

SCIENTIFIC REPORTS



OPEN

Statistically strong label-free quantitative identification of native fluorophores in a biological sample

Saabah B. Mahbub¹, Martin Plöschner¹, Martin E. Gosnell^{1,2}, Ayad G. Anwer¹ & Ewa M. Goldys¹

Bioimaging using endogenous cell fluorescence, without any external biomarkers makes it possible to explore cells and tissues in their original native state, also *in vivo*. In order to be informative, this label-free method requires careful multispectral or hyperspectral recording of autofluorescence images followed by unsupervised extraction (unmixing) of biochemical signatures. The unmixing is difficult due to the scarcity of biochemically pure regions in cells and also because autofluorescence is weak compared with signals from labelled cells, typically leading to low signal to noise ratio. Here, we solve the problem of unsupervised hyperspectral unmixing of cellular autofluorescence by introducing the Robust Dependent Component Analysis (RoDECA). This approach provides sophisticated and statistically robust quantitative biochemical analysis of cellular autofluorescence images. We validate our method on artificial images, where the addition of varying known level of noise has allowed us to quantify the accuracy of our RoDECA analysis in a way that can be applied to real biological datasets. The same unsupervised statistical minimisation is then applied to imaging of mouse retinal photoreceptor cells where we establish the identity of key endogenous fluorophores (free NADH, FAD and lipofuscin) and derive the corresponding molecular abundance maps. The pre-processing methodology of image datasets is also presented, which is essential for the spectral unmixing analysis, but mostly overlooked in the previous studies.

Fluorescence microscopy with its various modalities is a powerful tool for studying biological processes in cells and tissues¹. In order to reveal vital information about the imaged specimen, targeted labelling of multiple structures with extraneous fluorophores is usually performed. However, there are many applications where a more subtle, non-invasive approach might be desired, for example in the studies of fragile early embryos². Fortunately, many chemical compounds in biological specimens, including nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADPH), flavin adenine dinucleotide (FAD), flavin mononucleotide (FMN), retinol, cytochrome complex (Cyt. c), collagen, elastin and many other metabolites, enzymes, co-factors and vitamins in cells naturally exhibit fluorescence². Importantly, the same compounds play a vital role in cellular metabolism² and can therefore, provide highly informative fluorescent signatures³. Revealing the spatial distribution of cellular fluorophores and quantifying their concentration is therefore of utmost importance for the monitoring of biological processes⁴⁻⁶.

Native fluorescence in cells comprises a mixture of broad featureless spectra⁷ from multiple distinct molecular species. Each location in an imaged sample has a somewhat different chemical composition. This spatially-varying biochemistry is reflected in the spectral diversity of native fluorescence at various locations in the image. In this case, hyperspectral imaging and unsupervised linear unmixing make it possible to discriminate individual distinct fluorophores despite their highly overlapping spectra⁸, also in the presence of unavoidable image noise.

A range of approaches to unsupervised unmixing has been previously developed, mostly in the area of remote sensing⁹⁻¹¹. The most popular unmixing methods such as N-FINDR^{12,13} require the presence of “pure” pixels that are the locations where only a single endmember is present. When pixel purity cannot be guaranteed, such as in our case of autofluorescence from biological samples, or in the presence of noise, alternative numerical analysis methods must be used, such as Minimum-Volume Transform (MVT)¹⁴, Sparsity Promoting ICE (SPICE)¹⁵,

¹ARC Centre of Excellence for Nanoscale Biophotonics, Macquarie University, North Ryde, 2109, NSW, Australia.

²Quantitative Pty Ltd, ABN 17165684186, 116-118 Great Western Highway, Mt. Victoria, NSW, 2786, Australia. Correspondence and requests for materials should be addressed to S.B.M. (email: saabah.mahbub@gmail.com) or E.M.G. (email: ewa.goldys@mq.edu.au)

Minimum-Volume Enclosing Simplex (MVES) algorithms¹⁶, Minimum-Volume Simplex (MVSA)¹⁷, Dirichlet mixture Unmixing via Split Augmented Lagrangian (DUSAL)¹⁸ and Simplex Identification by variable Splitting and Augmented Lagrangian tools (SISAL)¹⁹. However, none of them except MVT^{2,3} have been previously applied to cell autofluorescence. They have only been applied to reflectance, where absolute reference points (such as in our case the fluorescence spectra of pure compounds) are not applicable. Moreover, the reflectance spectra must be less than unity, while fluorescence spectra are not similarly constrained; this is because fluorescence intensity can take arbitrary values. Furthermore, autofluorescence is usually more strongly affected by background noise than reflectance, with lower signal to noise ratio (SNR). In comparison, in conventional fluorescently labelled microscopy images of cells and tissues a typical range of SNR is 20–40 dB²⁰. However, in this study, the autofluorescent images of retina cells are characterised by a far lower SNR of ~5 dB; similar SNR values were observed in previous investigations^{2,3,21}.

The aim of this paper is to develop a state-of-the-art method to unmix cell autofluorescence. Here, unmixing with statistical minimisation¹⁸ is applied to biological hyperspectral fluorescence images for the first time. Earlier work accomplished the unmixing of cell autofluorescence by the MVT method^{2,3}. Our approach of RODECA is based on DECA and DUSAL^{18,22} (Section 3.4) with appropriate modifications namely introducing SISAL and normalization of the spectra (see Supplementary Material, Section 4). By using artificial data with varying degree of known signal noise we estimate the unmixing accuracy (uncertainty of fluorophore spectra and their abundances), which was impossible in our previous approach². We then employ this technique to unmix real biological autofluorescence images, we separate cellular autofluorescence into individual biochemical constituents and provide their cellular abundance maps.

Background of hyperspectral unmixing. In hyperspectral imaging, two-dimensional images (N pixels each; here $N \approx 10^6$) of the same sample are captured in L different spectral channels ($L = 18$ in our study). The recorded fluorescence signals form the hyperspectral dataset described by a matrix $\mathbf{y} = [y_{ki}]$, where y_{ki} is the pixel value in channel k , ($k = 1, \dots, L$) of image pixel i , ($i = 1, \dots, N$)²³. This matrix can be considered as a sum of the noiseless signal matrix \mathbf{x} and the matrix of image noise \mathbf{n} , where $\mathbf{y} = \mathbf{x} + \mathbf{n}$. Further analysis is based on a linear mixing model (LMM) which postulates that the observed spectrum in each pixel is a linear combination of a small number p of component spectra called “endmembers” with respective weights (that is concentration of the spectrum component) called “abundance” fractions^{23–26}. For the unmixing to be accurate, the number of components needs to be smaller than the number of channels ($p < L$). Algebraically, in the LMM, the noiseless signals \mathbf{x} are expressed with the aid of an endmember matrix $\mathbf{M} = [M_{kj}]$, (where, $k = 1, \dots, L, j = 1, \dots, p$) is weighted with abundance fractions specified in the abundance matrix $\mathbf{s} = [s_{ji}]$, ($j = 1, \dots, p, i = 1, \dots, N$) according to

$$\mathbf{x} = [x_{ki}] = \sum_{j=1}^p M_{kj} s_{ji} = \mathbf{M}\mathbf{s}. \quad (1)$$

Due to physical considerations, the abundance fractions in each pixel must be positive ($s_{ji} \geq 0$) and they must add to unity $\sum_{j=1}^p s_{ji} = 1$ ^{18,27}. The abundance matrix \mathbf{s} contains the abundance fractions of each endmember for all N pixels and this makes it possible to easily construct the abundance map for the imaged sample. The L -dimensional vectors \vec{y}_i forming columns of the matrix \mathbf{y} , represent the pixel spectra, while the L -dimensional vectors \vec{M}_j forming columns of matrix \mathbf{M} represent the endmember spectra. Equation (1) can be expressed as

$$\vec{y}_i = \vec{M}_1 s_{1i} + \dots + \vec{M}_p s_{pi} \quad (2)$$

This formulation makes it possible to see that the Equation (2) describes a convex set in L -dimensions span by the vectors \vec{M}_j . Therefore the pixel spectral vectors from the matrix \mathbf{x} form a convex hull (a simplex) whose vertices are determined by the endmember spectral vectors²⁸. There is p such vectors, so the convex hull has p vertices. The presence of noise \mathbf{n} creates a fuzzy boundary of this convex hull.

In the unsupervised unmixing approach both matrices \mathbf{M} and \mathbf{s} are unknown, and, additionally, we do not know the number of endmembers p . As previously indicated, the observed pixel signals in the hyperspectral dataset \mathbf{y} are affected by the sensor noise coming from the image sensor and there may be other experimental sources of errors which are impossible to eliminate. Equation (2) then becomes:

$$\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{n}, \quad (3)$$

The noise matrix \mathbf{n} is also unknown. The procedure of unsupervised unmixing aims to establish the number of component spectra of a linear mixture and their identity, specifically to estimate optimal p , \mathbf{M} and \mathbf{s} none of which are known *a-priori*. Such statistical optimisation can give accurate results because hyperspectral imaging generates large datasets, so a large number N of experimental pixel spectra are available for analysis.

Methods

Hyperspectral Hardware Setup. We used a fluorescence microscope (Olympus iX71TM) with a 40× water U12TM series objective, with the wide transmission in UV range. Selected bands of excitation wavelengths (centred at 334, 365, 375, 385, 395, 405, 415, 425, 435, 455, 475, 495 nm, each about 10 nm wide) are used to excite cell autofluorescence. Three epifluorescence filter cubes (details are given in Supplementary Material, Section 5) are available to measure single photon-excited emission of biological samples. With these 12 excitation sources and 3 filters, we created a total of 18 specific channels. Optical powers at the objective range from 0.01 μW (at 495 nm excitation with 587 nm emission, channel 15) to 42.8 μW (at 475 nm excitation with 587 nm emission, channel 14). The excitation sources are coupled by an optical fibre bundle with a 5 mm fused silica hexagonal

homogenizer²⁹. The excitation sources produce a reasonably flat approximately Gaussian distribution of illumination over the sample plane, whose flatness is further corrected digitally. All images are captured by Andor iXON™ camera (EMCCD, iXON 885 DU, Andor Technology Ltd., UK) operated below -65°C to reduce sensor-induced noise. Some of the underpinning noise mechanisms depend on illumination level and they cannot be reduced by sensor cooling. The sensor size is 1002×1004 pixels.

Data pre-processing. At the beginning of each experiment, reference images including calibration, water, and dark images are taken using our hyperspectral microscope system. These reference images are then used to pre-process the sample images. The pre-processing steps include image equalization, primary denoising with removing undetectable pixels and outliers (spikes or dips), background illumination flattening and cell segmentation. All this is carried out without changing the mathematical structure of the dataset. The pixel identification (image number, pixel coordinates, spectral channel etc.) are separately retained for the reconstruction of two-dimensional fluorophore abundance maps. The following pre-processing steps are carried out:

Image equalisation. In the image equalisation procedure, the intensity count at every channel is converted into the units of photons per pixel per second (PPS). This calculation helps to standardize images taken with different acquisition parameters, most notably electron-multiplication (EM) gain, and acquisition time. For our Andor iXON™ camera, the sample signal expressed in terms of photon per second, $y_{raw [PPS]}$ is given by:

$$y_{raw [PPS]}(k, i) = \frac{(y_{raw [digital]}(k, i) - BO_{[digital]}) \times se}{G_{EM} \times QE \times t_{exp}} \quad (4)$$

here $y_{k,i [digital]}$ denotes the measured digital counts (in the range $0-2^{14}$). The bias offset ($BO_{[digital]}$) used in our setup is 100 counts. The camera sensitivity (se) for our readout rate of 13 MHz is 0.89. The EM gain (G_{EM}) and exposure time (t_{exp}) are adjusted by the operator taking into account the sample signals and they are generally different in different channels³⁰. The quantum efficiency (QE) of camera sensor is also different for different channels (see Supplementary Material, Section 5).

Removing undetectable pixels. Two sets of dark images (acquired, respectively, with the microscope shutter open and closed) are taken to remove the undetectable pixels. Such undetectable pixels could be due to light blockages (e.g. by dust) between the sensor and the sample plane, or inactive camera pixels. The average of these two dark images is subtracted from all sample, water and calibration images, to correct for any pixels that are unresponsive.

Removing outliers (spikes). Abnormal behaviour of sensor pixels in combination with high EM sensor gain may cause random sharp spikes or sharp dips in the image. To remove these outliers, a ‘threshold limiting window’ is scanned over all the images to locate these spikes or dips^{31,32}. Then these specific data points are replaced with the values interpolated from immediately adjacent nine pixels^{31,33}.

Image smoothing. The main sources of noise from EMCCD camera include illumination independent and illumination-dependent noise³⁴. The illumination independent noises (e.g. dark-current shot noise, readout noise etc.) are minimised by using low sensor temperature (below -65°C). Illumination-dependent noise (e.g. photon shot noise, clock induced charge noise, EM gain register noise etc.) are considered as multiplicative temporal and spatial noise³⁵. The overall noise in autofluorescence images is a combination of illumination dependent noise which is approximately Poissonian, while the noise from the illumination independent sources can be modelled as a Gaussian noise³⁶. We used Gaussian noise in our simulation as a proxy for the overall noise, because the Poisson’s noise amplitude cannot be modified independently from the signal. Here, we used the customised wavelet filter to remove the image noise for smoothing^{2,29} which facilitate improved capture of spectral information from a signal compared with standard frequency spectra produced by Fourier analysis². The wavelets help to divide the signal into different scale components³⁷ and thus these customized wavelet filters have proved to be a computationally efficient method of capturing textural information from filters or banks of filters with attractive attributes with potentially lossless coverage of the frequency spectrum².

Removing background autofluorescence. The images are also affected by the unavoidable autofluorescence signals from the microscope slide, Petri dishes, dirt on sensors etc. These signals make additive contributions to all images. To remove these contributions we take two hyperspectral images of water in the petri dish used for imaging³⁸. The smoothed average of these two images is denoted by $B(k, i)$. This smoothed average image, different for each channel, is subtracted from each sample image in this specific channel. See Section 3.2.7 for the description of smoothing.

Measurements of reference images and reference spectra. The microscope system is calibrated by taking hyperspectral images of a ‘calibration fluid’ which in our experiments is a mixture of $30 \mu\text{M}$ NADH (quantum yield 0.019) and $5 \mu\text{M}$ riboflavin (quantum yield 0.24). Its composition has been adjusted so that the spectrum of the calibration fluid has non-zero response across all our spectral channels. The smoothed image of the calibration fluid is denoted by $C_{raw}(k, i)$. The excitation and emission spectra of the calibration fluid are also separately measured with fluorimeter, in our case a Cary Eclipse Fluorescence Spectrophotometer™, making sure these spectra span each channel used in hyperspectral image acquisition. The normalized spectrum, $f(k)$ of the calibration fluid thus obtained is used to correlate the hyperspectral images with the fluorescence spectra measured on a fluorimeter. This is necessary to be able to correctly assign the unmixed fluorophores, by using the spectra of pure

reference fluorophores measured using the same fluorimeter with the same settings, in particular the same grating as used with the calibration fluid.

Relating the hyperspectral images to fluorescence spectra, image flattening and final image smoothing. Finally, the raw sample image, $y_{raw}(k, i)$, is corrected by using the averaged and smoothed background image $B(k, i)$ and multiplied by the normalized spectrum of the calibration fluid, $f(k)$. Furthermore, the smoothed image of the calibration fluid is used to correct for the somewhat uneven (approximately Gaussian) illumination of the field of view³⁹. This is done by dividing the sample image in each channel (after subtracting of the smoothed water image) by the relevant smoothed image of the calibration fluid². These corrections are specified in Equation 5:

$$y(k, i) = \frac{f(k) \times (y_{raw}(k, i) - B(k, i))}{(C_{raw}(k, i) - B(k, i))} \quad (5)$$

Cell segmentation. In order to calculate cellular features, cellular images need to be segmented into individual cells. In this work, retina cells were selected manually with the overlaid DIC image. The normalized autofluorescence intensity in each cell is documented in the y_{ki} matrix. The image number, pixel coordinates i and the spectral channel indices k are saved separately for the reconstruction of two-dimensional fluorophore abundance maps.

Identifying the optimised vector subspace by HySime and data decorrelation. Hyperspectral datasets pose two main problems for the analysis. Firstly, the number of channels L may be large and this generates high volume of data which then leads to computational complexity. Secondly, hyperspectral datasets are highly correlated. This is apparent in the images in spectrally close channels which look closely similar (as illustrated in Fig. 1f–w). In this work, we address these two problems of dimensionality reduction and decorrelation by a standard approach called HySime^{17,28}. In HySime we first calculate the $(L \times L)$ signal covariance matrix which is then diagonalised. In the next step, all possible subsets of its orthogonal eigenvectors are individually examined. A numerical criterion derived in Supplementary Material (Section 1, Equation 14) is then evaluated for each of these subsets. This criterion uses the data matrix y only, and it minimises the mean square error between the unknown noiseless signal matrix x , and the noisy observed signal $y = x + n$. The subset which best satisfies this criterion is then selected as the set of p basis vectors (E_1, \dots, E_p) spanning the optimised “signal” subspace (details are presented in Supplementary Material Section 1.3). The number of these new basis vectors is generally less than L . The original data vectors y (pixel spectra) are then orthogonally projected onto this optimised subspace. The resulting new p -dimensional data vectors form an optimal low-dimensional representation of the original data. These data vectors provide the input to the DECA procedure.

DECA procedure and its modifications. In this work, the Robust Dependent Component Analysis (RoDECA) algorithm is used for extracting the native fluorophores and their corresponding abundance matrix which is a modified version of DECA¹⁸. DECA is an unsupervised hyperspectral unmixing method based on the assumption of the linear mixing model. DECA is well suited to highly mixed datasets, where there are no pure pixels present. In the DECA analysis, the intermediate $p \times p$ endmember matrix A is initially estimated blindly, by simulating the abundance matrix s as a mixture of Dirichlet densities¹⁸ (details are given in Supplementary Material, Section 3). The main advantage of using the Dirichlet densities is that this approach automatically enforces the constraints of non-negativity and unity sum, and thus it accounts for statistical dependence usually found in hyperspectral data which satisfy the LMM conditions. The Dirichlet densities are also well suited to model complex distributions in which the ‘mass probability function’¹⁸, is scattered over a number of k -clusters inside the simplex⁴⁰. Defining $W \equiv A^{-1}$ we then obtain that the abundance map s in the subspace is given by:

$$s = Wx. \quad (6)$$

The maximum likelihood estimation of $W^{41,42}$, is carried out using an iterative expectation maximization (EM) algorithm¹⁸. This algorithm iterates between the “expectation” step and “maximisation” step⁴³ following the General Expectation Maximisation (GEM) approach^{28,43} (details are given in Supplementary Material, Section 3.3 and Section 3.4). The optimization problem in the case of biological autofluorescence is very challenging owing to the nonconvex term $\log|\det(W)|$, originally introduced by the DUSAL function in DECA and that is why in this work we put forward a more relaxed optimization process.

In our RoDECA analysis⁴⁴, a soft negative function is introduced to determine the $\log|\det(W)|$ by simplex identification via the split augmented Lagrangian method (SISAL)¹⁹ (details are given in Supplementary Material, Section 4.1, Equation 38). SISAL optimises the hard nonconvex problem (where there is no sharp vertices) by using a soft negative function¹⁹ for the non-smooth convex hull^{45,46}. The SISAL method was chosen because it helps to maintain the full rank with account of physical constraints (non-negativity and sum-to-one), in which the pure pixel assumption is violated¹⁹. Normalization has been introduced to the extracted endmember matrix M (after SISAL), ensuring the non-negativity and normalized values at each channel (details are given in Supplementary Material, Section 4.2). When the input data has been normalized, the extracted data should be normalised as well to prevent undesired discrimination of the same biochemical with different intensities⁴⁷, which cannot be achieved by the SISAL or DUSAL alone.

Cell preparation and imaging conditions. In this study, we used the RGC-5 cell line, a neuronal cell line of retinal origin⁴⁸ derived from the mouse retinal photoreceptor cells⁴⁹. The cells were maintained in Dulbecco’s

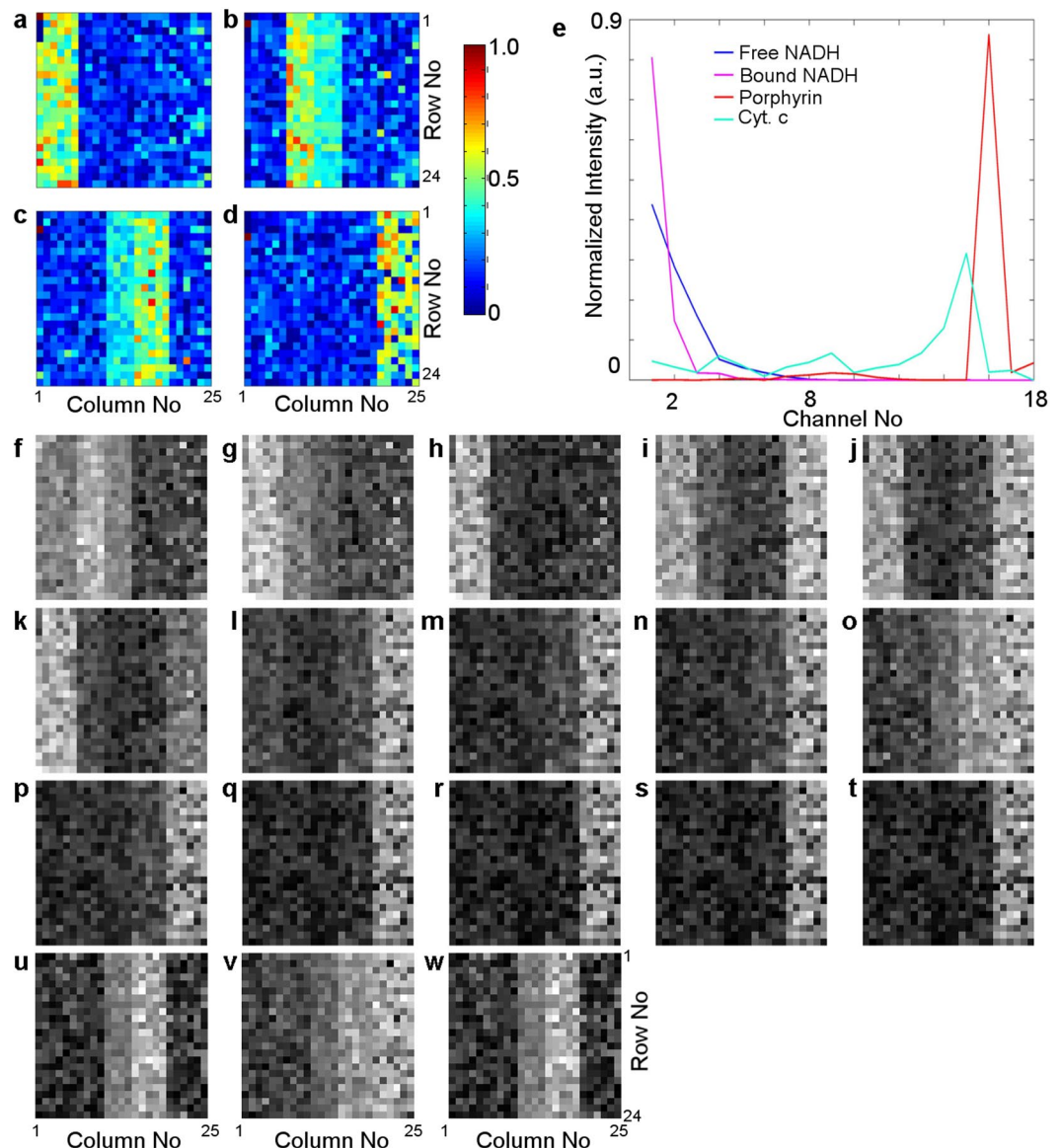


Figure 1. (a–d) The artificial abundance map $s_{[4 \times 600]}$ of the total of 600 (24×25) pixels containing the fractions of each of four fluorophores (a) free NADH, (b) bound NADH, (c) porphyrin and (d) Cyt. c. This abundance map corresponds to case D. (e) The normalized reference spectra are shown for free NADH (blue), bound NADH (magenta), Cyt. c (cyan) and porphyrin (red). (f–w) Simulated hyperspectral image set ($y = Ms + n$) assuming zero noise ($n = 0$) taken in $L = 18$ different spectral channels. The corresponding channel numbers are given in Supplementary Material, Section 5. Images (f–w) are rescaled for better visualization as a grey colour-map, where the minimum value is displayed as black, and the maximum value (varying at each channel) is displayed as white (details of the colour ranges are given in Supplementary Table 2).

Modified Eagle's medium (DMEM-high glucose, Sigma Aldrich, D5796) containing 10% bovine serum (Sigma Aldrich, 12133C) at 37°C with 5% CO₂ incubator. Approximately, 2.0×10^5 cells were seeded in each culture dish 12 hours before imaging.

The RGC-5 cells were cultured into three 35 mm plastic culture dishes with 18 mm diameter wells and # 1.5 gridded coverslip bottoms (Cell E&G, USA, and # GDB0004-200) for the preparation of hyperspectral imaging. Then DMEM medium containing 10% fetal bovine serum was removed (as it has own autofluorescence) and the cells were washed with Dulbecco's phosphate-buffered saline (DPBS, Sigma Aldrich, D8537) three times. After that 1 ml of non-fluorescent Hanks Balanced Salt Solution (HBSS, Life Tech, Australia, 14025076), was added to each dish for imaging. During the imaging session, the temperature was maintained at 37°C. The cells located over a reference grid at the bottom of the dishes were imaged using 18 fluorescence channels. These gridded coverslips are also useful for further correlation of hyperspectral imaging with labelled imaging, not reported here. A single wavelength image is typically acquired for between 0.5–15 seconds depending on the sample and wavelength, approximately 3 minutes for the entire stack of images for all 18 spectral channels.

Results and Discussion

Overview of Hyperspectral Unmixing Analysis. The hyperspectral unmixing analysis involves several steps. First, hyperspectral images are taken with modified autofluorescence microscope setup (Section 3.1); these form the hyperspectral dataset \mathbf{y} . It is desirable to unmix all images from a biological experiment together, as this enables their more accurate comparisons, but single images can also be unmixed. The hyperspectral dataset \mathbf{y} then undergoes data pre-processing (Section 3.2). Subsequently, linear unmixing is carried out for the dataset \mathbf{y} (Sections 3.3 and 3.4). It involves several steps, including; decorrelation and estimation of the number of endmembers p by the HySime method²⁸, and estimation of the noise matrix \mathbf{n} ^{28,50} in the least square error sense⁵⁰ with an iterative statistical procedure (Section 3.3). This is followed by endmember extraction and generation of the abundance matrix with the RoDECA (Section 3.4). This step iteratively derives the abundance matrix from an initial mixture of Dirichlet densities and endmember spectra. The procedure is blind, and it does not require any geometrical algorithms^{9–11,18}.

Determining the accuracy of unmixing using simulated data. In this work, we first applied the RoDECA analysis to an artificial data set, in order to determine how accurately we can unmix fluorophores for varying amounts of deliberately added noise. Such artificial hyperspectral fluorescence data make it possible to generate noisy data with a varying signal to noise ratio. Mathematically, we generate a reasonably realistic “zero-noise” dataset $\mathbf{x} = \mathbf{M}\mathbf{s}$ by selecting real biological fluorophores and using their real spectra. Then we generate a suite of different white Gaussian noise matrices \mathbf{n} with different noise amplitudes and thus produce artificial signal matrices $\mathbf{y} = \mathbf{M}\mathbf{s} + \mathbf{n}$. We then calculate the \mathbf{M} and \mathbf{s} by using our developed RoDECA procedure and compare them with original \mathbf{M} and \mathbf{s} for further analysis.

Preparation of artificial data sets. We generated pre-determined abundance maps \mathbf{s} , by creating two-dimensional images $N = 24 \times 25$ pixels representing maps of abundance fractions and assigned them to each of the four chosen fluorophores free NADH, bound NADH, porphyrin and cytochrome complex, Cyt. c (see example in Fig. 1a–d) and free NADH, FAD, porphyrin and cytochrome complex, Cyt. c (see example in Fig. 2a–d). These fluorophores are commonly found in biological samples, and they are related to cell metabolism. Among them, the ratio of NADH and FAD reflects the metabolic rate in cells^{4,51}. The spectra of chosen fluorophores in 18 channels are shown in Fig. 1(e) and in Fig. 2(e). We emphasise that these spectra are neither emission nor excitation spectra, they are proportional to fluorescence intensities of the respective fluorophores in the spectral channels whose specifications are given in Supplementary Material, Section 5. The abundance map \mathbf{s} satisfies physical constraints, namely that all four abundance matrices in each pixel add to unity and they are all positive.

The simulated dataset \mathbf{x} with zero noise ($\mathbf{n} = 0$) has been obtained by multiplying the endmembers spectra \mathbf{M} , shown in Fig. 1(e), by the abundance matrix \mathbf{s} , the abundance matrix of case D shown in Fig. 1(a–d). The simulated data set is shown in Fig. 1(f–w), where we present images in all spectral channels with image scaling for better visualisation. During the image scaling process, the highest pixel value was assigned as white colour and the lowest pixel value was black (details of colour scales are given in Supplementary Table 2). The hyperspectral dataset cube with 18 channel images was then reshaped to form the y_{ki} matrix used in the unmixing analysis. The pixel coordinates (k, i) were saved separately for further reconstruction of the abundance maps from the extracted s_{ji} .

Four sets of artificial data (referred to as cases A, B, C, D) were generated with varying degree of spatial and spectral overlap. In case A, the maps of the abundance fraction have been designed with distinctive stripes for ease of visualisation. We assigned high abundance (up to 80–85%) to each of the four fluorophores in a separate region of the image. These regions do not overlap (Fig. 2a–d). The abundances in our artificial data are randomly generated to be able to represent a wide variety of abundance mixtures in the real data. This enables the reader to better visualise the convex hull surrounding the data points. This process is separate to the addition of the Gaussian image noise.

The remaining three datasets (B, C, D) are characterised by a higher degree of spatial overlap (between two fluorophores - Case B and three and four fluorophores - Case C). These results are presented in Section 6.2 and 6.3 in the Supplementary Material. The dataset with high spectral overlap (between free NADH and bound NADH - Case D) is presented in Fig. 1).

In case B, FAD and porphyrin spatially overlap in columns 11 through to 16 with the abundance fractions of around 50% each. The simplexes of the artificial data, as well as the spectra of extracted fluorophores compared with reference spectra in the presence of added Gaussian white noise with the SNR of 10 dB, 20 dB, 40 dB and 60 dB, are shown in Supplementary Figs 9, 10, 11 and 12 respectively.

In case C, FAD and porphyrin spatially overlap in column 11 through to 16 which produces the abundance fraction of around 50% for both FAD and porphyrin. Moreover, the regions of free NADH and Cyt. c spatially overlap in rows 11 through to 16 with the abundance fractions of around 50% each. RoDECA unmixing was carried out with added Gaussian white noise for the SNR of 10 dB, 20 dB, 40 dB and 60 dB. The results are shown in Supplementary Figs 14, 15, 16 and 17 respectively.

In Case D, we chose two spectrally similar fluorophores, free NADH and bound NADH (Fig. 1e). Moreover, in this example, we introduced significant spatial overlap between bound NADH and porphyrin, in column 11 through to 16, with abundance fractions of around 50% each. In order to establish the robustness of RoDECA unmixing, the unmixing analysis was carried over with a broad range of added Gaussian white noise with SNR from 10 dB to 60 dB. The simplexes of the artificial data, as well as the spectra of extracted fluorophores compared with reference spectra for the SNR of 10 dB, 20 dB, 40 dB and 60 dB, are shown in Supplementary Figs 18, 19, 20 and 21 respectively.

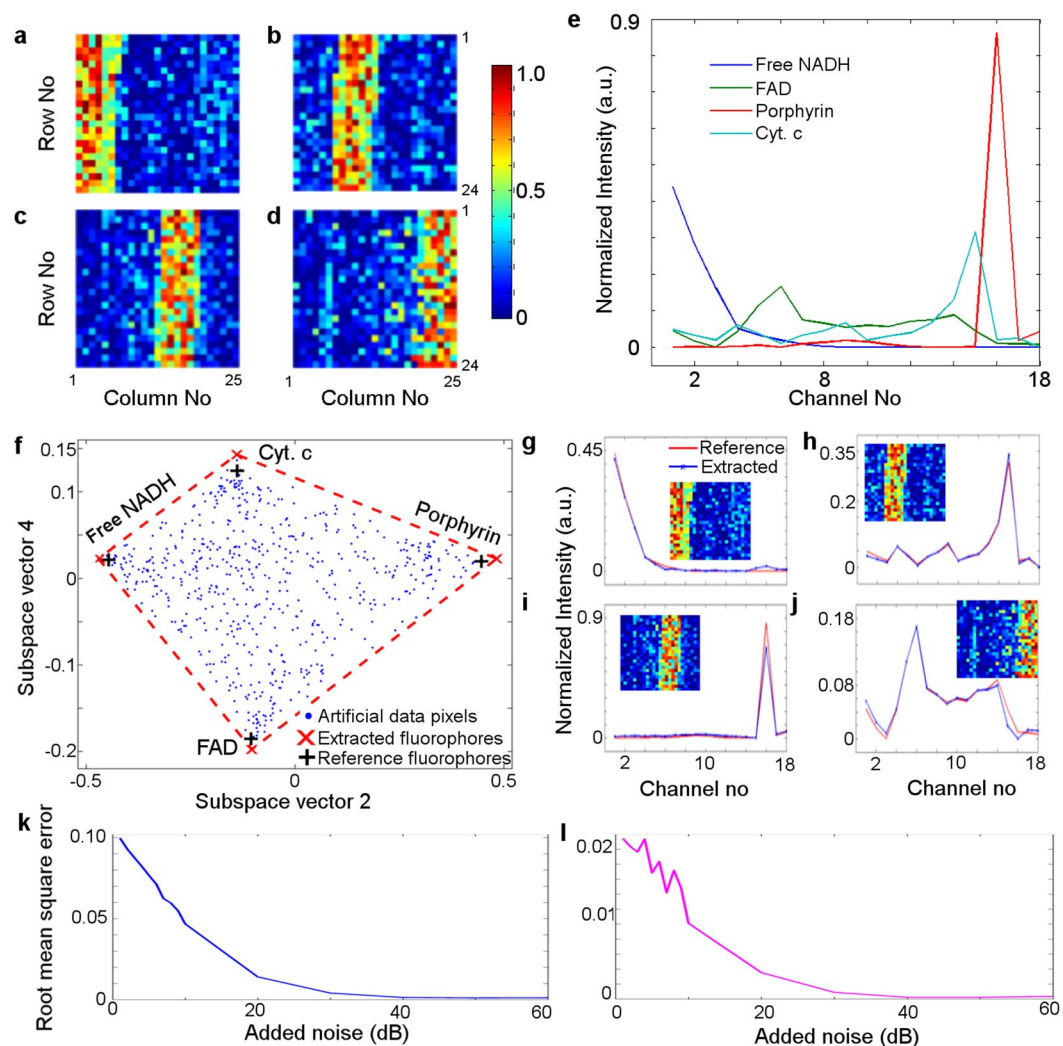


Figure 2. (a–d) The artificial abundance map $s_{4 \times 600}$ of the total of 600 (24×25) pixels containing the fractions of each of four fluorophore (a) free NADH, (b) FAD, (c) porphyrin and (d) Cyt. c. This abundance map corresponds to case A. (e) The normalized reference spectra are shown for free NADH (blue), FAD (green), Cyt. c (cyan) and porphyrin (red). (f) Simplex representation of simulated data corresponds to case A. The blue dots represent 600 data points projected onto the plane span by subspace eigenvectors 2 and 4, as in this plane all 4 vertices of the simplex were clearly visible. The red crosses represent the endmembers extracted by unsupervised unmixing using modified DECA with red enclosure marking the surface of the simplex and black crosses representing the reference spectra of free NADH, FAD, porphyrin and Cyt. c. (g–j) The endmembers spectra extracted using DECA (in blue) are compared with the reference spectra of (g) Free NADH, (h) FAD, (i) porphyrin and (j) Cyt. c (red). In Figures (f–j) the SNR is 20 dB. (k–l) Root mean square error calculations of the extracted (k) abundance matrices (in blue) and (l) endmember spectra (in magenta) for various levels of applied Gaussian noise n .

DECA unmixing of the artificial data set. We first discuss the results of unmixing of the artificial dataset (Case A) $y_{[18 \times 600]}$ with added Gaussian white noise yielding the signal to noise (SNR) ratio of 20 dB. The dataset y is processed with HySime, to obtain the number of fluorophores (in this case $p = 4$), and the corresponding subspace vectors E_1, \dots, E_p . In this subspace, the dimensionality of the input data y_{ki} is reduced, and the input data projected onto the subspace is now denoted y_{ji} ($j = 1, \dots, p$) (Section 3.3). Following the DECA analysis of this dataset, we obtain both M and s , both these matrices also projected onto the same subspace.

Both abundance maps (in Fig. 2a–d) and corresponding spectra (in Fig. 2e) are presented for this study. For display purposes, the artificial dataset y_{ji} was further projected onto the signal subspace vectors 2 and 4, see Fig. 2(f) (details are given in Supplementary Material, Section 1.3 and Section 2). The channel signals (spectra) of reference fluorophores were similarly projected (black crosses). As expected, the data points (blue dots, $N = 600$), form a convex simplex with four vertices. The vertices correspond to the extracted endmember spectra projected onto the chosen subspace. It can be seen that the endmembers and the original fluorophores closely coincide. Figure 2(g–j) shows the spectra of endmembers extracted by using DECA analysis compared with reference

fluorophore spectra for the added noise of SNR 20 dB, and the corresponding abundance maps. As anticipated, for this high signal to noise ratio the unmixed endmembers and reference fluorophores are practically identical.

In order to establish the robustness of DECA unmixing, this analysis was extended to cover a broad range of added Gaussian white noise with SNR from 1 dB to 60 dB. The simplexes of the artificial data (Case A), as well as the spectra of extracted fluorophores compared with reference spectra for the SNR of 60 dB, 40 dB and 10 dB, are shown in Supplementary Figs 4, 5 and 6. The root mean square error of abundance ξ_s is determined by using Equation 7 (see¹⁸). Here $s_{cal}(j, i)$ denotes the calculated abundance matrix and $s_{org}(j, i)$ is the original abundance matrix.

$$\xi_s = \frac{1}{p} \sum_{j=1}^p \frac{1}{N} \sum_{i=1}^N \sqrt{s_{cal}(j, i) - s_{org}(j, i)^2} \quad (7)$$

The spectrum error ξ_M is calculated as per Equation 8. Here, $M_{cal}(k, j)$ and $M_{org}(k, j)$ represent the calculated and original matrices of endmember spectra, respectively.

$$\xi_M = \frac{1}{p} \sum_{j=1}^p \frac{1}{N} \sum_{k=1}^L \sqrt{M_{cal}(k, j) - M_{org}(k, j)^2} \quad (8)$$

Figure 2(k,l) shows the root mean square error of the abundance fraction and for the endmember spectra for different levels of added Gaussian white noise. They confirm the expected trends such as lower errors at a lower value of added noise. As expected, at high SNR the individual maps of abundance fractions obtained from the RoDECA analysis and the original abundance maps are almost identical, with the abundance error for the SNR of 40 dB and 20 dB of 0.0014 and 0.0142 respectively. The spectrum error for the SNR of 40 dB and 20 dB is also very small with the respective values of 0.0002 and 0.0035.

Similar analysis was carried out using artificial data in cases B, C and D. The root mean square error of abundance ξ_s for Cases B - D is presented in Supplementary Figure 22 (blue coloured cross, diamond and asterisk symbols). The root mean square error of spectrum ξ_M for Case B - D is presented in Supplementary Figure 23 in red coloured cross, diamond and asterisk symbols.

We also performed an additional study to benchmark RoDECA with other well-established methods including SISAL¹⁹, VCA⁵², DUSAL¹⁸ and MVSA¹⁷. The results are shown in Supplementary Figure 24. The root mean square error of spectrum ξ_M of RoDECA is the smallest (0.0165), while it is 0.0867, 0.0566, 0.0897 and 0.0667 for SISAL, VCA, MVSA, and DUSAL respectively.

DECA unmixing in a biological experiment. In our experiment, we collected the hyperspectral images of retina cells, and we also took a series of reference images of the calibration fluid and water as well as dark images. Figure 3(a) shows a raw single channel image (1004 × 1002 pixels) of the retina cells in channel 5 where the digital counts in the hyperspectral images were converted into equivalent units of photons per pixel per second (PPS). The images have then been pre-processed as described in the Methods section. The pre-processing steps included image equalization, removing undetectable pixels, removing background fluorescence and flattening the images. In this experiment higher EM gains and longer acquisition time with multiple averaging is used to optimise image quality. The average of two dark images is then applied to remove the undetectable pixels. The image is flattened with a normalized spectrum of the calibration fluid, shown in Fig. 3(b) and multiplied by the appropriate value from the spectrum of the calibration fluid. After that, a customized wavelet filter is used to smooth the raw image, as presented in Fig. 3(c). This procedure increases the SNR to 8.5 dB from the original value of 4.6 dB. The insets of Fig. 3(a,b,c) show the intensity value across the image (in blue) for the ease of comparison between these pre-processing steps. Figure 3(d) shows the intensity plot of the smoothed image of the retina cell, presented in Fig. 3(c), whereas Fig. 3(e) shows the DIC image for the same region. Finally, the image was segmented manually using the superimposed DIC, Channel 2 and Channel 5 images.

After pre-processing, the DECA analysis was used for the unmixing of this hyperspectral dataset. Figure 3(f) shows the simplex for the analysed dataset from the retinal cells investigated here. The extracted fluorophores are shown as red crosses, whereas the corresponding reference spectra are shown with black crosses. The vertices are not clearly pronounced, compared with the artificial data, which is indicative of highly mixed fluorophores. The analysis of such dataset using alternative minimum volume methods MVT, MVES or MVSA^{2,3} would be challenging due to a higher degree of associated complex noise and highly mixed fluorophores. Since the endmembers are highly mixed, i.e., there is no spectral vector near the vertices nor near facets, the minimum volume methods are unable to satisfy their underpinning assumption of at least $p - 1$ spectral vectors on each facet of the data simplex¹⁸. In this case, statistical methods such as DUSAL¹⁸ or our RoDECA are necessary to handle such highly mixed data sets. However, with RoDECA analysis, the three most abundant fluorophores have been successfully identified. Figure 4a,e,i compares the (excitation) spectra of the extracted and reference fluorophores which shows that the extracted endmembers are similar to FAD, NADH and lipofuscin. The root mean square error of spectrum error was measured 0.0275 with respect to all 16 channels and root mean square error of individual spectra for FAD, NADH and lipofuscin is 0.0234, 0.0484 and 0.0107 respectively, presented in Fig. 4a,e,i.

These assignments are in agreement with earlier works in retina and retinal cells which reported the presence of NAD(P)H, FAD, and retinol^{53,54}, identified by the two photon microscopy^{53,55} or *ex vivo* fluorescence microscopy⁵⁶⁻⁵⁸, while lipofuscin, has been detected using single photon imaging methods⁵⁹⁻⁶¹. Our RoDECA unmixing methodology makes it possible to identify NADH and FAD by single photon fluorescence in a statistically robust way. We note that lipofuscin in retinal cells is composed of many different fluorophores with a broad excitation and emission profile⁶², which gives a very rounded vertex in our simplex representation in Fig. 3(f).

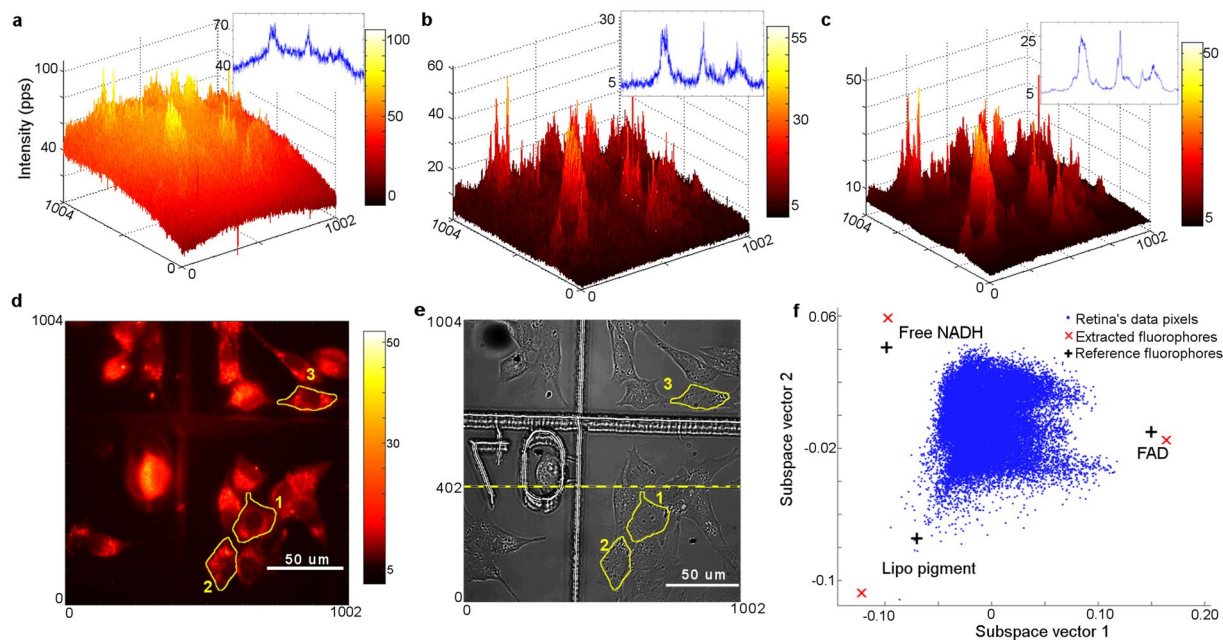


Figure 3. (a–e) Selected single channel image (1004×1002 pixels) of the retina cells in consecutive stages of pre-processing. (a) Raw data (y_{raw}) of retina cells in Channel 5 before any image pre-processing. Intensity is given in PPS units, as explained in Section 3.2.1, (b) the same image after background subtraction (Section 3.2.5) and image flattening (Section 3.2.7) by using the image of the calibration fluid. Intensity is given in PPS units. (c) Smoothed image after wavelet smoothing (Section 3.2.4). The intensity scale is now normalized. (d) The intensity map for Fig. 3(c), where cells are now clearly distinguishable in the smoothed image. Three cells are outlined (yellow line) and numbered for individual cell analysis. (e) The corresponding differential interference contrast (DIC) image with the three cells outlined in Fig. 3(d). Note that the grid itself and the grid number seen in the DIC image is not observed in the fluorescence image. (f) The simplex representation of the retina data on projection on the subspace 1 and 2 of the signal subspace. Data pixels - blue dots, extracted endmembers - red crosses. The location of the reference spectra of free NADH, FAD and lipofuscin are marked with black crosses. The total of 209028 pixels have been analysed.

The obtained abundance maps (Fig. 4b,c,d,f,g,h,j,k,l) show cell nuclei (dark ovals), confirming that we have captured aspects of cellular morphology. These abundance maps are expected to reflect relative proportions of fluorophore concentrations weighted with their fluorescence quantum yield. We note however, that some variations of the fluorescence quantum yield may be expected, as a result of the highly diversified chemical environments within a cell and pH variations. The FAD and NADH are found to be more pronounced in the perinuclear region of the cytosol. This is consistent with their mitochondrial localisation. Both FAD and NADH are the essential component of electron transport chain and used as an indication of mitochondrial function⁶³ also in retina⁶⁴, where they can serve as functional indicators of the ganglion cell layer^{65–67}. The ratio of the fluorescence intensity of these fluorophores (NADH/FAD), the NADH redox ratio, is a marker of the metabolic state⁶⁸. The third extracted fluorophore, lipofuscin is also to be expected in the retinal cells⁶⁹, because these cells accumulate this by-product of phagocytosis^{62,70,71}. The extracted abundance map of lipofuscin at Fig. 4(g) clearly show the lipofuscin granules. Lipofuscin granules are usually found in retinal pigment epithelial cells (RPE)⁶⁹, which allows monitoring the health conditions of the retina cells⁷². The density of lipofuscin granules in RPE cells in mice increased with age⁷³, and it is regarded as the key pathogenic indicator for age-related macular degeneration (AMD). As the eye is an ideal organ for non-invasive imaging, the presented methodology might contribute to improved quantification and evaluation of lipofuscin for clinical *in vivo* diagnostics^{58,74}.

Conclusions

This paper presents the mathematics and experimental application of the unsupervised unmixing analysis applied to biological autofluorescence images which contain significant levels of noise. Earlier hyperspectral imaging experiments analysed by unsupervised unmixing used mostly reflectance, however, reflectance is a significantly different modality from fluorescence due to the availability of the absolute reference spectra in fluorescence. Earlier studies of decomposition of multiply labelled fluorescence images of cells and tissues^{75,76} were limited to much higher SNR than in the current work, ranging from 20 dB to 40 dB³⁰. The image pre-processing steps include the standardization of the measured fluorescence intensity counts. This is followed by the subtraction of background autofluorescence and image flattening. These help to reduce illumination independent artefacts. The illumination-dependent complex noise is treated by a customized wavelet filter² for primary smoothing. The dataset is then decorrelated by HySime allowing the data dimensionality to be reduced. Subsequently, both unknown a-priori unknown endmembers and abundances are extracted with the RoDECA analysis, by introducing

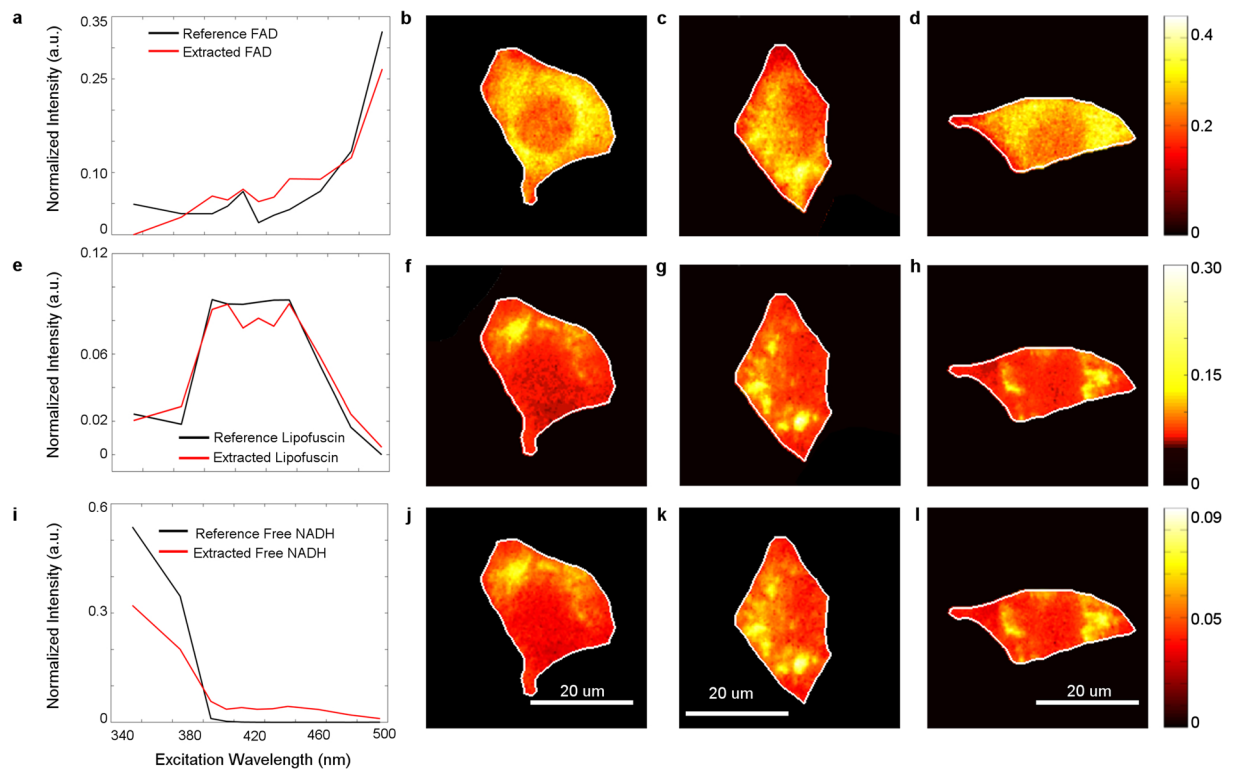


Figure 4. Abundance map representation of three retina cells, marked (as 1, 2, 3) in the DIC image in Fig. 3(e). First column (panels a,e,i) shows the extracted and reference spectra in red and black respectively. Figures (b,f,j) (c,g,k) and (d,h,l) represent the three abundance maps for the first, second and third cell, respectively. Here the first row shows the abundance map of FAD (b,c and d), the second row illustrates the abundance map of Lipofuscin (f,g and h) and third row presents the abundance map of Free NADH (j,k and l).

SISAL and the normalization of the spectra. Unmixing of artificial data has been conducted to quantify the noise-dependent unmixing error for both abundances and extracted spectra, which help evaluate the credibility of our RoDECA analysis. The real biological dataset of retina cells then has been tested to identify individual biochemical components of cellular autofluorescence and provide their relative abundance maps. These maps reflect the cellular morphology and correctly localise core cellular features. With its robust unmixing statistical analysis, this study reports a versatile unsupervised unmixing method for label-free quantitative identification of native fluorophores in a biological sample and it yields new insights into cytometry of retinal cells.

Quantitative analysis of autofluorescence features presented here has a vast array of potential applications in the analysis of cells and tissues, from an understanding of metabolic activity to developing future clinical tools for diagnostics and therapy monitoring². In this context, it is especially important that spectral fluorescence microscopy is a relatively inexpensive imaging methodology. As native fluorescence is such a universal and informative biological characteristics, the hyperspectral unmixing technology of autofluorescence has a potential for impact on the field of life sciences and biomedicine.

References

- Bates, M. A New Approach to Fluorescence Microscopy. *Science* **330**, 1334–1335, <https://doi.org/10.1126/science.1200252> (2010).
- Gosnell, M. E. *et al.* Quantitative non-invasive cell characterisation and discrimination based on multispectral autofluorescence features. *Scientific Reports* **6**, 23453, <https://doi.org/10.1038/srep23453> (2016).
- Gosnell, M. E., Anwer, A. G., Cassano, J. C., Sue, C. M. & Goldys, E. M. Functional hyperspectral imaging captures subtle details of cell metabolism in olfactory neurosphere cells, disease-specific models of neurodegenerative disorders. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1863**, 56–63 (2016).
- Quinn, K. P. *et al.* Quantitative metabolic imaging using endogenous fluorescence to detect stem cell differentiation. *Scientific Reports* **3**, 3432, <https://doi.org/10.1038/srep03432> (2013).
- Reyes, J. M. *et al.* Metabolic changes in mesenchymal stem cells in osteogenic medium measured by autofluorescence spectroscopy. *Stem Cells* **24**, 1213–1217 (2006).
- Renkoski, T. E. *et al.* Ratio images and ultraviolet C excitation in autofluorescence imaging of neoplasms of the human colon. *Journal of Biomedical Optics* **18**, 016005–016005 (2013).
- Pu, Y. *et al.* Native fluorescence spectroscopy reveals spectral differences among prostate cancer cell lines with different risk levels. *Journal of biomedical optics* **18**, 087002–087002 (2013).
- Zimmermann, T., Rietdorf, J. & Pepperkok, R. Spectral imaging and its applications in live cell microscopy. *FEBS Letters* **546**, 87–92 (2003).
- Biucas-Dias, J. M. & Plaza, A. In *Remote Sensing*. 78300A-78315 (International Society for Optics and Photonics, 2010).

10. Bioucas-Dias, J. M. *et al.* Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of* **5**, 354–379, <https://doi.org/10.1109/JSTARS.2012.2194696> (2012).
11. Bioucas-Dias, J. M. & Plaza, A. In *Geoscience and Remote Sensing Symposium (IGARSS), IEEE International*. 1135–1138 (2011).
12. Winter, M. E. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics*. 266–275 (1999).
13. Ambikapathi, A., Tsung-Han, C., Wing-Kin, M. & Chong-Yung, C. Chance-Constrained Robust Minimum-Volume Enclosing Simplex Algorithm for Hyperspectral Unmixing. *Geoscience and Remote Sensing, IEEE Transactions on* **49**, 4194–4209, <https://doi.org/10.1109/TGRS.2011.2151197> (2011).
14. Craig, M. D. Minimum-volume transforms for remotely sensed data. *Geoscience and Remote Sensing, IEEE Transactions on* **32**, 542–552 (1994).
15. Zare, A. & Gader, P. Sparsity Promoting Iterated Constrained Endmember Detection in Hyperspectral Imagery. *Geoscience and Remote Sensing Letters, IEEE* **4**, 446–450, <https://doi.org/10.1109/LGRS.2007.895727> (2007).
16. Tsung-Han, C., Chong-Yung, C., Yu-Min, H. & Wing-Kin, M. A Convex Analysis-Based Minimum-Volume Enclosing Simplex Algorithm for Hyperspectral Unmixing. *Signal Processing, IEEE Transactions on* **57**, 4418–4432, <https://doi.org/10.1109/TSP.2009.2025802> (2009).
17. Li, J. & Bioucas-Dias, J. M. In *Geoscience and Remote Sensing Symposium (IGARSS), IEEE International*. III-250–253 (IEEE, 2008).
18. Nascimento, J. M. P. & Bioucas-Dias, J. M. Hyperspectral Unmixing Based on Mixtures of Dirichlet Components. *Geoscience and Remote Sensing, IEEE Transactions on* **50**, 863–878, <https://doi.org/10.1109/TGRS.2011.2163941> (2012).
19. Bioucas-Dias, J. M. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, First Workshop on*. 1–4 (2009).
20. Anderson, N., Prabhat, P. & Erdogan, T. Spectral Modeling in Fluorescence Microscopy. 1–20 (Semrock, Rochester, NY).
21. Rehman, A. U. *et al.* Fluorescence quenching of free and bound NADH in HeLa cells determined by hyperspectral imaging and unmixing of cell autofluorescence. *Biomedical Optics Express* **8**, 1488–1498, <https://doi.org/10.1364/BOE.8.001488> (2017).
22. Nascimento, J. M. P. & Bioucas-Dias, J. M. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, First Workshop on*. 1–5 (2009).
23. Keshava, N. & Mustard, J. F. Spectral unmixing. *Signal Processing Magazine, IEEE* **19**, 44–57, <https://doi.org/10.1109/79.974727> (2002).
24. Keshava, N., Kerekes, J., Manolakis, D. & Shaw, G. An algorithm taxonomy for hyperspectral unmixing. (2000).
25. Keenan, M. R., Timlin, J. A., Van Benthem, M. H. & Haaland, D. M. In *International Symposium on Optical Science and Technology*. 193–202 (International Society for Optics and Photonics, 2002).
26. Keshava, N. A survey of spectral unmixing algorithms. *Lincoln Laboratory Journal* **14**, 55–78 (2003).
27. Manolakis, D., Siracusa, C. & Shaw, G. Hyperspectral subpixel target detection using the linear mixing model. *Geoscience and Remote Sensing, IEEE Transactions on* **39**, 1392–1409, <https://doi.org/10.1109/36.934072> (2001).
28. Nascimento, J. M. & Bioucas-Dias, J. M. In *Geoscience and Remote Sensing Symposium, 2007, IEEE International*. 3225–3228 (IEEE, 2007).
29. Gosnell, M. E. *Unlocking the potential of spectral imaging for the characterization of cell and stem cell populations* Ph.D. thesis, Macquarie University (2014).
30. North, A. J. Seeing is believing? A beginners' guide to practical pitfalls in image acquisition. *The Journal of cell biology* **172**, 9–18 (2006).
31. Behrend, C. J., Tarnowski, C. P. & Morris, M. D. Identification of outliers in hyperspectral Raman image data by nearest neighbor comparison. *Applied spectroscopy* **56**, 1458–1461 (2002).
32. Vidal, M. & Amigo, J. M. Pre-processing of hyperspectral images. *Essential steps before image analysis. Chemometrics and Intelligent Laboratory Systems* **117**, 138–148 (2012).
33. Jones, H. D. *et al.* Weighting hyperspectral image data for improved multivariate curve resolution results. *Journal of Chemometrics* **22**, 482–490 (2008).
34. Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. A technique for evaluation of CCD video-camera noise. *IEEE Transactions on Circuits and Systems for Video Technology* **18**, 280–284 (2008).
35. Nakamura, J. *Image sensors and signal processing for digital still cameras*. (CRC press, 2016).
36. Haider, S. A. *et al.* Fluorescence microscopy image noise reduction using a stochastically-connected random field model. *Scientific reports* **6**, 20640, <https://doi.org/10.1038/srep20640> (2016).
37. Clausi, D. A. & Zhao, Y. In *Geoscience and Remote Sensing Symposium, IEEE International*. 2453–2455 (IEEE, 2002).
38. Lakowicz, J. R. & Masters, B. R. Principles of fluorescence spectroscopy. *Journal of Biomedical Optics* **13**, 029901 (2008).
39. Healey, G. & Kondepudy, R. In *Computer Vision and Pattern Recognition, 1992 IEEE Computer Society Conference on*. 90–95 (1992).
40. Minka, T. Estimating a Dirichlet distribution. (Technical report, MIT, 2000).
41. McLachlan, G. & Peel, D. *Finite mixture models*. (John Wiley & Sons, 2004).
42. Figueiredo, M. A. T. & Jain, A. K. Unsupervised learning of finite mixture models. *IEEE Transactions on pattern analysis and machine intelligence* **24**, 381–396 (2002).
43. McLachlan, G. & Krishnan, T. *The EM algorithm and extensions*. Vol. 382 (John Wiley & Sons, 2007).
44. Mahbub, S. B. *Unsupervised hyperspectral unmixing analysis for label-free quantitative identification of native fluorophores in a biological sample by a Robust Dependent Component Analysis (RoDECA)* Ph.D thesis, Macquarie University, (2017).
45. Goldstein, T. & Osher, S. The split Bregman method for L1-regularized problems. *SIAM journal on imaging sciences* **2**, 323–343 (2009).
46. Afonso, M. V., Bioucas-Dias, J. M. & Figueiredo, M. A. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing* **19**, 2345–2356 (2010).
47. Gross, W., Schilling, H. & Middelman, W. In *Geoscience and Remote Sensing Symposium, IEEE International*. 4714–4717 (IEEE, 2012).
48. Gupta, V. K., You, Y., Li, J. C., Klistorner, A. & Graham, S. L. Protective effects of 7, 8-dihydroxyflavone on retinal ganglion and RGC-5 cells against excitotoxic and oxidative stress. *Journal of Molecular Neuroscience* **49**, 96–104 (2013).
49. Tan, E. *et al.* Expression of cone-photoreceptor-specific antigens in a cell line derived from retinal tumors in transgenic mice. *Investigative ophthalmology & visual science* **45**, 764–768 (2004).
50. Bioucas-Dias, J. M. & Nascimento, J. M. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing* **46**, 2435–2445 (2008).
51. Skala, M. C. *et al.* In vivo multiphoton microscopy of NADH and FAD redox states, fluorescence lifetimes, and cellular morphology in precancerous epithelia. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19494–19499, <https://doi.org/10.1073/pnas.0708425104> (2007).
52. Nascimento, J. M. & Bioucas Dias, J. M. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on* **43**, 898–910 (2005).
53. Rossi, E. *et al.* Imaging retinal mosaics in the living eye. *Eye* **25**, 301–308 (2011).
54. Chen, C. *et al.* Reduction of all-trans retinal to all-trans retinol in the outer segments of frog and mouse rod photoreceptors. *Biophysical journal* **88**, 2278–2287 (2005).

55. He, S., Ye, C., Sun, Q., Leung, C. K. & Qu, J. Y. Label-free nonlinear optical imaging of mouse retina. *Biomedical optics express* **6**, 1055–1066 (2015).
56. Reinert, K. C., Dunbar, R. L., Gao, W., Chen, G. & Ebner, T. J. Flavoprotein autofluorescence imaging of neuronal activation in the cerebellar cortex *in vivo*. *Journal of neurophysiology* **92**, 199–211 (2004).
57. Kuznetsov, A. V. *et al.* Functional imaging of mitochondria in saponin-permeabilized mice muscle fibers. *The Journal of cell biology* **140**, 1091–1099 (1998).
58. Balendra, S., Normando, E., Bloom, P. & Cordeiro, M. Advances in retinal ganglion cell imaging. *Eye* **29**, 1260–1269 (2015).
59. Gray, D. C. *et al.* *In vivo* fluorescence imaging of primate retinal ganglion cells and retinal pigment epithelial cells. *Optics Express* **14**, 7144–7158 (2006).
60. Morgan, J. I., Dubra, A., Wolfe, R., Merigan, W. H. & Williams, D. R. *In vivo* autofluorescence imaging of the human and macaque retinal pigment epithelial cell mosaic. *Investigative ophthalmology & visual science* **50**, 1350–1359 (2009).
61. Morgan, J. I. *et al.* Light-induced retinal changes observed with high-resolution autofluorescence imaging of the retinal pigment epithelium. *Investigative ophthalmology & visual science* **49**, 3715–3729 (2008).
62. Delori, F. C. *et al.* *In vivo* fluorescence of the ocular fundus exhibits retinal pigment epithelium lipofuscin characteristics. *Investigative ophthalmology & visual science* **36**, 718–729 (1995).
63. Elner, S. G. *et al.* Retinal flavoprotein autofluorescence as a measure of retinal health. *Transactions of the American Ophthalmological Society* **106**, 215 (2008).
64. Kageyama, G. H. & Wong-Riley, M. The histochemical localization of cytochrome oxidase in the retina and lateral geniculate nucleus of the ferret, cat, and monkey, with particular reference to retinal mosaics and ON/OFF-center visual channels. *Journal of Neuroscience* **4**, 2445–2459 (1984).
65. Schweitzer, D. *et al.* Towards metabolic mapping of the human retina. *Microscopy research and technique* **70**, 410–419 (2007).
66. Peters, S., Hammer, M. & Schweitzer, D. In *European Conference on Biomedical Optics*. 808605 (Optical Society of America, 2011).
67. Sharma, R., Williams, D. R., Palczewska, G., Palczewski, K. & Hunter, J. J. Two-Photon Autofluorescence Imaging Reveals Cellular Structures Throughout the Retina of the Living Primate Eye Two-Photon Autofluorescence Imaging. *Investigative ophthalmology & visual science* **57**, 632–646 (2016).
68. Maleki, S. *et al.* Optical imaging of mitochondrial redox state in rodent model of retinitis pigmentosa. *Journal of biomedical optics* **18**, 016004–016004 (2013).
69. Kennedy, C. J., Rakoczy, P. E. & Constable, I. J. Lipofuscin of the retinal pigment epithelium: a review. *Eye* **9**, 763–771 (1995).
70. Radu, R. A. *et al.* Treatment with isotretinoin inhibits lipofuscin accumulation in a mouse model of recessive Stargardt's macular degeneration. *Proceedings of the National Academy of Sciences* **100**, 4742–4747 (2003).
71. Wolf, G. Lipofuscin and macular degeneration. *Nutrition reviews* **61**, 342–346 (2003).
72. Xu, G. *et al.* In *International Symposium on Biophotonics, Nanophotonics and Metamaterials*. 68–71 (IEEE, 2006).
73. Xu, H., Chen, M., Manivannan, A., Lois, N. & Forrester, J. V. Age-dependent accumulation of lipofuscin in perivascular and subretinal microglia in experimental mice. *Aging cell* **7**, 58–68 (2008).
74. Ach, T. *et al.* Quantitative Autofluorescence and Cell Density Maps of the Human Retinal Pigment Epithelium. *Investigative Ophthalmology & Visual Science* **55**, 4832–4841, <https://doi.org/10.1167/iovs.14-14802> (2014).
75. Neher, R. A. *et al.* Blind source separation techniques for the decomposition of multiply labeled fluorescence images. *Biophysical journal* **96**, 3791–3800 (2009).
76. Pengo, T., Muñoz-Barrutia, A., Zudaire, I. & Ortiz-de Solorzano, C. Efficient Blind Spectral Unmixing of Fluorescently Labeled Samples Using Multi-Layer Non-Negative Matrix Factorization. *PLOS ONE* **8**, e78504, <https://doi.org/10.1371/journal.pone.0078504> (2013).

Acknowledgements

The author would like to acknowledge that primary user-defined functions (e.g. HySime, VCA, MVSA, DUSAL) were kindly made available by their respective author, Jose M. P. Nascimento, under GNU licence. Moreover, the author would like to acknowledge Dr. Vivek Gupta, Faculty of Medicine and Health Sciences, Macquarie University for providing the RGC-5 cell line. M. P. acknowledges the support of an ARC DECRA fellowship (DE170100241).

Author Contributions

S.B.M., M.E.G. and E.M.G. designed this study and S.B.M. and E.M.G. designed unmixing methodology. A.G.A. prepared the retina cells experiments with Dr. Vivek Gupta, while S.B.M. performed image acquisition. S.B.M., M.E.G. analysed the retina data and S.B.M., E.M.G. analysed the artificial data. S.B.M., M.P. and E.M.G. wrote the manuscript. All authors have reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-15952-y>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017