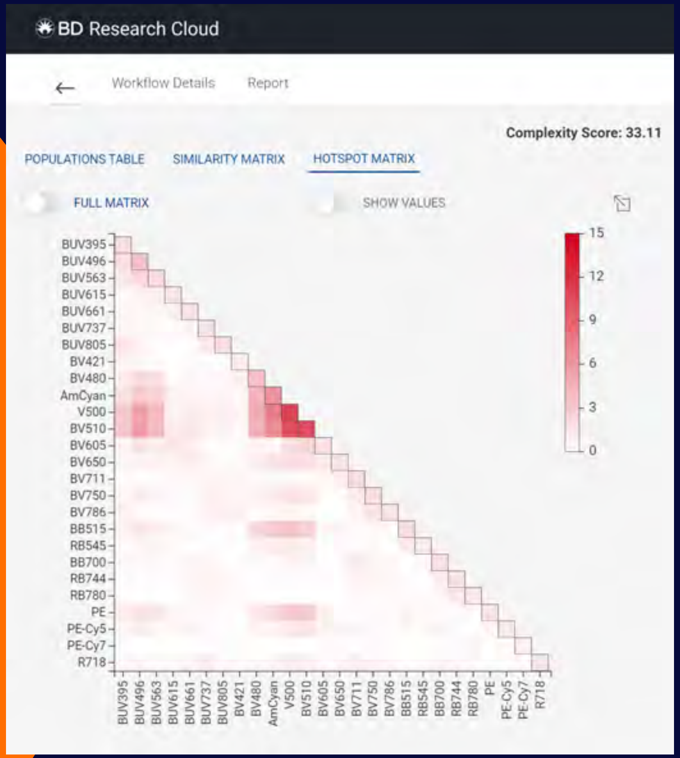
BD Biosciences
Laura Ferrer
Gisele Baracho
Xiaoshan Shi
Jessica Stokes
Lissette Wilensky
Stephanie Widmann
Emilie Jalbert
Aaron Middlebrook
Chip Lomas
Tri Le
Mirko Corselli
Aaron Tyznik

ETH Zurich
Florian Mair

Fred Hutchinson Cancer Center
Andrew Konecny
Martin Prlic

Preprint:

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY



Questions?